MOLECULAR BIOMARKERS FOR CANCER CONTROL

EDITED BY: Xiaofeng Dai, Dongqing Wei and Jianying Zhang

PUBLISHED IN: Frontiers in Genetics and Frontiers in Neuroscience

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-200-4
DOI 10.3389/978-2-88974-200-4

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

MOLECULAR BIOMARKERS FOR CANCER CONTROL

Topic Editors:
Xiaofeng Dai, Jiangnan University, China
Dongqing Wei, Shanghai Jiao Tong University, China
Jianying Zhang, The University of Texas at El Paso, United States

Citation: Dai, X., Wei, D., Zhang, J., eds. (2022). Molecular Biomarkers for Cancer Control. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-200-4

Table of Contents

frontiers in Neuroscience

BRIEF RESEARCH REPORT
published: 11 August 2020
doi: 10.3389/fnins.2020.00822

A Panel of Synapse-Related Genes as a Biomarker for Gliomas

*Xiangwen Ji[1,2], Hongwei Zhang[3] and Qinghua Cui[1,2]**

[1] Department of Biomedical Informatics, Center for Non-coding RNA Medicine, MOE Key Lab of Cardiovascular Sciences, School of Basic Medical Sciences, Peking University, Beijing, China, [2] Department of Physiology and Pathophysiology, Center for Non-coding RNA Medicine, MOE Key Lab of Cardiovascular Sciences, School of Basic Medical Sciences, Peking University, Beijing, China, [3] Department of Neurosurgery, Sanbo Brain Hospital, Capital Medical University, Beijing, China

Gliomas are the most common primary brain cancers. In recent years, *IDH* mutation and 1p/19q codeletion have been suggested as biomarkers for the diagnosis, treatment, and prognosis of gliomas. However, these biomarkers are only effective for a part of glioma patients, and thus more biomarkers are still emergently needed. Recently, an electrochemical communication between normal neurons and glioma cells by neuro-glioma synapse has been reported. Moreover, it was discovered that breast-to-brain metastasis tumor cells have pseudo synapses with neurons, and these synapses were indicated to promote tumor progression and metastasis. Based on the above observations, we first curated a panel of 17 synapse-related genes and then proposed a metric, synapse score to quantify the "stemness" for each sample of 12 glioma gene expression datasets from TCGA, CGGA, and GEO. Strikingly, synapse score showed excellent predictive ability for the prognosis, diagnosis, and grading of gliomas. Moreover, being compared with the two established biomarkers, *IDH* mutation and 1p/19q codeletion, synapse score demonstrated independent and better predictive performance. In conclusion, this study proposed a quantitative method, synapse score, as an efficient biomarker for monitoring gliomas.

Keywords: glioma, synapse, biomarker, survival, WHO grade

OPEN ACCESS

Edited by:
Dongqing Wei,
Shanghai Jiao Tong University, China

Reviewed by:
Oksana Sorokina,
University of Edinburgh,
United Kingdom
Anatoly Sorokin,
University of Liverpool,
United Kingdom

****Correspondence:***
Qinghua Cui
cuiqinghua@hsc.pku.edu.cn

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Neuroscience

Received: *04 March 2020*
Accepted: *14 July 2020*
Published: *11 August 2020*

Citation:
Ji X, Zhang H and Cui Q (2020) A Panel of Synapse-Related Genes as a Biomarker for Gliomas. Front. Neurosci. 14:822. doi: 10.3389/fnins.2020.00822

INTRODUCTION

Brain and other nervous system cancers are estimated to take up 1.4% of new cancers but 2.9% of cancer deaths in 2019 (Brain and Other Nervous System Cancer, 2019). Gliomas are the most frequent of these cancers, including astrocytoma (including glioblastoma), oligodendroglioma, ependymoma, oligoastrocytoma (mixed glioma), malignant glioma, not otherwise specified (NOS) glioma, and a few rare histologies (Ostrom et al., 2016). The World Health Organization (WHO) classified gliomas into grades I to IV and introduced biomarkers of *IDH* mutation and 1p/19q codeletion in the 2016 edition (Louis et al., 2007; Wesseling and Capper, 2018). Glioblastoma (WHO grade IV) accounts for about half of gliomas, with a median survival of less than 2 years (Gramatzki et al., 2016; Ostrom et al., 2016). Gliomas with lower grade have a diverse prognosis, either progressing to be as poor as glioblastoma or living more than 10 years after effective treatment (Ruda and Soffietti, 2017).

Over the years, with the fast improvement of omics and big data technology, RNA sequencing has been developing toward lower cost and higher throughput, producing a large amount of biological and medical data, which provides great convenience for life science research (Bolouri et al., 2016). Impelled by advantage of big data analysis, numerous biomarkers have been found in the diagnosis and prognosis of gliomas (Kros et al., 2015). Gene set enrichment analysis (GSEA) provides a facility to extract effective information from a large number of RNA expression data (Subramanian et al., 2005). Moreover, single sample GSEA (ssGSEA) can calculate without group information and give every sample an enrichment score (Barbie et al., 2009). The Biomarkers such as *IDH* mutation and 1p/19q codeletion provided help for monitoring the development and prognosis of gliomas but are only effective for a part of patients (Aibaidula et al., 2017). Therefore, given the enormous severity of gliomas, more biomarkers are emergently needed.

It is recently reported that neuron and glioma have electrochemical communication through AMPA receptor-dependent synapses between presynaptic neurons and postsynaptic glioma cells (Venkataramani et al., 2019; Venkatesh et al., 2019). These observations suggest that the neural synaptic electrochemical connections promote glioma progression. Simultaneously, an appearance of glutamatergic "pseudo-tripartite" synapses between breast-to-brain metastasis tumor cells and neurons was observed (Zeng et al., 2019). Based on these anatomical and cytological findings, we hypothesized that the synapse-related genes can be used as a biomarker for glioma prognosis. To confirm this hypothesis, here we first curated a list of genes involved in synapse-related functions and then performed ssGSEA analysis for glioma gene expression datasets from the Cancer Genome Atlas (TCGA), the Chinese Glioma Genome Atlas (CGGA), and the Gene Expression Omnibus (GEO). Strikingly, these synapse-related genes were found to be an independent and effective biomarker for gliomas.

MATERIALS AND METHODS

Gene Expression Datasets and Analysis

RNAseq data, normalized in fragments per kilo-base per million mapped fragments, as well as sample and clinical information were obtained from TCGA data portal[1]. WHO grade, *IDH* mutation status, and 1p/19q codeletion status were obtained from the study by Ceccarelli et al. (2016). CGGA[2] provides tumor gene expression data for thousands of glioma patients (including one microarray and two RNAseq batches), as well as corresponding clinical data. The calculation and presentation of the results will be conducted separately due to different platforms and batches. In addition, glioma microarray gene expression profiling data (GSE4290, GSE16011, GSE50161, GSE52009, GSE54004, GSE61374, and GSE107850) were available at GEO datasets[3]. Gene expression data were structured with gene symbols as row names and sample IDs as column names; duplicate gene symbols were averaged using their median value.

Synapse-Related Genes Screening

Gene ontology (GO) terms, which were related to synapse, neuron, neurotransmitter transport, glutamate receptor, or cell junction, were selected from NCBI[4]. Using ssGSEA, we calculated enrichment scores (ESs) for each GO term and each sample in two CGGA RNAseq batches. The ssGSEA algorithm was performed by python (v3.6.8) package gseapy (v0.9.13), which is a python wrapper for GSEA and ssGSEA. The minimum number of genes in the gene set was set as 10, and the maximum was 1,000. As a result, 163 of 581 terms were retained. Cox regression models were used to calculate the hazard ratios (HRs) and *p*-values for ESs of each GO term. We used CoxPHFitter from python package lifelines (v0.23.7) to fit Cox models. Default parameters were used except the data frame and the column names of survival times and events. *P*-values were adjusted using Benjamini-Hochberg method. The false discovery rates (FDRs) of the two batches are multiplied to calculate the combined FDR (**Supplementary File S1**). The terms with different directions in two batches (HR < 1 in one batch and HR > 1 in the other) were excluded. Terms with top 10 smallest combined FDR values, except "peripheral nervous system neuron development" (GO:0048935) as gliomas are located in the central nervous system, are used for subsequent analysis (ionotropic glutamate receptor signaling pathway, AMPA glutamate receptor complex, regulation of short-term neuronal synaptic plasticity, dopaminergic synapse, synapse maturation, excitatory postsynaptic potential, parallel fiber to Purkinje cell synapse, synapse organization, and regulation of AMPA receptor activity). Next, we evaluated the HRs and *p*-values of 171 genes (eight genes are not in the datasets) from these nine GO terms and calculated the combined FDRs (**Supplementary File S2**). One hundred forty-four genes were filtered out with the same directions in two data batches. Then we obtained ESs of genes with top n ($n = 1, 2, \ldots, 144$) smallest combined FDRs for each sample in two data batches. After we evaluated the combined FDRs of every gene set, the gene set with the top 17 genes were selected (**Supplementary File S3**). Finally, using this synapse-related 17-gene set, we performed ssGSEA (default parameters) and calculated ESs for samples of TCGA, CGGA, and GEO datasets. We defined the ES as synapse score.

Statistical Analysis

Kaplan–Meier (K–M) curves and Cox proportional hazards regression were performed by R packages survival (v2.44-1.1) and survminer (v0.4.6) and python package lifelines (v0.23.7). Log rank test was used to calculate the difference between two K–M curves. Significance of difference between two groups of continuous variables was analyzed by two-sided Wilcoxon rank sum test. Receiver operating characteristic (ROC) curve and area under ROC curve (AUROC) were processed by R package pROC (Robin et al., 2011) (v1.15.3). All statistical significances above were calculated by R (v3.5.2). Spearman's correlation analysis was

[1] https://portal.gdc.cancer.gov/

[2] http://www.cgga.org.cn/

[3] https://www.ncbi.nlm.nih.gov/gds/

[4] https://ftp.ncbi.nih.gov/gene/DATA/

applied to evaluate the correlation using python package scipy (v1.2.1). *P*-values < 0.05 were considered significant.

RESULTS

The Screening of Synapse-Related Genes

In order to investigate whether synapse-related genes can be biomarkers for glioma patients, we first curated a list of GO terms associated with synapse, neuron, neurotransmitter transport, glutamate receptor, or cell junction. After excluding the terms with less than 10 or more than 1,000 genes, 163 terms were retained. Then we evaluated the survival prediction performances of these gene sets in two CGGA RNAseq batches using ssGSEA and Cox regression (**Supplementary File S1**). Most (118/163) of the terms were found to have HR < 1 in both data batches. The 10 best performed terms were further selected, and "peripheral nervous system neuron development" (GO:0048935) is excluded as gliomas are located in the central nervous system (**Figure 1A**). As a result, 171 genes were collected.

Afterward, we assessed the prognostic performances of these genes using Cox regression (**Figure 1B** and **Supplementary File S2**). 144 genes were filtered out with the same directions (both HRs < 1 or both HRs > 1) in two data batches. To further trim the gene set, we calculated ESs of gene sets which include the top n ($n = 1, 2, \ldots, 144$) best performed genes and evaluated their survival prognostic abilities (**Figure 1C** and **Supplementary File S3**). In most (142/144) cases, the gene sets performed better than any of the 144 genes on its own. Finally, the gene set with 17 genes was selected (**Figure 1B** and **Supplementary Table S1**), including profilin 1 (*PFN1*), SH3 and multiple ankyrin repeat domains 2 (*SHANK2*), calcium voltage-gated channel auxiliary subunit gamma 2 (*CACNG2*), tenascin R (*TNR*), shisa family member 7 (*SHISA7*), cholinergic receptor nicotinic beta 2 subunit (*CHRNB2*), glutamate ionotropic receptor NMDA type subunit 3A (*GRIN3A*), mitogen-activated protein kinase eight interacting protein 2 (*MAPK8IP2*), glutamate ionotropic receptor delta type

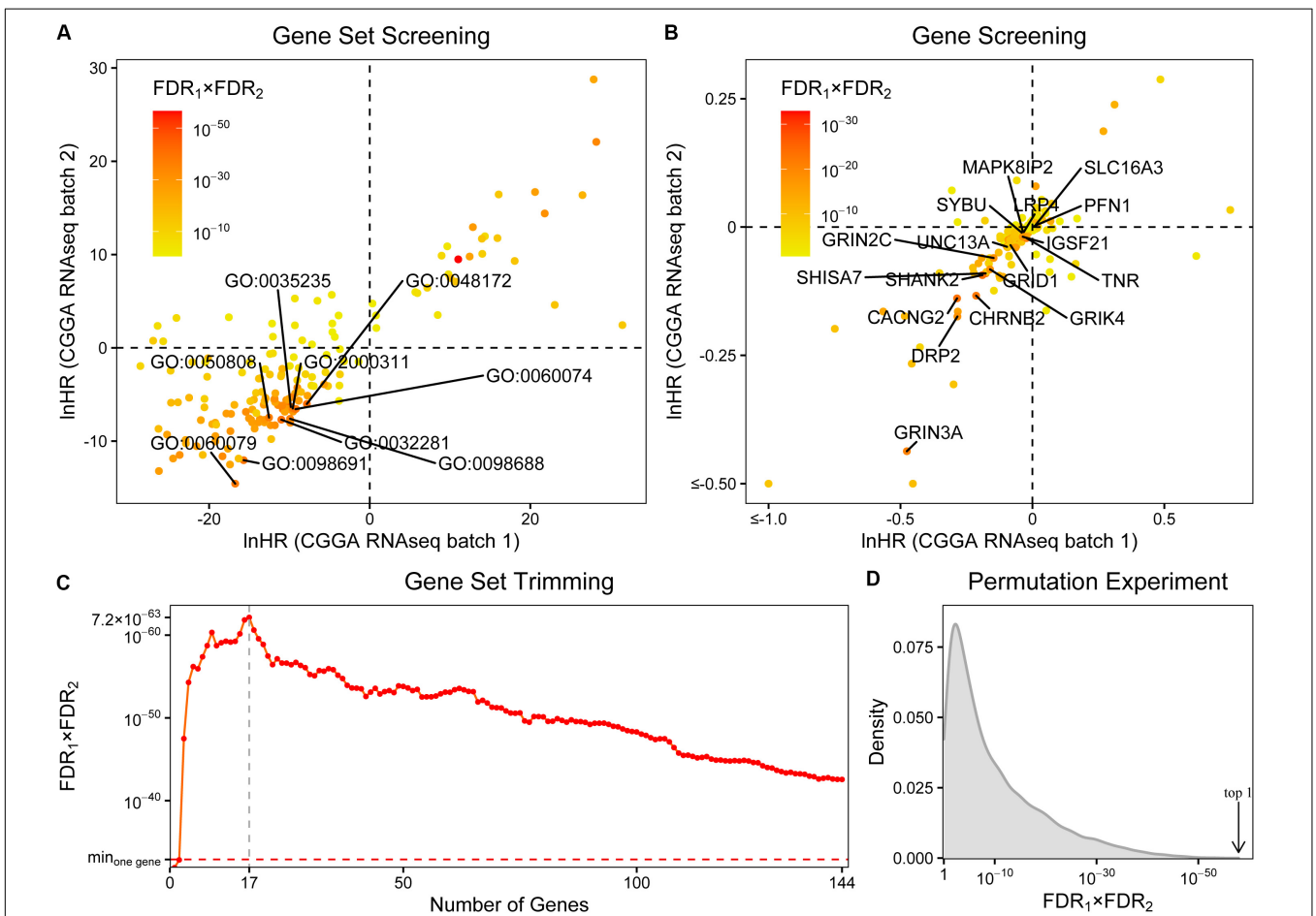

FIGURE 1 | The screening of synapse-related genes. **(A)** The survival prognostic performances of 163 synapse-related gene ontology terms. The chosen terms were labeled. **(B)** The prognostic performances of 171 collected genes. The finally selected genes were labeled. **(C)** The prognostic performances of gene sets with different sizes. $\text{Min}_{onegene}$: the best performance of single collected gene. The prognostic performances were evaluated by hazard ratios (HRs) of Cox regression. FDRs were calculated by Benjamini-Hochberg method. **(D)** The distribution of combined FDRs of 10,000 random 17-gene sets. Arrow: the combined FDR of our synapse-related 17-gene set.

subunit 1 (*GRID1*), unc-13 homolog A (*UNC13A*), LDL receptor-related protein 4 (*LRP4*), syntabulin (*SYBU*), solute carrier family 16 member 3 (*SLC16A3*), dystrophin-related protein 2 (*DRP2*), glutamate ionotropic receptor kainate type subunit 4 (*GRIK4*), glutamate ionotropic receptor NMDA type subunit 2C (*GRIN2C*), and immunoglobin superfamily member 21 (*IGSF21*).

As the next few gene sets, neurotransmitter uptake (GO:0001504), glutamate receptor signaling pathway (GO:0007215), NMDA selective glutamate receptor complex (GO:0017146), and glutamate receptor activity (GO:0008066) have combined FDRs of similar magnitudes ($<5 \times 10^{-35}$); the choice of top 10 terms could be too arbitrary. It may be useful to include them in subsequent analyses. We took these terms into consideration one by one and performed the same steps of screening and trimming described above. The inclusion of the term neurotransmitter uptake did not change the final result, and the same 17 genes were screened out. As for the other three terms, they all resulted in a 20-gene set, adding potassium voltage-gated channel subfamily B member 1 (*KCNB1*), nicastrin (*NCSTN*), and phospholipase C beta 1 (*PLCB1*) to the previous 17-gene set. But its combined *p*-value (2.38×10^{-64}) was a little worse than the previous 17-gene set (3.70×10^{-66}). Although there are still many significant terms, like focal adhesion (GO:0005925) at #15, we could not consider more due to the time complexity of subsequent screening and trimming. Finally, we decided to use the 17-gene set for future validations.

To further verify the efficiency of the 17-gene set, a permutation experiment was performed. After randomly selecting 10,000 sets with 17 genes from all the 23,271 genes that exist in both batches of datasets, we tested their prognostic abilities by ssGSEA and Cox regression. As a result, the combined FDR of the selected 17-gene set ranked first in all random gene sets ascendingly (**Figure 1D** and **Supplementary File S4**).

The Panel of Synapse-Related Genes Serves as a Novel Biomarker for Gliomas

Using the 17 collected synapse-related genes, we performed ssGSEA to TCGA, CGGA, and GEO datasets, and the ESs, defined as synapse score, were used for survival analysis. The results show that glioma patients with higher synapse scores have longer overall survival time (**Figure 2**). Cox regression analysis also shows the same results (**Table 1**). Moreover, patients with higher WHO grade have significantly lower synapse scores (**Figure 3** and **Supplementary Figure S1**), which agrees with the survival analysis. In addition, it is worthy to mention that there were normal brain samples in datasets GSE4290 (**Figure 3C**), GSE16011 (**Figure 3D**), and GSE50161 (**Supplementary Figure S1c**). The synapse scores of normal samples were significantly higher than glioma samples, suggesting that the synapse score shows an ability to distinguish between glioma and normal brain tissue by giving a cutoff value, which reveals a potential diagnostic application of synapse score. ROC analyses were further used to evaluate the diagnostic ability; the areas under the curve (AUCs) of GSE4290, GSE16011, and GSE50161 datasets are 0.89, 0.94, and 0.99, respectively (**Supplementary Figure S2**).

Comparison of Synapse Score With Established Biomarkers

IDH mutation and 1p/19q codeletion are two established biomarkers for gliomas. Both biomarkers provided great help for monitoring glioma development, but both are effective on only some patients. Therefore, it is interesting to explore whether synapse score is an independent biomarker and whether synapse score is better than the established biomarkers or not. For doing so, we first analyzed the relationship of synapse scores with *IDH* mutation and 1p/19q codeletion status. We found that *IDH*-mut gliomas were associated with significantly higher synapse scores than *IDH*-wt ones (**Supplementary Figures S3a–d**). And 1p/19q codeletion gliomas represent higher synapse scores than non-codeletion ones (**Supplementary Figures S3e–h**). Moreover, after removing the effects of the two established biomarkers using multivariate Cox regression model, we revealed that synapse score is an independent biomarker for predicting prolonged overall survival in gliomas (**Table 1**). In addition, the grading ability of synapse score is also independent of *IDH* mutation

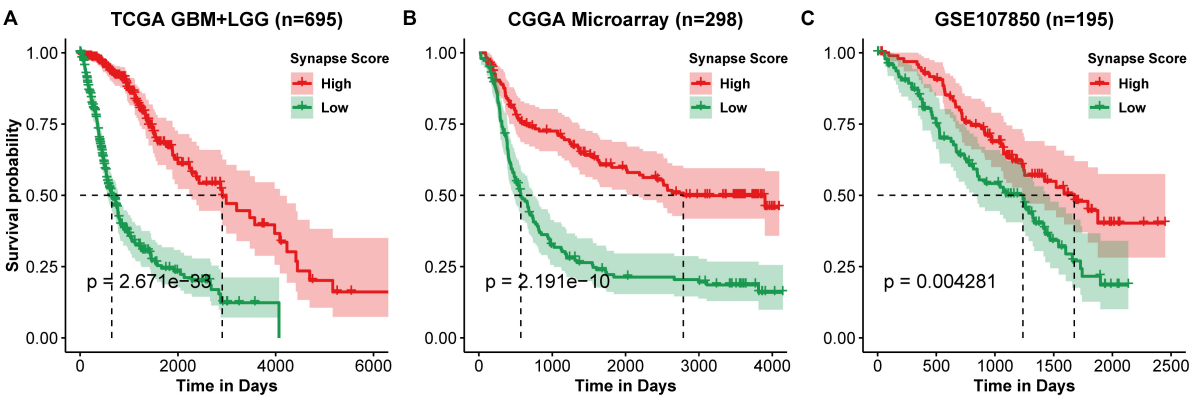

FIGURE 2 | Kaplan–Meier curve of overall survival. **(A)** TCGA lower grade glioma (LGG) and glioblastoma multiforme (GBM). **(B)** CGGA Microarray. **(C)** GSE107850 from GEO datasets. Group was separated by the median value of synapse scores. Differences between two curves were estimated by log-rank test.

TABLE 1 | The predictive ability of synapse score adjusted using WHO grade, *IDH* mutation, and 1p/19q codeletion.

Datasets	Hazard ratio (95% CI) (no. of samples)			
	Unadjusted	**Grade-adjusted**	***IDH* Status-adjusted**	**1p/19q Codeletion-adjusted**
CGGA Microarray	0.0379 (0.0151–0.0947)*** (n = 298)	0.270 (0.0852–0.856)* (n = 295)	0.0645 (0.0249–0.167)*** (n = 296)	0.0491 (9.04×10^{-3} to 0.267)*** (n = 91)
CGGA RNAseq batch 1	5.77×10^{-5} (1.15×10^{-5} to 2.89×10^{-4})*** (n = 311)	1.04×10^{-3} (1.57×10^{-4} to 6.85×10^{-3})*** (n = 307)	2.97×10^{-5} (3.49×10^{-6} to 2.52×10^{-4})*** (n = 310)	2.93×10^{-4} (4.61×10^{-5} to 1.86×10^{-3})*** (n = 303)
CGGA RNAseq batch 2	3.31×10^{-4} (9.18×10^{-5} to 1.20×10^{-3})*** (n = 619)	9.17×10^{-3} (1.99×10^{-3} to 0.0423)*** (n = 619)	5.62×10^{-3} (1.16×10^{-3} to 0.0272)*** (n = 574)	7.85×10^{-4} (1.72×10^{-4} to 3.58×10^{-3})*** (n = 555)
TCGA GBM + LGG	2.84×10^{-7} (4.00×10^{-8} to 2.02×10^{-6})*** (n = 695)	8.18×10^{-3} (4.13×10^{-4} to 0.162)** (n = 634)	4.83×10^{-3} (2.96×10^{-4} to 0.0788)*** (n = 685)	9.48×10^{-7} (1.03×10^{-7} to 8.77×10^{-6})*** (n = 688)
GSE107850	1.53×10^{-4} (2.20×10^{-6} to 0.0107)*** (n = 195)	3.09×10^{-4} (3.03×10^{-6} to 0.0314)*** (n = 180)		

*Hazard ratio (HR) and 95% confidence interval (95% CI) of synapse score using univariate and multivariate Cox proportional hazards regression models for gliomas were shown. HR with 95% CI that does not include one is considered significant. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

and 1p/19q codeletion (**Supplementary Figure S4**). Finally, we compared the survival predictive performance of synapse score, *IDH* mutation, and 1p/19q codeletion status (**Supplementary Table S2**). In most instances, synapse score outperforms *IDH* mutation and 1p/19q codeletion.

DISCUSSION

Given the recently revealed roles of neuro-glioma synapse in glioma development, here we curated a panel of 17 synapse-related genes and proposed the synapse score as a biomarker for the prognosis, grading, and diagnosis of gliomas. The synapse score was validated by more than 3,000 samples of 12 datasets from TCGA, CGGA, and GEO.

AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor, one type of glutamate receptors, was focused on in recent studies of neuron-glioma synapses (Venkataramani et al., 2019; Venkatesh et al., 2019). In our study, several AMPA glutamate receptor-related terms, such as ionotropic glutamate receptor signaling pathway (GO:0035235), AMPA glutamate receptor complex (GO:0032281), and regulation of AMPA receptor activity (GO:2000311), were filtered out to have strong survival predictive capacities, suggesting a significant role of AMPA receptor in gliomas.

In addition to AMPA receptor, other ionotropic glutamate receptor genes are also used in the 17-gene set, including N-Methyl-D-aspartate (NMDA) receptor [*GRIN3A* (Marco et al., 2013), *GRIN2C* (Collingridge et al., 2009)], kainate receptor [*GRIK4* (Arora et al., 2018)], and non-classical glutamate receptor such as glutamate delta-1 receptor [*GRID1* (Gupta et al., 2015)], suggesting that other ionotropic glutamate receptors also perform important functions in gliomas. Meanwhile, a gene from other synaptic receptors such as nicotinic acetylcholine receptor [*CHRNB2* (Diaz-Otero et al., 2008)] was also collected. More genes do not belong to receptors, and they perform neuronal-specific synthesis and glycosylation [*TNR* (Woodworth et al., 2004)], signal transduction [*MAPK8IP2* (Kennedy et al., 2007)], neurotransmission [*UNC13A* (Reddy-Alla et al., 2017), *LRP4* (Sun et al., 2016)], synapse formation [*LRP4* (Karakatsani et al., 2017)], inhibitory synapse differentiation [*IGSF21* (Tanabe et al., 2017)], and other functions in synapses (**Supplementary Table S1**).

Many of the selected genes have been found to be associated with neurological diseases, including autism [*SHANK2* (Monteiro and Feng, 2017; Won et al., 2012)], chronic pain [*CACNG2* (Bortsov et al., 2019; Nissenbaum et al., 2010)], epilepsy [*CHRNB2* (Diaz-Otero et al., 2008)], Huntington's disease [*GRIN3A* (Marco et al., 2013)], and neurodegenerative diseases [*SYBU* (Bereczki et al., 2018)] (**Supplementary Table S1**). However, only a few genes have been studied in gliomas. For example, *PFN1* has been found to be involved in tumor angiogenesis in glioblastoma (Fan et al., 2014) and was also found to be associated with poor prognosis in our study (HR > 1) (**Figure 1B**). According to a proteomics

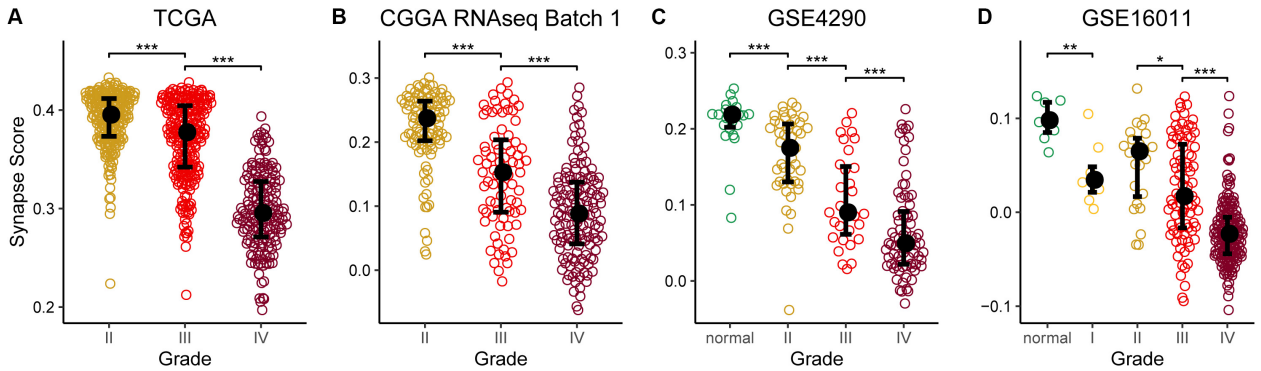

FIGURE 3 | Synapse scores were significantly lower in higher grade gliomas. **(A)** TCGA lower grade glioma (LGG) and glioblastoma multiforme (GBM). **(B)** CGGA RNAseq batch 1. **(C)** GSE4290. **(D)** GSE16011. Significances of difference between two groups were analyzed by two-side Wilcoxon rank sum test. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

study of gliomas (Bi et al., 2017), *TNR* is down-regulated in glioblastomas. A similar result was found in our study, that low expression of this gene was correlated with poor prognosis (HR $<$ 1). These studies validate our findings and suggest the research and application values of other synapse-related genes in gliomas.

When screening GO terms, there are 17 terms with the opposite directions (HR $>$ 1 in one batch and HR $<$ 1 in the other). Interestingly, all of these terms are negative in the batch 1 dataset and positive in the batch two dataset. There are 4 terms that are not significant in both datasets (FDR $\geq$ 0.05), which may be random effects. In addition, 10 terms are only significant in batch one, while one term is only significant in batch two, which may be caused by batch effect and differences of samples. Moreover, there are two terms that are significant but have opposite directions in two batches (neuroblast proliferation and neuron maturation). Given their low FDR ranking (FDR1: 140th, FDR2: 126th for neuroblast proliferation, FDR1: 144th, FDR2: 121st for neuron maturation out of 163 terms, ascendingly), these could be false positives. The practical effects of these terms need to be widely validated in future studies.

There are 121 significant (FDR1 $<$ 0.05 and FDR2 $<$ 0.05) synapse-related terms with the same direction of HRs in two batches of datasets, suggesting important roles of synapse-related genes in gliomas. But we could not consider all of the terms and genes due to time complexity. Finding the best gene set is a non-deterministic polynomial-time (NP) hard problem. In this paper, we used heuristic algorithms to find the optimal gene set by adding genes one by one in ascending order of combined FDR. It is known that heuristic algorithms do not always get the best results. There could be a gene set and a machine learning method with better prognostic ability using the synapse-related genes. Although our 17-gene set may not be the best result, it is still validated by a permutation experiment and 10 additional datasets and showed better prognostic capability than traditional biomarkers, *IDH* mutation, and 1p/19q codeletion, revealing the extensive research and application value of synapse-related genes in gliomas.

In spite of its ability as glioma biomarker for the identified synapse-related gene panel, it should be especially noted that the result seems the opposite of existing knowledge. That is, it was reported that neuro-glioma synapse could promote tumor progression and metastasis (Venkataramani et al., 2019; Venkatesh et al., 2019; Zeng et al., 2019), which thus can infer that synapse-related genes should result in a poorer prognosis. However, we revealed it is associated with a better but not poorer prognosis. One possible reason is that the more severe the disease is, the less the normal neurons exist. Molecular processes may play different roles in various cells, organs, and diseases. For example, as an important discovery in glioma research, *IDH* mutation is identified as one of the early events of gliomas, and the epigenetic changes caused by *IDH* mutation are considered as a main tumor driver (Turkalp et al., 2014). Nevertheless, clinical studies have found that *IDH* mutation can lead to a longer survival time (Cancer Genome Atlas Research Network et al., 2015). Similarly, immunotherapy, which has been widely used, was criticized for producing serious side effects (Moslehi et al., 2018). These instances suggest that the synapse-related gene panel could also have multiple aspects.

Analogously, *IDH*-mut and 1p/19q codeletion are typically biomarkers that promote glioma progression but benefit prognosis. Existing studies have focused on mechanisms that promote glioma, but the reasons for better prognosis are generally reported by clinical studies, such as better chemoradiotherapy sensitivity (Chen et al., 2017). We conjectured that synapses, *IDH* mutation, and 1p/19q codeletion shared a part of the mechanism that resulted in the observed phenomenon. The causations in synapses, mutations, and gliomas remain to be explored.

In summary, although the mechanism is unclear, we revealed that the proposed synapse score is an independent and potentially better biomarker for glioma overall survival and shows a predictive capacity in different grade gliomas

and normal brain tissues, which could be useful in the prognosis, grading, and diagnosis of gliomas.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/, http://firebrowse.org/, http://www.cgga.org.cn/, and https://www.ncbi.nlm.nih.gov/gds/.

AUTHOR CONTRIBUTIONS

QC conceived the project. XJ performed the analysis and conducted the experiments. XJ, HZ, and QC wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the Natural Science Foundation of China (81670462, 81970440, and 81921001 to QC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2020.00822/full#supplementary-material

REFERENCES

Aibaidula, A., Chan, A. K., Shi, Z., Li, Y., Zhang, R., and Yang, R. (2017). Adult IDH wild-type lower-grade gliomas should be further stratified. *Neuro Oncol.* 19, 1327–1337. doi: 10.1093/neuonc/nox078

Arora, V., Pecoraro, V., Aller, M. I., Roman, C., Paternain, A. V., and Lerma, J. (2018). Increased Grik4 gene dosage causes imbalanced circuit output and human disease-related behaviors. *Cell Rep.* 23, 3827–3838. doi: 10.1016/j.celrep.2018.05.086

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., and Dunn, I. F. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462, 108–112. doi: 10.1038/nature08460

Bereczki, E., Branca, R. M., Francis, P. T., Pereira, J. B., Baek, J. H., Hortobagyi, T., et al. (2018). Synaptic markers of cognitive decline in neurodegenerative diseases: a proteomic approach. *Brain* 141, 582–595. doi: 10.1093/brain/awx352

Bi, B. B., Li, F., Guo, J. S., Li, C. L., Jing, R. R., Lv, X., et al. (2017). Label-free quantitative proteomics unravels the importance of rna processing in glioma malignancy. *Neuroscience* 351, 84–95. doi: 10.1016/j.neuroscience.2017.03.023

Bolouri, H., Zhao, L. P., and Holland, E. C. (2016). Big data visualization identifies the multidimensional molecular landscape of human gliomas. *Proc. Natl. Acad. Sci. U.S.A.* 113, 5394–5399. doi: 10.1073/pnas.1601591113

Bortsov, A. V., Devor, M., Kaunisto, M. A., Kalso, E., Brufsky, A., Kehlet, H., et al. (2019). CACNG2 polymorphisms associate with chronic pain after mastectomy. *Pain* 160, 561–568. doi: 10.1097/j.pain.0000000000001432

Brain and Other Nervous System Cancer (2019). *Cancer Stat Facts*. Atlanta: CDC.

Cancer Genome Atlas Research Network, Brat, D. J., Verhaak, R. G., Aldape, K. D., Yung, W. K., Salama, S. R., et al. (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* 372, 2481–2498. doi: 10.1056/nejmoa1402121

Ceccarelli, M., Barthel, F. P., Malta, T. M., Sabedot, T. S., Salama, S. R., and Murray, B. A. (2016). Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* 164, 550–563.

Chen, R., Smith-Cohn, M., Cohen, A. L., and Colman, H. (2017). Glioma subclassifications and their clinical significance. *Neurotherapeutics* 14, 284–297. doi: 10.1007/s13311-017-0519-x

Collingridge, G. L., Olsen, R. W., Peters, J., and Spedding, M. (2009). A nomenclature for ligand-gated ion channels. *Neuropharmacology* 56, 2–5. doi: 10.1016/j.neuropharm.2008.06.063

Diaz-Otero, F., Quesada, M., Morales-Corraliza, J., Martinez-Parra, C., Gomez-Garre, P., and Serratosa, J. M. (2008). Autosomal dominant nocturnal frontal lobe epilepsy with a mutation in the CHRNB2 gene. *Epilepsia* 49, 516–520.

Fan, Y., Potdar, A. A., Gong, Y. Q., Eswarappa, S. M., Donnola, S., Lathia, J. D., et al. (2014). Profilin-1 phosphorylation directs angiocrine expression and glioblastoma progression through HIF-1 alpha accumulation. *Nat. Cell Biol.* 16, 445–U144.

Gramatzki, D., Dehler, S., Rushing, E. J., Zaugg, K., Hofer, S., Yonekawa, Y., et al. (2016). Glioblastoma in the Canton of Zurich, Switzerland revisited: 2005 to 2009. *Cancer* 122, 2206–2215. doi: 10.1002/cncr.30023

Gupta, S. C., Yadav, R., Pavuluri, R., Morley, B. J., Stairs, D. J., and Dravid, S. M. (2015). Essential role of GluD1 in dendritic spine development and GluN2B to GluN2A NMDAR subunit switch in the cortex and hippocampus reveals ability of GluN2B inhibition in correcting hyperconnectivity. *Neuropharmacology* 93, 274–284. doi: 10.1016/j.neuropharm.2015.02.013

Karakatsani, A., Marichal, N., Urban, S., Kalamakis, G., Ghanem, A., and Schick, A. (2017). Neuronal LRP4 regulates synapse formation in the developing CNS. *Development* 144, 4604–4615. doi: 10.1242/dev.150110

Kennedy, N. J., Martin, G., Ehrhardt, A. G., Cavanagh-Kyros, J., Kuan, C. Y., Rakic, P., et al. (2007). Requirement of JIP scaffold proteins for NMDA-mediated signal transduction. *Gene Dev.* 21, 2336–2346. doi: 10.1101/gad.1563107

Kros, J. M., Mustafa, D. M., Dekker, L. J., Sillevis Smitt, P. A., Luider, T. M., and Zheng, P. P. (2015). Circulating glioma biomarkers. *Neuro Oncol.* 17, 343–360.

Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., et al. (2007). The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* 114, 97–109.

Marco, S., Giralt, A., Petrovic, M. M., Pouladi, M. A., Martinez-Turrillas, R., and Martinez-Hernandez, J. (2013). Suppressing aberrant GluN3A expression rescues synaptic and behavioral impairments in Huntington's disease models. *Nat. Med.* 19, 1030–1038. doi: 10.1038/nm.3246

Monteiro, P., and Feng, G. P. (2017). SHANK proteins: roles at the synapse and in autism spectrum disorder. *Nat. Rev. Neurosci.* 18, 147–157. doi: 10.1038/nrn.2016.183

Moslehi, J. J., Salem, J. E., Sosman, J. A., Lebrun-Vignes, B., and Johnson, D. B. (2018). Increased reporting of fatal immune checkpoint inhibitor-associated myocarditis. *Lancet* 391:933. doi: 10.1016/s0140-6736(18)30533-6

Nissenbaum, J., Devor, M., Seltzer, Z., Gebauer, M., Michaelis, M., and Tal, M. (2010). Susceptibility to chronic pain following nerve injury is genetically affected by CACNG2. *Genome Res.* 20, 1180–1190. doi: 10.1101/gr.104976.110

Ostrom, Q. T., Gittleman, H., Xu, J., Kromer, C., Wolinsky, Y., Kruchko, C., et al. (2016). CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2009-2013. *Neuro Oncol.* 18, v1–v75. doi: 10.1093/neuonc/now207

Reddy-Alla, S., Bohme, M. A., Reynolds, E., Beis, C., Grasskamp, A. T., and Mampell, M. M. (2017). Stable positioning of Unc13 restricts synaptic vesicle fusion to defined release sites to promote synchronous neurotransmission. *Neuron* 95, 1350–1364.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. doi: 10.1186/1471-2105-12-77

Ruda, R., and Soffietti, R. (2017). Controversies in management of low-grade gliomas in light of new data from clinical trials. *Neuro Oncol.* 19, 143–144. doi: 10.1093/neuonc/now275

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Sun, X. D., Li, L., Liu, F., Huang, Z. H., Bean, J. C., and Jiao, H. F. (2016). Lrp4 in astrocytes modulates glutamatergic transmission. *Nat. Neurosci.* 19, 1010–1018. doi: 10.1038/nn.4326

Tanabe, Y., Naito, Y., Vasuta, C., Lee, A. K., Soumounou, Y., Linhoff, M. W., et al. (2017). IgSF21 promotes differentiation of inhibitory synapses via binding to neurexin2 alpha. *Nat. Commun.* 8:408.

Turkalp, Z., Karamchandani, J., and Das, S. (2014). IDH mutation in glioma: new insights and promises for the future. *JAMA Neurol.* 71, 1319–1325.

Venkataramani, V., Tanev, D. I, Strahle, C., Studier-Fischer, A., and Fankhauser, L. T. (2019). Glutamatergic synaptic input to glioma cells drives brain tumour progression. *Nature* 573, 532–538. doi: 10.1038/s41586-019-1564-x

Venkatesh, H. S., Morishita, W., Geraghty, A. C., Silverbush, D., Gillespie, S. M., and Arzt, M. (2019). Electrical and synaptic integration of glioma into neural circuits. *Nature* 573, 539–545.

Wesseling, P., and Capper, D. (2018). WHO 2016 Classification of gliomas. *Neuropathol. Appl. Neurobiol.* 44, 139–150. doi: 10.1111/nan.12432

Won, H., Lee, H. R., Gee, H. Y., Mah, W., Kim, J. I., and Lee, J. (2012). Autistic-like social behaviour in Shank2-mutant mice improved by restoring NMDA receptor function. *Nature* 486, 261–265. doi: 10.1038/nature11208

Woodworth, A., Pesheva, P., Fiete, D., and Baenziger, J. U. (2004). Neuronal-specific synthesis and glycosylation of tenascin-R. *J. Biol. Chem.* 279, 10413–10421. doi: 10.1074/jbc.m312466200

Zeng, Q., Michael, I. P., Zhang, P., Saghafinia, S., Knott, G., Jiao, W., et al. (2019). Synaptic proximity enables NMDAR signalling to promote brain metastasis. *Nature* 573, 526–531. doi: 10.1038/s41586-019-1576-6

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.

frontiers in Genetics

SYSTEMATIC REVIEW
published: 11 August 2020
doi: 10.3389/fgene.2020.00789

Serum-Derived microRNAs as Prognostic Biomarkers in Osteosarcoma: A Meta-Analysis

Huan Luo[1,2], Peng Wang[1,2], Hua Ye[1,2], Jianxiang Shi[2,3], Liping Dai[2,3], Xiao Wang[2,3], Chunhua Song[1,2], Jianying Zhang[1,2,3] and Jitian Li[4]**

[1] College of Public Health, Zhengzhou University, Zhengzhou, China, [2] Henan Key Laboratory of Tumor Epidemiology, Zhengzhou University, Zhengzhou, China, [3] Zhengzhou University, Henan Academy of Medical and Pharmaceutical Sciences, Zhengzhou, China, [4] Laboratory of Molecular Biology, Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), Zhengzhou, China

Recent reports suggest that microRNAs (miRNAs) may serve as prognostic biomarkers in osteosarcoma. Due to osteosarcoma's early metastasis and poor prognosis, it is very important to find novel prognostic biomarkers for improving osteosarcoma's prognosis. Herein we propose a meta-analysis for serum miRNA's prognostic value in osteosarcoma. In this study, the literature available from PubMed, Web of Science, Embase, and Cochrane Library databases was reviewed. The pooled hazard ratios (HRs) with their 95% confidence intervals (CIs) were calculated to evaluate miRNAs prognostic values. A total of 20 studies investigating serum miRNAs were included in this meta-analysis; the initial terminal point of these reports included overall survival (OS), progression-free survival (PFS), disease-free survival (DFS), and recurrence-free survival (RFS). For prognostic meta-analyses, the pooled HR for terminal events of higher expression of miRNAs and lower expression of miRNAs were 5.68 (95% CI 4.73–6.82, $P < 0.05$) and 3.78 (95% CI 3.27–4.37, $P < 0.05$), respectively. Additionally, subgroup analyses were conducted based on the analysis methods applied and clinicopathological features reported. In the pooled analyses, the miRNA expression levels are associated with poor prognosis according to both univariate and multivariate analyses. Furthermore, serum miRNAs (miRNA-195, miRNA-27a, miRNA-191, miRNA-300, miRNA-326, miRNA-497, miRNA-95-3p, miRNA-223, miRNA-491-5p, miRNA-124, miRNA-101, miRNA-139-5p, miRNA-194) were associated with poor OS and found to be closely correlated with clinical stage and distant metastasis in osteosarcoma. The results illustrate that low or high expression of these specific miRNAs are both potentially useful as prognostic serum biomarkers in osteosarcoma, and miRNAs (miRNA-195, miRNA-27a, miRNA-191, miRNA-300, miRNA-326, miRNA-497, miRNA-95-3p, miRNA-223, miRNA-491-5p, miRNA-124, miRNA-101, miRNA-139-5p, miRNA-194) may indicate clinical stage and metastasis in this form of cancer.

Keywords: miRNA, osteosarcoma, prognosis, serum, biomarker, meta-analysis

OPEN ACCESS

Edited by:
Dongqing Wei,
Shanghai Jiao Tong University, China

Reviewed by:
Leda Torres,
National Institute of Pediatrics, Mexico
Debashis Sahoo,
University of California, San Diego, United States

***Correspondence:**
Jianying Zhang
jianyingzhang@hotmail.com
Jitian Li
jitianlee@hotmail.com

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *31 March 2020*
Accepted: *02 July 2020*
Published: *11 August 2020*

Citation:
Luo H, Wang P, Ye H, Shi J, Dai L, Wang X, Song C, Zhang J and Li J (2020) Serum-Derived microRNAs as Prognostic Biomarkers in Osteosarcoma: A Meta-Analysis. Front. Genet. 11:789. doi: 10.3389/fgene.2020.00789

INTRODUCTION

Patient survival in osteosarcoma has improved in recent decades. Osteosarcoma is the most common malignant bone tumor, with a worldwide incidence of approximately one to three cases annually per million (Kansara et al., 2014). Current therapies include surgical resection and combination neoadjuvant chemotherapy, which is reported to have a curative effect in ∼70% of patients (Collins et al., 2013). Metastasis and recurrence are common challenges in refractory osteosarcoma, that worsen patient prognosis (Bielack et al., 2002). The highly malignant nature of osteosarcoma, as well as its high rates of recurrence and lung metastasis represent strong concerns (Jones et al., 2012; Ogawa et al., 2013). Clinically, histological examination of the biopsy specimens is preferred for the diagnosis or prognostic evaluation of osteosarcoma. However, such invasive tests may be burdensome when monitoring the progression of the disease, and the accuracy of diagnosis and prognostic evaluation may vary because of differences in sample collection and personnel. Therefore, it is essential to develop novel approaches for the timely diagnosis of osteosarcoma in order to achieve better prognosis (Gu et al., 2014).

MiRNAs are small (about 21-nucleotide-long) non-coding RNAs which can regulate gene expression (Filipowicz et al., 2008). Elevated or downregulated miRNAs may act as oncogenes or tumor suppressors in various cancers (Hayashita et al., 2005; He et al., 2005; Kent and Mendell, 2006; Tian et al., 2019). Additionally, miRNAs that are stable in serum or plasma, or in other biological samples, may have potential utility as diagnostic or prognostic biomarkers in different cancers (Calin and Croce, 2006; Esquela-Kerscher and Slack, 2006; Mitchell et al., 2008; Zhou et al., 2016). These findings show that miRNAs warrant attention as potential novel biomarkers for diagnosis or prognosis in osteosarcoma.

Although numerous recent studies have reported a correlation between prognosis in osteosarcoma and miRNA expression, none have demonstrated sufficient evidence for clinical translation of their findings. For instance, two previous meta-analyses have concluded the prognostic value of miRNA expression in osteosarcoma (Cheng et al., 2017; Kim et al., 2017); however, in these studies, either tissue or both tissue and blood were used as samples. Tissue samples' obtainment are invasive for patients than serum samples. To optimally obtain samples from patients and increase patients' acceptability, it is important for us to find novel serum biomarker for osteosarcoma. To the best of our knowledge, very few studies have provided robust evidence on the potential prognostic utility of serum miRNAs in osteosarcoma. Therefore, in the present work, we conducted a meta-analysis of studies in which serum samples were analyzed, to explore the prognostic value of miRNAs in osteosarcoma. Following which, subgroup analyses included analysis method and clinicopathological features were also explored to better analyze the prognostic value of various groups.

METHODS

This study was implemented according to the guidelines of the Meta-analysis of Observational Studies in Epidemiology (MOOSE) (Stroup et al., 2000), and the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Moher et al., 2009). We have completed the prognostic value of serum microRNA. In constructing the prognostic value of serum microRNA, we comply with the population, interventions, comparators, out-comes, and study designs (PICOS) principle to complete the research design.

Selection of Studies

The literature available in PubMed, Web of Science, Embase, and Cochrane Library databases, up to June 20, 2020, was investigated. The combination of search terms used was (osteosarcoma OR osteogenic sarcoma) AND (microRNA OR miRNA OR miR) AND (prognosis OR survival OR prognostic OR outcome). Only studies of the Chinese population published in English were included, and studies analyzing samples other than serum were excluded.

Inclusion and Exclusion Criteria

Inclusion criteria for studies in this review were as follows: (1) studies investigating the utility of miRNAs for evaluating prognosis in osteosarcoma, (2) serum miRNAs' assay method based on quantitative real-time polymerase chain reaction, (3) studies presenting sufficient data to allow calculation of HR and 95% CI, and (4) studies in which a cut-off value was defined. Studies were subject to the following exclusion criteria: (1) studies reporting duplicate data; studies in non-Chinese populations, (2) non-English publications, review articles, or meta-analysis, (3) studies reporting insufficient data for pooled analysis, and (4) studies of tissue, cell lines, or animal experiments.

Quality Assessment and Data Extraction

For prognostic meta-analyses, the quality of included studies was assessed using the Newcastle–Ottawa Scale (NOS), based on the following categories: selection, comparability, and outcome; the highest score was 9, with scores $\geq$6 indicating studies of high quality (Stang, 2010). The extracted data and information included were as follows: the first author, the year of publication, the country of origin, osteosarcoma sample size, sample type, cut off value, miRNAs characteristics, analysis methods, clinical outcomes, and detection methods. Two investigators retrieved and assessed the literature, respectively, and disagreements were resolved by extensive discussion.

Statistical Methods

All analyses were performed using STATA 12.0 software. Based on the information provided in the included studies, the pooled HRs with 95% CIs were calculated using this meta-analysis model. Forest plots were used to estimate the effect of miRNA expression on overall survival (OS), progression-free survival (PFS), disease-free survival (DFS), and recurrence free-survival

(RFS) (Hong et al., 2014; Zhang et al., 2014a,b; Cai et al., 2015; Tang et al., 2015; Wang N. G. et al., 2015; Wang T. et al., 2015; Yang et al., 2015; Cao et al., 2016; Dong et al., 2016; Liu et al., 2016; Niu et al., 2016; Pang et al., 2016; Wang S. N. et al., 2017; Wang Z. et al., 2017; Cong et al., 2018; Li et al., 2018; Yao et al., 2018; Zhou et al., 2018; Shi et al., 2020). I^2 index was used to assess the between-study heterogeneity, with $I^2 > 50\%$ indicating a large degree of heterogeneity; in this case, a random effect model was applied. $I^2 \leq 50\%$ implied that there was no significant heterogeneity, and the fixed effect model was used. Next, subgroup analyses were conducted to identify potential sources of heterogeneity and assess the prognostic value of different subgroups; the level of significance was set at $P < 0.05$. In addition, Begg's test (Begg and Mazumdar, 1994) and Egger's test (Egger et al., 1997) were performed to assess the publication bias; values of $P < 0.05$ indicated significant publication bias.

RESULTS

Characteristics of the Included Studies and Quality Assessment

The screening process for the studies is shown in detail in **Figure 1**. A primary search of the PubMed, Web of Science,

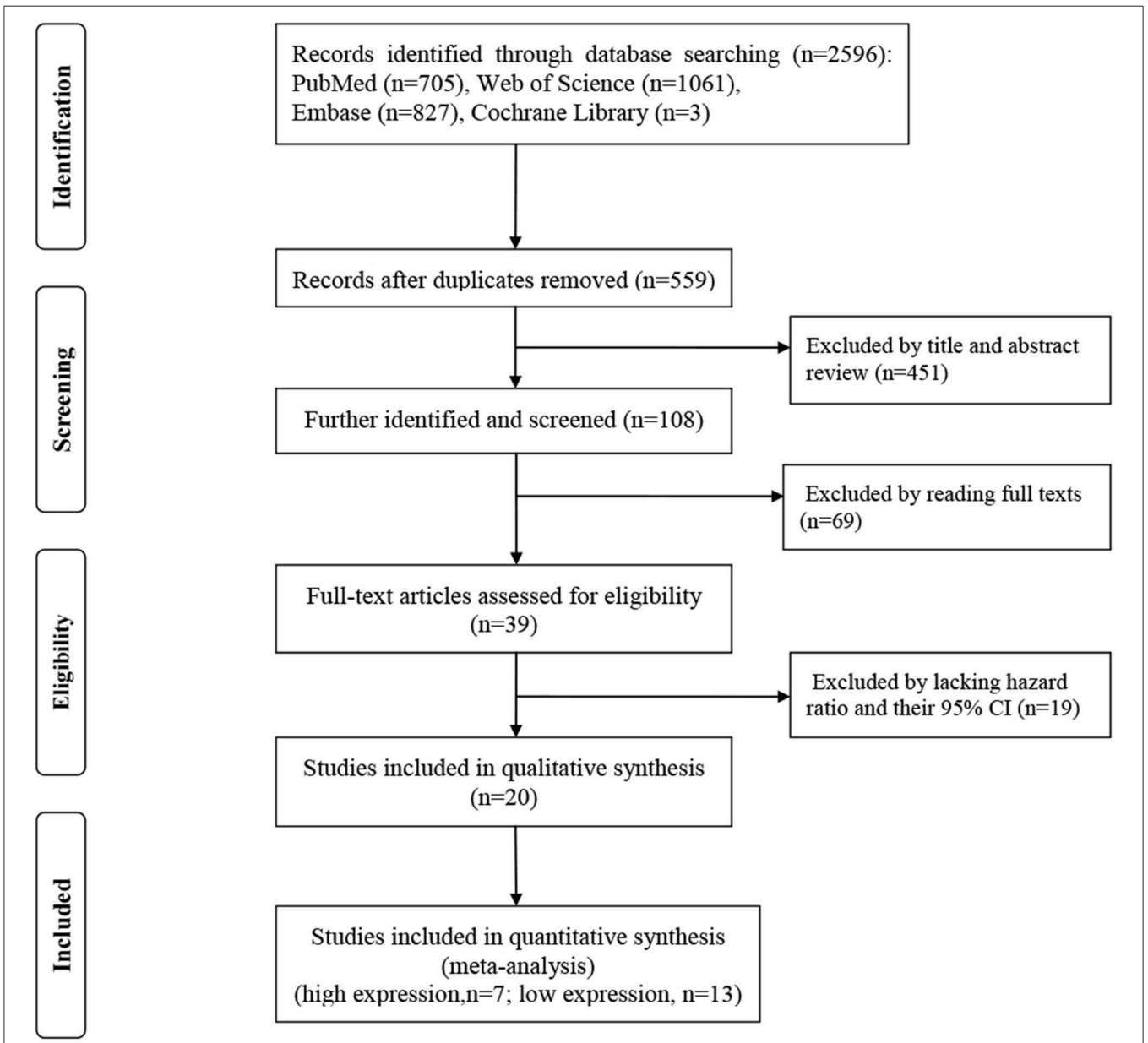

FIGURE 1 | Flow-process diagram of the study selection process.

Embase, and Cochrane Library databases, using the search strategy described, identified 2,596 articles. The innovative contribution of this work is focused on studies which using serum sample, herein studies using plasma or tissue were excluded, as such, all included studies examined serum samples. The data extracted from the included studies, the quality of the reports and heterogeneity are shown in **Tables 1**, **2**. The osteosarcoma sample size ranged from 60 to 185 subjects. The assay method was based on qRT-PCR. Cut off values were defined in the included studies to differentiate between high-expression miRNAs and low-expression miRNAs, and multivariate or univariate analyses were performed. The quality of each study was high according to the NOS (Stang, 2010). A total of 20 studies and 2,242 osteosarcoma patients were included in this prognostic meta-analysis.

Prognostic Accuracy and Subgroup Analyses

To analyze the prognostic value of miRNA expression in osteosarcoma, forest plots of data from the 19 studies, in accordance with HRs and their 95% CIs, are shown in **Figure 2**. The HRs were calculated on the basis of low-expression or high-expression miRNAs, respectively. HR >1 or <1 implied poor or good prognosis for patients with osteosarcoma, respectively. The pooled HRs for low- and high-expression miRNAs were 3.78 (95% CI 3.27–4.37, $P < 0.05$) and 5.68 (95% CI 4.73–6.82, $P < 0.05$), respectively, and both tended to be associated with a poorer outcome. Additionally, low-expression miRNAs were stratified by outcomes, including OS (pooled HR = 3.59, 95% CI 3.02–4.26, $P < 0.05$), DFS (pooled HR = 4.25, 95% CI 3.14–5.76, $P < 0.05$), and RFS (pooled HR = 4.34, 95% CI 2.48–7.60, $P < 0.05$). Furthermore, high-expression miRNAs were classified by outcomes, including OS (pooled HR = 5.98, 95% CI 4.58–7.80, $P < 0.05$), DFS (pooled HR = 4.80, 95% CI 3.53–6.53, $P < 0.05$), RFS (pooled HR = 6.82, 95% CI 2.13–21.88, $P < 0.05$), and PFS (pooled HR = 6.95, 95% CI 4.34–11.12, $P < 0.05$), and these miRNAs were also associated with poor prognosis in osteosarcoma.

Subgroup analyses were performed according to analysis method and clinicopathological features in order to explore the correlation of miRNA expression on prognosis in osteosarcoma, as shown in **Figures 3**, **4**, respectively. As shown in **Figure 3A**, subgroup analyses reveal that low expression levels of miRNA were significantly correlated with poor prognosis in osteosarcoma according to both multivariate (pooled HR = 3.88, 95% CI 3.29–4.58, $P < 0.05$) and univariate analyses (pooled HR = 3.47, 95% CI 2.57–4.68, $P < 0.05$). Similar results are shown in **Figure 3B**: multivariate (pooled HR = 5.28, 95%

TABLE 1 | The extracted data and quality assessment of literature on the prognostic utility of miRNAs in osteosarcoma.

Author/miRNA	Year	Country	Sample size	Sample type	Cut off value	miRNA expression with poor prognosis	Assay method	Analysis method	Outcome	NOS score
Hong miRNA-29a/29b	2014	China	80	Serum	2.85/3.27	High	qRT-PCR	M	OS, DFS	8
Zhang miRNA-196a/196b	2014	China	100	Serum	4.86/5.48	High	qRT-PCR	M	OS, DFS	8
Cai miRNA-195	2014	China	166	Serum	1.44	Low	qRT-PCR	M	OS, DFS	8
Zhang miRNA-133b/206	2014	China	100	Serum	2.66/2.84	Low	qRT-PCR	M	OS, DFS	8
Yang miRNA-221	2015	China	108	Serum	2.42	High	qRT-PCR	M	RFS, OS	7
Wang miRNA-152	2015	China	80	Serum	NR	Low	qRT-PCR	M	OS	6
Tang miRNA-27a	2015	China	166	Serum	3.70	High	qRT-PCR	M	OS, DFS	8
Wang miRNA-191	2015	China	100	Serum	3.56	High	qRT-PCR	M	OS, DFS	7
Dong miRNA-223	2016	China	112	Serum	1.21	Low	qRT-PCR	M	OS	6
Liu miRNA-300	2016	China	114	Serum	NR	High	qRT-PCR	U/M	OS/DFS	6
Cao miRNA-326	2016	China	60	Serum	Mean level	Low	qRT-PCR	M	OS	6
Niu miRNA-95-3p	2016	China	133	Serum	0.75	Low	qRT-PCR	M	OS	7
Pang miRNA-497	2016	China	185	Serum	4.80	Low	qRT-PCR	U/M	OS	7
Li miRNA-542-3p	2017	China	76	Serum	0.87	High	qRT-PCR	U/M	OS, PFS	7
Wang miRNA-491	2017	China	102	Serum	Mean level	Low	qRT-PCR	U/M	OS	7
Wang miRNA-491-5p	2017	China	72	Serum	NR	Low	qRT-PCR	M	OS, DFS	6
Cong miRNA-124	2017	China	114	Serum	0.37-fold	Low	qRT-PCR	M	OS, DFS	6
Yao miRNA-101	2018	China	152	Serum	Median level	Low	qRT-PCR	U/M	OS, RFS	6
Zhou miRNA-139-5p	2018	China	98	Serum	Median level	Low	qRT-PCR	U/M	OS	7
Shi miRNA-194	2020	China	124	Serum	Median level	Low	qRT-PCR	M	OS,DFS	7

M, Multivariate analysis; U, Univariate analysis; DFS, Disease-free survival; OS, Overall survival; RFS, Recurrence-free survival; PFS, Progression-free survival; NR, Not reported; high, high expression; low, low expression.

TABLE 2 | Prognostic value of the microRNAs expression profile mentioned in the literature.

microRNA	HR (95% CI)	Heterogeneity test		Sample size	Expression	Outcome	Number of microRNA
		I^2 (%)	Chi^2 (*P*)				
miR-133b	5.53 (2.58–11.83)	0.0%	0.939	100	Low expression	OS/DFS	2
miR-206	5.66 (2.69–11.88)	0.0%	0.914	100	Low expression	OS/DFS	2
miR-133b/206	9.48 (4.59–19.57)	0.0%	0.953	100	Low expression	OS/DFS	2
miR-195	4.23 (2.31–7.73)	0.0%	0.568	166	Low expression	OS/DFS	2
miR-152	0.13 (0.02–0.70)	–	–	80	Low expression	OS	1
miR-223	4.59 (1.84–11.45)	–	–	112	Low expression	OS	1
miR-326	3.90 (1.13–12.53)	–	–	60	Low expression	OS	1
miR-95-3p	4.22 (2.31–8.07)	–	–	133	Low expression	OS	1
miR-497	3.96 (2.39–6.58)	0.0%	0.868	185	Low expression	OS	2
miR-491	3.06 (1.56–6.00)	0.0%	0.928	102	Low expression	OS	2
miR-124	3.73 (2.27–6.12)	0.0%	0.841	114	Low expression	OS/DFS	2
miR-491-5p	2.68 (1.66–4.32)	0.0%	0.951	72	Low expression	DFS/OS	2
miR-101	4.16 (2.80–6.19)	0.0%	0.995	152	Low expression	OS/RFS	4
miR-139-5p	3.03 (2.17–4.23)	0.0%	0.707	98	Low expression	OS	2
miR-29a	5.68 (2.50–12.92)	0.0%	0.903	80	High expression	OS/DFS	2
miR-29b	5.71 (2.58–12.67)	0.0%	0.904	80	High expression	OS/DFS	2
miR-196a	6.59 (3.08–14.11)	0.0%	0.896	100	High expression	OS/DFS	2
miR-196b	6.64 (3.09–14.24)	0.0%	0.900	100	High expression	OS/DFS	2
miR-196a/196b	9.99 (4.82–20.69)	0.0%	0.979	100	High expression	OS/DFS	2
miR-221	7.26 (3.29–16.04)	0.0%	0.886	108	High expression	RFS/OS	2
miR-27a	3.36 (1.90–5.95)	0.0%	0.851	166	High expression	OS/DFS	2
miR-191	3.05 (1.75–5.31)	0.0%	0.593	100	High expression	OS/DFS	2
miR-300	5.07 (3.41–7.54)	0.0%	0.958	114	High expression	OS/DFS	4
miR-542-3p	7.83 (5.41–11.34)	0.0%	0.431	76	High expression	OS/PFS	4
miR-194	4.01 (2.53–6.36)	0.0%	0.695	124	Low expression	OS/DFS	2

DFS, Disease-free survival; OS, Overall survival; RFS, Recurrence-free survival; PFS, Progression-free survival; –, Not available.

CI 4.29–6.51, $P < 0.05$) and univariate analyses (pooled HR = 7.23, 95% CI 4.93–10.59, $P < 0.05$) both indicate that high expression of miRNA are correlated with poor prognosis in osteosarcoma. Additionally, as shown in **Figure 4**, the level of expression of serum miRNAs is closely correlated with distant metastasis (pooled HR = 3.30, 95% CI 2.77–3.94, $P < 0.05$), and clinical stage (pooled HR = 3.48, 95% CI 2.91–4.15, $P < 0.05$) in osteosarcoma.

Publication Bias and Sensitivity Analysis

The *P*-values for Begg's tests of low-expression miRNAs and high-expression miRNAs were 0.028 and 0.602, respectively, and the corresponding *P*-values for Egger's tests were 0.544 and 0.283. Furthermore, the funnel plots of Begg's and Egger's are all symmetrical demonstrating that there is no significant publication bias in this research (**Figure 5**). Sensitivity analyses revealed that none of the studies were outliers, suggesting that the pooled results of this research are credible.

DISCUSSION

Although osteosarcoma is the common malignant bone tumor (Kansara et al., 2014) and extensive progress has been made in the development of effective therapies (Bielack et al., 2002; Collins et al., 2013), patient prognosis remains unsatisfactory. Therefore, for improved treatment, management, and patient prognosis in osteosarcoma, the identification of novel prognostic biomarkers is critical. The potential utility of circulating miRNAs as non-invasive biomarkers has been demonstrated in several types of cancers (Zhou et al., 2016; Tan et al., 2019). Furthermore, blood testing is more easily accepted by patients than other invasive tests. Therefore, we conducted a meta-analysis of studies investigating the prognostic capacity of serum miRNAs in osteosarcoma. A more detailed subgroup analysis was also conducted to further examine the association between analysis methods, clinical stage, metastasis, and miRNA expression level in osteosarcoma.

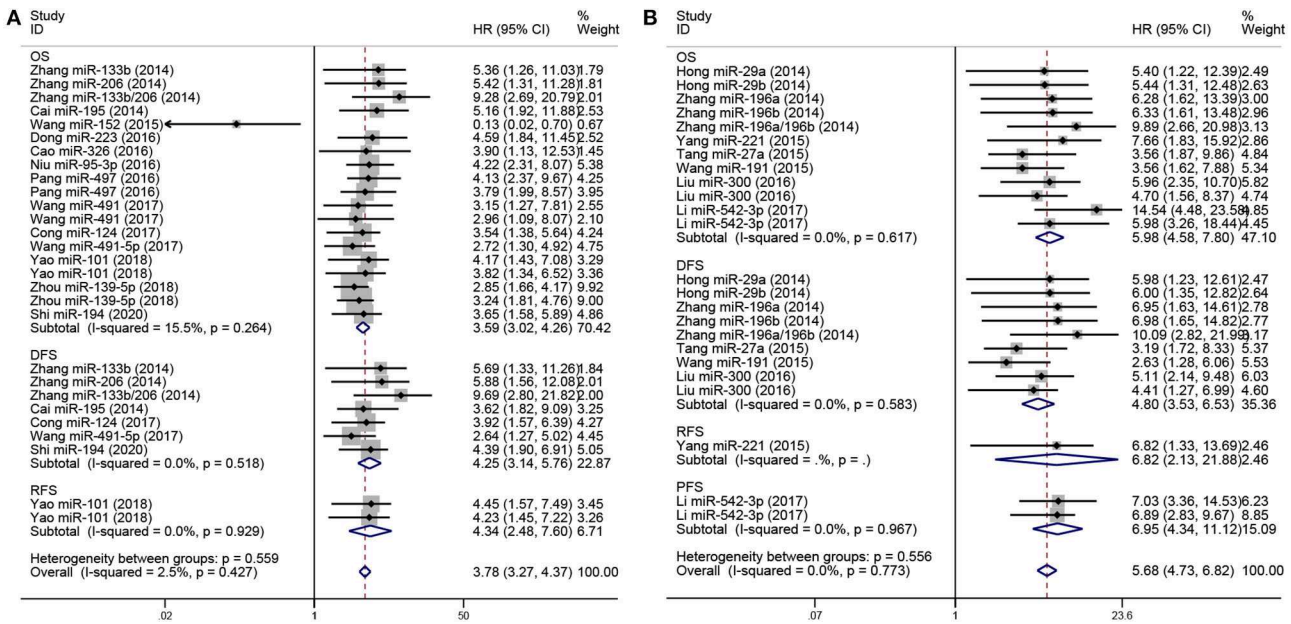

FIGURE 2 | Forest plot for miRNA expression and prognosis in osteosarcoma, stratified by outcomes included (OS, DFS, RFS, PFS). **(A)** Low-expression miRNAs correlated with poor prognosis in osteosarcoma, stratified by outcomes (OS, DFS, RFS). **(B)** High-expression miRNAs correlated with poor prognosis in osteosarcoma, stratified by event times (OS, DFS, RFS, PFS).

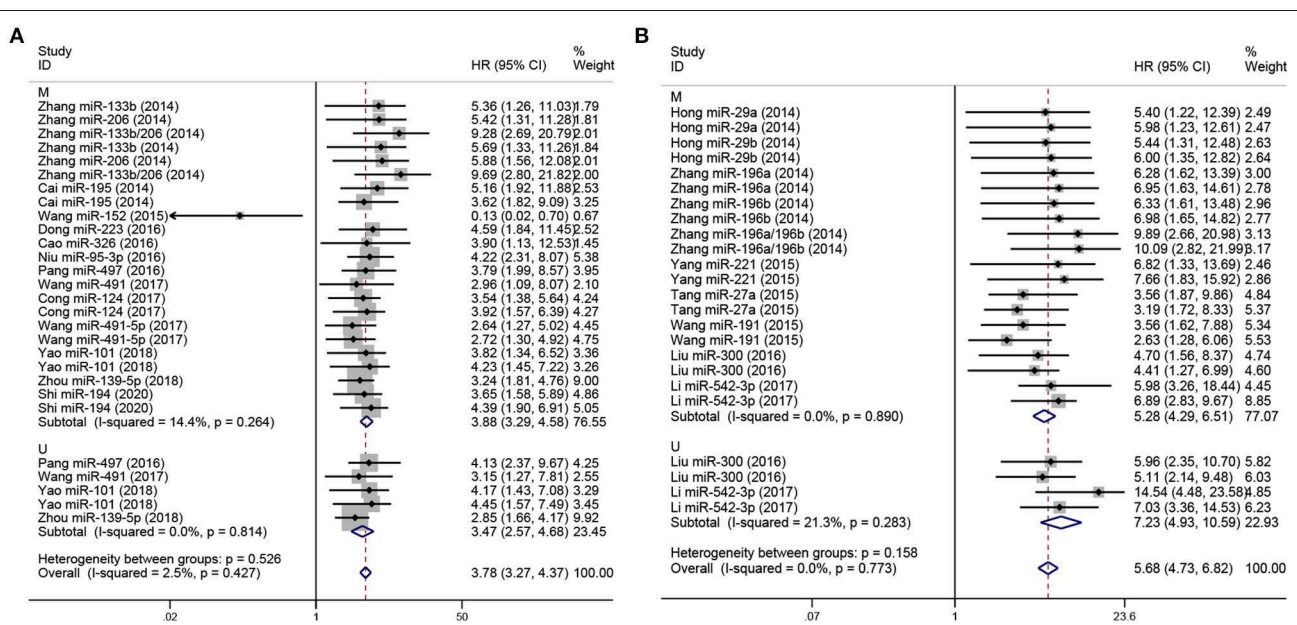

FIGURE 3 | Forest plot for miRNA expression and prognosis of osteosarcoma, stratified by analysis method. **(A)** Low-expression miRNAs correlated with poor prognosis in osteosarcoma, stratified by analysis method (M and U). **(B)** High-expression miRNAs with poor prognosis in osteosarcoma, stratified by analysis method (M and U). The *p*-values for heterogeneity of HR by subgroup, and overall, are shown. M, Multivariate analysis; U, Univariate analysis.

We investigated 20 studies on 23 different miRNAs in osteosarcoma in this meta-analysis; these included 9 highly expressed miRNAs (miRNA-29a, miRNA-29b, miRNA-196a, miRNA-196b, miRNA-221, miRNA-27a, miRNA-191, miRNA-542-3p, and miRNA-300) (Hong et al., 2014; Zhang et al., 2014a; Tang et al., 2015; Wang T. et al., 2015; Yang et al.,

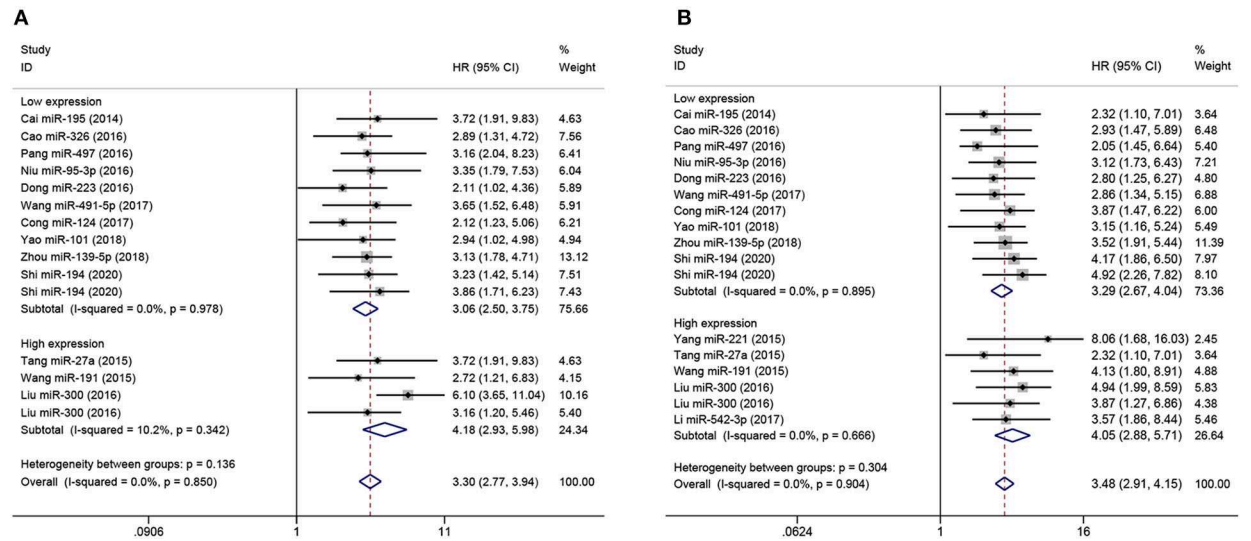

FIGURE 4 | Forest plot of the correlation between metastasis, clinical stage, and miRNA expression in studies with poor overall survival. **(A)** Correlation between miRNA level and tumor metastasis in osteosarcoma with poor overall survival, stratified by miRNAs expression (high-expression miRNA or low-expression miRNA). **(B)** Correlation between miRNA level and clinical stage in osteosarcoma with poor overall survival, stratified by miRNA expression (high-expression miRNA or low-expression miRNA).

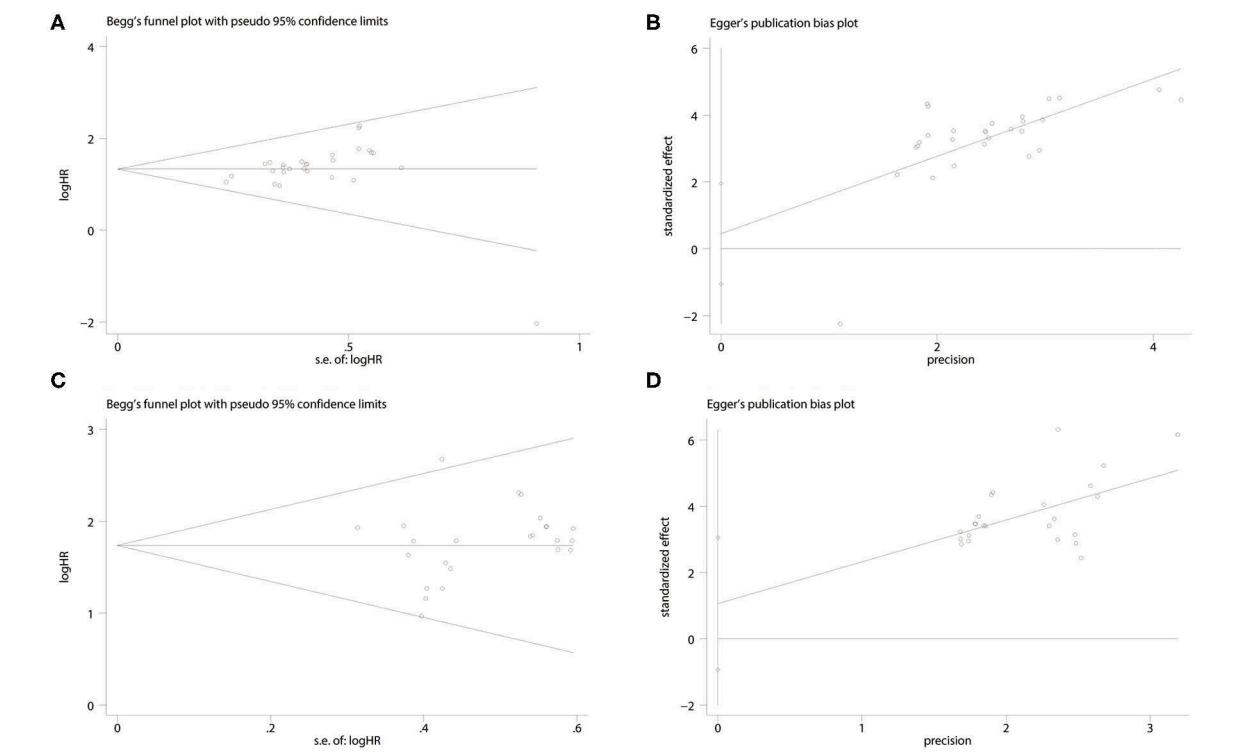

FIGURE 5 | Forest plot of the publication bias. **(A)** Begg's funnel plot of publication bias for the association between miRNA low expression and poor prognosis. **(B)** Egger's test of publication bias for the association between miRNA low expression and poor prognosis. **(C)** Begg's funnel plot of publication bias for the association between miRNA high expression and poor prognosis. **(D)** Egger's test of publication bias for the association between miRNA high expression and poor prognosis.

2015; Li et al., 2018), and 14 miRNAs with low expression (miRNA-195, miRNA-223, miRNA-497, miRNA-491, miRNA-124, miRNA-101, miRNA-139-5p, miRNA-326, miRNA-133b, miRNA-206, miRNA-152, miRNA-95-3p, and miRNA-491-5p, miR-194) (Cai et al., 2015; Cao et al., 2016; Dong et al., 2016; Pang et al., 2016; Wang S. N. et al., 2017; Cong et al., 2018; Yao et al., 2018; Zhou et al., 2018; Shi et al., 2020). A previous meta-analysis has reported that aberrant expression of miRNAs, in terms of both elevated and downregulated expression, are associated with poor prognosis in osteosarcoma (Cheng et al., 2017). Similar to the above findings, our data confirm the observation that serum miRNAs with aberrantly elevated or downregulated levels of expression are strongly correlated with poor prognosis for osteosarcoma. Among osteosarcoma patients, high expression of serum miRNAs (miRNA-196a and miRNA-196b) and combined expression of miRNA-196a/miRNA-196b were independent prognostic factors for OS and DFS (Zhang et al., 2014a). Further, Frampton et al. (2014) reported that miRNA-21, miRNA-23a, and miRNA-27a are highly expressed in pancreatic tumor, and their combination could serve as a prognostic biomarker in this tumor. The above results suggest that the development of an optimal panel of miRNA expression would be useful biomarker to improve prognosis in osteosarcoma.

The levels of miRNA-27a, miRNA-191, miRNA-195, miRNA-497, miRNA-223, miRNA-124, miRNA-101, miRNA-139-5p, miRNA-326, miRNA-95-3p, miRNA-491-5p, and miRNA-300, miRNA-194 (Cai et al., 2015; Tang et al., 2015; Wang T. et al., 2015; Cao et al., 2016; Dong et al., 2016; Liu et al., 2016; Niu et al., 2016; Pang et al., 2016; Wang Z. et al., 2017; Cong et al., 2018; Yao et al., 2018; Zhou et al., 2018; Shi et al., 2020) are potentially associated with clinical stage as well as the tumor metastasis in osteosarcoma. miRNA-27a, miRNA-191, and miRNA-300 are highly expressed in osteosarcoma, while miRNA-195, miRNA-497, miRNA-223, miRNA-124, miRNA-101, miRNA-139-5p, miRNA-326, miRNA-95-3p, miRNA-194, and miRNA-491-5p are expressed at low levels. In particular, the high expression of miR-191 may affect cancer progression through various pathways, and is associated with therapeutic outcomes or poor prognosis in cancers (Elyakim et al., 2010; Li et al., 2017). As such, miR-27a and miR-191 serve as regulatory factors in osteosarcoma. Low expression of miR-195 may act as a regulator in hepatocellular carcinoma cells, with potential utility in cancer therapy (Xu et al., 2009). In addition, low expression of serum miR-497 may be associated with tumor development (Kong et al., 2015). miR-124 was found to be expressed at low levels and correlated with invasion and metastasis in hepatocellular carcinoma (HCC), which may predict poor prognosis (Zheng et al., 2012). Low expression of miR-101 has been shown to inhibit the progression of bladder transitional cell carcinoma, and may serve as a tumor suppressor gene in this disease (Friedman et al., 2009). miR-139-5p has been shown to be expressed at low levels, and is considered to act as a regulator of cell proliferation, metastasis, apoptosis, and the cell cycle (Zhang et al., 2014c). In glioblastomas, miR-326 shows low expression and acts as a tumor suppressor (Nawaz et al., 2016). These results demonstrate, at a molecular level, that miRNAs have diagnostic or prognostic utility that could be extended to other tumors. miR-195, miR-497, miR-223, miR-124, miR-101, miR-139-5p, and miR-326 act as tumor suppressors in cancers, which is consistent with the results obtained in our study. Therefore, the conclusions that high-expression or low-expression miRNAs are potential novel biomarkers for predicting prognosis in osteosarcoma are reliable.

However, this study has some limitations. All relevant publications may not have been included in the databases, and specific subgroup analyses showed mild heterogeneity. HRs and RRs were merged into HRs in the included literature, potentially leading to slight logical errors, finally the included studies' population limited to Chinese. Despite these limitations, this study suggests that miRNAs have potential utility as novel prognostic markers in osteosarcoma. Further investigation of the dynamic expressional profile of miRNAs during the entire course of the development and treatment of osteosarcoma is needed for clinical implementation of miRNAs as biomarkers in either diagnosis or prognosis.

Despite these limitations, the results of this study suggest that serum miRNAs represent excellent biomarkers of prognosis in osteosarcoma. We conclude that miRNAs in this systematic review with aberrantly low or high expression are indicative of poor prognosis in osteosarcoma. Further, the expression of miRNAs(miRNA-195, miRNA-27a, miRNA-191, miRNA-300, miRNA-326, miRNA-497, miRNA-95-3p, miRNA-223, miRNA-491-5p, miRNA-124, miRNA-101, miRNA-139-5p, miRNA-194) is correlated with clinical stage and tumor metastasis in this disease. However, the implementation of miRNAs as biomarkers for monitoring cancer progression, clinical stage and distant metastasis and guiding therapeutic interventions improving poor prognosis. Future studies should aim to address these critical aspects.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

HL designed the study and collected data. HL and PW drafted the manuscript. PW, HY, JS, LD, XW, and CS contributed to the writing. JZ and JL contributed to the writing and review of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Science and Technology Major Project of China (No. 2018ZX10302205), the Major Project of Science and Technology in Henan Province (No. 161100311400), the Program of Natural Science Foundation of

Henan Province (No. 182300410009), and the Major Project of TCM research in Henan Province (No. 2018ZYZD01).

ACKNOWLEDGMENTS

The authors gratefully acknowledge JZ (Henan Academy of Medical and Pharmaceutical Sciences, Zhengzhou University) for providing financial support and language help, and JL (Laboratory of Molecular Biology, Henan Luoyang Orthopedic Hospital) for providing assistance with language and writing of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00789/full#supplementary-material

REFERENCES

Begg, C. B., and Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50, 1088–1101. doi: 10.2307/2533446

Bielack, S. S., Kempf-Bielack, B., Delling, G., Exner, G. U., Flege, S., Helmke, K., et al. (2002). Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. *J. Clin. Oncol.* 20, 776–790. doi: 10.1200/jco.2002.20.3.776

Cai, H., Zhao, H., Tang, J., and Wu, H. (2015). Serum miR-195 is a diagnostic and prognostic marker for osteosarcoma. *J. Surg. Res.* 194, 505–510. doi: 10.1016/j.jss.2014.11.025

Calin, G. A., and Croce, C. M. (2006). MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res.* 66, 7390–7394. doi: 10.1158/0008-5472.Can-06-0800

Cao, L., Wang, J., and Wang, P. Q. (2016). MiR-326 is a diagnostic biomarker and regulates cell survival and apoptosis by targeting Bcl-2 in osteosarcoma. *Biomed. Pharmacother.* 84, 828–835. doi: 10.1016/j.biopha.2016.10.008

Cheng, D., Qiu, X., Zhuang, M., Zhu, C., Zou, H., and Liu, Z. (2017). MicroRNAs with prognostic significance in osteosarcoma: a systemic review and meta-analysis. *Oncotarget* 8, 81062–81074. doi: 10.18632/oncotarget.19009

Collins, M., Wilhelm, M., Conyers, R., Herschtal, A., Whelan, J., Bielack, S., et al. (2013). Benefits and adverse events in younger versus older patients receiving neoadjuvant chemotherapy for osteosarcoma: findings from a meta-analysis. *J. Clin. Oncol.* 31, 2303–2312. doi: 10.1200/jco.2012.43.8598

Cong, C., Wang, W., Tian, J., Gao, T., Zheng, W., and Zhou, C. (2018). Identification of serum miR-124 as a biomarker for diagnosis and prognosis in osteosarcoma. *Cancer Biomark. Sect. A Dis. Mark.* 21, 449–454. doi: 10.3233/CBM-170672

Dong, J., Liu, Y., Liao, W., Liu, R., Shi, P., and Wang, L. (2016). miRNA-223 is a potential diagnostic and prognostic marker for osteosarcoma. *J. Bone Oncol.* 5, 74–79. doi: 10.1016/j.jbo.2016.05.001

Egger, M., Davey Smith, G., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315, 629–634. doi: 10.1136/bmj.315.7109.629

Elyakim, E., Sitbon, E., Faerman, A., Tabak, S., Montia, E., Belanis, L., et al. (2010). hsa-miR-191 is a candidate oncogene target for hepatocellular carcinoma therapy. *Cancer Res.* 70, 8077–8087. doi: 10.1158/0008-5472.Can-10-1313

Esquela-Kerscher, A., and Slack, F. J. (2006). Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer* 6, 259–269. doi: 10.1038/nrc1840

Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* 9, 102–114. doi: 10.1038/nrg2290

Frampton, A. E., Castellano, L., Colombo, T., Giovannetti, E., Krell, J., Jacob, J., et al. (2014). MicroRNAs cooperatively inhibit a network of tumor suppressor genes to promote pancreatic tumor growth and progression. *Gastroenterology* 146, 268–277. doi: 10.1053/j.gastro.2013.10.010

Friedman, J. M., Liang, G., Liu, C. C., Wolff, E. M., Tsai, Y. C., Ye, W., et al. (2009). The putative tumor suppressor microRNA-101 modulates the cancer epigenome by repressing the polycomb group protein EZH2. *Cancer Res.* 69, 2623–2629. doi: 10.1158/0008-5472.Can-08-3114

Gu, J., Li, J., Huang, M., Zhang, Z., Li, D., Song, G., et al. (2014). Identification of osteosarcoma-related specific proteins in serum samples using surface-enhanced laser desorption/ionization-time-of-flight mass spectrometry. *J. Immunol. Res.* 2014:649075. doi: 10.1155/2014/649075

Hayashita, Y., Osada, H., Tatematsu, Y., Yamada, H., Yanagisawa, K., Tomida, S., et al. (2005). A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res.* 65, 9628–9632. doi: 10.1158/0008-5472.Can-05-2352

He, L., Thomson, J. M., Hemann, M. T., Hernando-Monge, E., Mu, D., Goodson, S., et al. (2005). A microRNA polycistron as a potential human oncogene. *Nature* 435, 828–833. doi: 10.1038/nature03552

Hong, Q., Fang, J., Pang, Y., and Zheng, J. (2014). Prognostic value of the microRNA-29 family in patients with primary osteosarcomas. *Med. Oncol.* 31:37. doi: 10.1007/s12032-014-0037-1

Jones, K. B., Salah, Z., Del Mare, S., Galasso, M., Gaudio, E., Nuovo, G. J., et al. (2012). miRNA signatures associate with pathogenesis and progression of osteosarcoma. *Cancer Res.* 72, 1865–1877. doi: 10.1158/0008-5472.Can-11-2663

Kansara, M., Teng, M. W., Smyth, M. J., and Thomas, D. M. (2014). Translational biology of osteosarcoma. *Nat. Rev. Cancer* 14, 722–735. doi: 10.1038/nrc3838

Kent, O. A., and Mendell, J. T. (2006). A small piece in the cancer puzzle: microRNAs as tumor suppressors and oncogenes. *Oncogene* 25, 6188–6196. doi: 10.1038/sj.onc.1209913

Kim, Y. H., Goh, T. S., Lee, C. S., Oh, S. O., Kim, J. I., Jeung, S. H., et al. (2017). Prognostic value of microRNAs in osteosarcoma: a meta-analysis. *Oncotarget* 8, 8726–8737. doi: 10.18632/oncotarget.14429

Kong, X. J., Duan, L. J., Qian, X. Q., Xu, D., Liu, H. L., Zhu, Y. J., et al. (2015). Tumor-suppressive microRNA-497 targets IKKbeta to regulate NF-kappaB signaling pathway in human prostate cancer cells. *Am. J. Cancer Res.* 5, 1795–1804.

Li, H., Zhou, Z. Q., Yang, Z. R., Tong, D. N., Guan, J., Shi, B. J., et al. (2017). MicroRNA-191 acts as a tumor promoter by modulating the TET1-p53 pathway in intrahepatic cholangiocarcinoma. *Hepatology* 66, 136–151. doi: 10.1002/hep.29116

Li, Q., Song, S., Ni, G., Li, Y., and Wang, X. (2018). Serum miR-542-3p as a prognostic biomarker in osteosarcoma. *Cancer Biomark. Sect. A Dis. Mark.* 21, 521–526. doi: 10.3233/cbm-170255

Liu, J. D., Xin, Q., Tao, C. S., Sun, P. F., Xu, P., Wu, B., et al. (2016). Serum miR-300 as a diagnostic and prognostic biomarker in osteosarcoma. *Oncol. Lett.* 12, 3912–3918. doi: 10.3892/ol.2016.5214

Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L., et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10513–10518. doi: 10.1073/pnas.0804549105

Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann. Intern. Med.* 151, 264–269. doi: 10.7326/0003-4819-151-4-200908180-00135

Nawaz, Z., Patil, V., Paul, Y., Hegde, A. S., Arivazhagan, A., Santosh, V., et al. (2016). PI3 kinase pathway regulated miRNome in glioblastoma: identification of miR-326 as a tumour suppressor miRNA. *Mol. Cancer* 15:74. doi: 10.1186/s12943-016-0557-8

Niu, J., Sun, Y., Guo, Q., Niu, D., and Liu, B. (2016). Serum miR-95-3p is a diagnostic and prognostic marker for osteosarcoma. *SpringerPlus* 5:1947. doi: 10.1186/s40064-016-3640-0

Ogawa, K., Seki, T., Onji, T., Adachi, N., Tanaka, S., Hide, I., et al. (2013). Mutant gammaPKC that causes spinocerebellar ataxia type 14 upregulates Hsp70, which protects cells from the mutant's cytotoxicity. *Biochem. Biophys. Res. Commun.* 440, 25–30. doi: 10.1016/j.bbrc.2013.09.013

Pang, P. C., Shi, X. Y., Huang, W. L., and Sun, K. (2016). miR-497 as a potential serum biomarker for the diagnosis and prognosis of osteosarcoma. *Eur. Rev. Med. Pharmacol. Sci.* 20, 3765–3769.

Shi, L., Xie, C., Zhu, J., and Chen, X. (2020). Downregulation of serum miR-194 predicts poor prognosis in osteosarcoma patients. *Ann. Diagn. Pathol.* 46:151488. doi: 10.1016/j.anndiagpath.2020.151488

Stang, A. (2010). Critical evaluation of the Newcastle–Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur. J. Epidemiol.* 25, 603–605. doi: 10.1007/s10654-010-9491-z

Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., et al. (2000). Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 283, 2008–2012. doi: 10.1001/jama.283.15.2008

Tan, C., Cao, J., Chen, L., Xi, X., Wang, S., Zhu, Y., et al. (2019). Noncoding RNAs Serve as diagnosis and prognosis biomarkers for hepatocellular carcinoma. *Clin. Chem.* 65, 905–915. doi: 10.1373/clinchem.2018.301150

Tang, J., Zhao, H., Cai, H., and Wu, H. (2015). Diagnostic and prognostic potentials of microRNA-27a in osteosarcoma. *Biomed. Pharmacother.* 71, 222–226. doi: 10.1016/j.biopha.2015.01.025

Tian, X. P., Huang, W. J., Huang, H. Q., Liu, Y. H., Wang, L., Zhang, X., et al. (2019). Prognostic and predictive value of a microRNA signature in adults with T-cell lymphoblastic lymphoma. *Leukemia* 33, 2454–2465. doi: 10.1038/s41375-019-0466-0

Wang, N. G., Wang, D. C., Tan, B. Y., Wang, F., and Yuan, Z. N. (2015). Down-regulation of microRNA152 is associated with the diagnosis and prognosis of patients with osteosarcoma. *Int. J. Clin. Exp. Pathol.* 8, 9314–9319.

Wang, S. N., Luo, S., Liu, C., Piao, Z., Gou, W., Wang, Y., et al. (2017). miR-491 inhibits osteosarcoma lung metastasis and chemoresistance by targeting alphaB-crystallin. *Mol. Ther. J. Am. Soc. Gene Ther.* 25, 2140–2149. doi: 10.1016/j.ymthe.2017.05.018

Wang, T., Ji, F., Dai, Z., Xie, Y., and Yuan, D. (2015). Increased expression of microRNA-191 as a potential serum biomarker for diagnosis and prognosis in human osteosarcoma. *Cancer Biomark. Sect. A Dis. Mark.* 15, 543–550. doi: 10.3233/cbm-150493

Wang, Z., Jiang, C., Chang, X., and Dai, Y. (2017). Low miR-491-5p is an unfavorable prognostic marker for osteosarcoma. *Int. J. Clin. Exp. Pathol.* 10, 3304–3309.

Xu, T., Zhu, Y., Xiong, Y., Ge, Y. Y., Yun, J. P., and Zhuang, S. M. (2009). MicroRNA-195 suppresses tumorigenicity and regulates G1/S transition of human hepatocellular carcinoma cells. *Hepatology* 50, 113–121. doi: 10.1002/hep.22919

Yang, Z., Zhang, Y., Zhang, X., Zhang, M., Liu, H., Zhang, S., et al. (2015). Serum microRNA-221 functions as a potential diagnostic and prognostic marker for patients with osteosarcoma. *Biomed. Pharmacother.* 75, 153–158. doi: 10.1016/j.biopha.2015.07.018

Yao, Z. S., Li, C., Liang, D., Jiang, X. B., Tang, J. J., Ye, L. Q., et al. (2018). Diagnostic and prognostic implications of serum miR-101 in osteosarcoma. *Cancer Biomark. Sect. A Dis. Mark.* 22, 127–133. doi: 10.3233/cbm-171103

Zhang, C., Yao, C., Li, H., Wang, G., and He, X. (2014a). Combined elevation of microRNA-196a and microRNA-196b in sera predicts unfavorable prognosis in patients with osteosarcomas. *Int. J. Mol. Sci.* 15, 6544–6555. doi: 10.3390/ijms15046544

Zhang, L., Dong, Y., Zhu, N., Tsoi, H., Zhao, Z., Wu, C. W., et al. (2014c). microRNA-139-5p exerts tumor suppressor function by targeting NOTCH1 in colorectal cancer. *Mol. Cancer* 13:124. doi: 10.1186/1476-4598-13-124

Zhang, C., Yao, C., Li, H., Wang, G., and He, X. (2014b). Serum levels of microRNA-133b and microRNA-206 expression predict prognosis in patients with osteosarcoma. *Int. J. Clin. Exp. Pathol.* 7, 4194–4203.

Zheng, F., Liao, Y. J., Cai, M. Y., Liu, Y. H., Liu, T. H., Chen, S. P., et al. (2012). The putative tumour suppressor microRNA-124 modulates hepatocellular carcinoma cell aggressiveness by repressing ROCK2 and EZH2. *Gut* 61, 278–289. doi: 10.1136/gut.2011.239145

Zhou, L., Ma, X., Yue, J., Chen, T., Wang, X. Y., Wang, Z. W., et al. (2018). The diagnostic effect of serum miR-139-5p as an indicator in osteosarcoma. *Cancer Biomark. Sect. A Dis. Mark.* 23, 561–567. doi: 10.3233/cbm-181744

Zhou, S., Wang, B., Hu, J., Zhou, Y., Jiang, M., Wu, M., et al. (2016). miR-421 is a diagnostic and prognostic marker in patients with osteosarcoma. *Tumour Biol. J. Int. Soc. Oncodev. Biol. Med.* 37, 9001–9007. doi: 10.1007/s13277-015-4578-5

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 September 2020
doi: 10.3389/fgene.2020.572983

Construction and Characterization of a Synergistic lncRNA–miRNA Network Reveals a Crucial and Prognostic Role of lncRNAs in Colon Cancer

*Bin Zhao[1†], Xiusheng Qu[2†], Xin Lv[3], Qingdong Wang[4], Deqiang Bian[5], Fan Yang[1], Xingwang Zhao[1], Zhiwu Ji[1], Jian Ni[1], Yan Fu[1], Guorong Xin[1] and Haitao Yu[1]**

[1] Department of Proctology, First Affiliated Hospital of Jiamusi University, Jiamusi, China, [2] Department of Chemoradiotherapy, First Affiliated Hospital of Jiamusi University, Jiamusi, China, [3] Department of General Surgery, Samii Medical Center, Shenzhen, China, [4] Department of Anesthesiology, First Affiliated Hospital of Jiamusi University, Jiamusi, China, [5] Scientific Research Departments, First Affiliated Hospital of Jiamusi University, Jiamusi, China

Non-coding RNAs such as long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) have been found to be indispensable factors in carcinogenesis and cancer development. Numerous studies have explored the regulatory functions of these molecules and identified the synergistic interactions among lncRNAs or miRNAs, while those between lncRNAs and miRNAs remain to be investigated. In this study, we constructed and characterized an lncRNA–miRNA synergistic network following a four-step approach by integrating the regulatory pairs and expression profiles. The synergistic interactions with more shared regulatory mRNAs were found to have higher interactional intensity. Through the analysis of nodes in the network, we found that lncRNAs played roles that are more central and had similar synergistic interactions with their neighbors when compared with miRNAs. In addition, known colon adenocarcinoma (COAD)-related RNAs were found to be enriched in this synergistic network, with higher degrees, betweenness, and closeness. Finally, we proposed a risk score model to predict the clinical outcome for COAD patients based on two prognostic hub lncRNAs, MEG3 and ZEB1-AS1. Moreover, the hierarchical networks of these two lncRNAs could contribute to the understanding of the biological mechanism of tumorigenesis. For each lncRNA–miRNA interaction in the hub-related subnetwork and two hierarchical networks, we performed RNAup method to evaluate their binding energy. Our results identified two important lncRNAs with prognostic roles in colon cancer and dissected their regulatory mechanism involving synergistic interaction with miRNAs.

Keywords: non-coding RNA, synergistic interaction, colon cancer, biological mechanism, prognostic biomarker

OPEN ACCESS

Edited by:
Xiaofeng Dai,
Jiangnan University, China

Reviewed by:
Padhmanand Sudhakar,
Earlham Institute (EI), United Kingdom
Nathan Weinstein,
Universidad Nacional Autónoma de México, Mexico

****Correspondence:***
Haitao Yu
15694549898@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *15 June 2020*
Accepted: *24 August 2020*
Published: *15 September 2020*

Citation:
Zhao B, Qu X, Lv X, Wang Q, Bian D, Yang F, Zhao X, Ji Z, Ni J, Fu Y, Xin G and Yu H (2020) Construction and Characterization of a Synergistic lncRNA–miRNA Network Reveals a Crucial and Prognostic Role of lncRNAs in Colon Cancer. Front. Genet. 11:572983. doi: 10.3389/fgene.2020.572983

INTRODUCTION

Colon adenocarcinoma (COAD) has emerged as one of the leading causes of cancer-related deaths worldwide, with an increasing prevalence (Dienstmann et al., 2015). COAD is a complex disease, whose initiation and progression is closely related with mechanisms of regulation of gene expression (Liu et al., 2017). In recent years, with the development of next-generation sequencing technologies, studies suggested that less than 2% of the human genome encoded protein-coding genes, whereas non-coding RNAs represented most of the human transcriptome (Tay et al., 2014; Levy and Myers, 2016). Non-coding transcripts are divided into various classes, among which long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) have attracted increasing attention. Notably, lncRNAs have been implicated in a diverse range of biological processes, including proliferation, migration, or genomic stability (Mercer et al., 2009; Fatica and Bozzoni, 2014). For instance, some lncRNAs have been reported to regulate gene expression through binding to PRC2 and acting as important controllers of cellular functions (Lee, 2012). miRNAs are small non-coding RNAs that also play a key role in gene regulation (Xu et al., 2019). miRNAs regulate gene expression mainly by binding to the 3′-untranslated regions of mRNAs and leading to their degradation or inhibition, and various studies have demonstrated that aberrant gene expression is closely linked to tumorigenesis, metastasis, and specific tumor stages (Liu et al., 2009).

Previous studies have demonstrated that lncRNAs interact with miRNAs to act on biological traits (Zhang et al., 2019): for example, Kallen et al. (2013) found that the lncRNA H19 modulated miRNA let-7 by performing *in vivo* crosslinking combined with affinity purification experiments. In summary, lncRNAs and miRNAs could interact by regulating mRNAs, thus playing critical roles in the pathological processes involved in disease development (Liao et al., 2019). However, the biological roles and functions of lncRNA–miRNA synergistic interactions have not yet been described in COAD and should be investigated to improve the efficiency of early diagnosis and treatment in the tumorigenesis and progression of this disease.

In this study, we constructed and characterized the lncRNA–miRNA synergistic network involved in COAD by integrating the lncRNA/miRNA–mRNA regulation pairs and the expression profiles of these RNA molecules. In total, we identified 305 positive and 294 negative synergistic lncRNA–miRNA interactions with significantly shared mRNAs. We observed that some of the synergistic lncRNAs and miRNAs were significantly enriched with cancer RNAs, and COAD-related lncRNAs were more important than COAD-related miRNAs. Finally, we proposed a risk score model to predict the clinical outcome of COAD patients based on two prognostic hub lncRNAs, MEG3 and ZEB1-AS1, which were identified by univariate Cox regression analysis. The biological mechanism involving these two lncRNAs was further analyzed. For synergistic lncRNA–miRNA interactions in the hub-related subnetwork and two prognostic lncRNAs related interactions, we provided the total free energy of binding evaluated by RNAup method. Altogether, our analysis provides new insight for exploring the molecular mechanisms of lncRNA–miRNA synergistic interactions and uncovering candidate lncRNA biomarkers for the prognosis of COAD.

MATERIALS AND METHODS

The lncRNA/miRNA–mRNA Regulation Pairs

The miRNA–mRNA target data were downloaded and filtered from StarBase (Li et al., 2014). We chose the miRNA–mRNA interactions which were predicted by at least three of seven target-predicting programs. These seven target-predicting programs included PITA, RNA22, miRmap, DIANA-microT, miRanda, PicTar, and TargetScan. Recent advances in high-throughput sequencing of immunoprecipitated RNAs after cross-linking (CLIP-Seq, HITS-CLIP, PAR-CLIP, CLASH, iCLIP) provide powerful ways to identify biologically relevant RNA-target interactions. To obtain the high-quality miRNA–mRNA datasets, we further selected the miRNA–mRNA interactions which were validated by at least three CLIP-seq data from above interactions as the final miRNA–mRNA interactions. Similarly, we obtained the lncRNA–mRNA interactions that have at least two interactions and supported by at least three CLIP-seq data (**Supplementary Table S1**). Moreover, we downloaded the experimentally validated lncRNA–mRNA interactions from LncReg and LncRNA2Target databases (**Supplementary Table S1**; Zhou et al., 2015; Cheng et al., 2019). Integrating the lncRNA–mRNA interactions downloaded from these three databases, we obtained the final lncRNA–mRNA interactions. This way, we obtained 1,336 lncRNA–mRNA and 202,712 miRNA–mRNA high-quality non-redundant interactions.

Expression Profiles and Clinical Data of COAD Samples

The lncRNA, miRNA, and mRNA expression data for COAD patients were downloaded from the Cancer Genome Atlas (TCGA) database (Tomczak et al., 2015). For each expression profile, RNAs with missing values in more than 30% samples were removed and each of the remaining missing value was imputed by the KNN Imputation. Then all expression values were log2 transformed to obtain the final expression profiles. We chose sample-matched miRNA and lncRNA/mRNA expression profiles for further analysis. The clinical data of COAD patients was also obtained from TCGA.

CpG sites with missing values in more than 30% samples were removed and each of the remaining missing value was imputed by the KNN Imputation.

Collection of COAD-Related lncRNAs and miRNAs

The COAD-related lncRNAs were downloaded from LncRNADisease (Bao et al., 2019) and lnc2Cancer (Gao et al., 2019). Similarly, we collected the COAD-related miRNAs from several databases, including miR2Disease (Jiang et al., 2009),

HMDD (Huang et al., 2019), SM2miR (Liu et al., 2013), and OncomiRDB (Wang et al., 2014).

Construction of the lncRNA–miRNA Synergistic Interaction Network

To identify the synergistic lncRNA–miRNA interactions, we developed a four-step computational method by integrating lncRNA–mRNA interactions, miRNA–mRNA interactions, and expression profiles of lncRNA, miRNA, and mRNA in COAD samples (**Supplementary Figure S1**).

First, high-quality lncRNA–mRNA and miRNA–mRNA interactions were downloaded from several databases and processed to obtain the non-redundant data as described above. Second, the regulatory networks of lncRNA–mRNA and miRNA–mRNA were constructed by filtering the lncRNA/miRNA–mRNA pairs obtained from StarBase with their expression profiles. The regulatory correlation between lncRNA/miRNA and mRNA was evaluated by Pearson correlation coefficient based on the matched lncRNA/miRNA and mRNA expression profiles. The pairs with significant correlation were saved as the lncRNA–mRNA (p-adjusted < 0.05) and miRNA–mRNA ($R < -0.4$, p-adjusted < 0.05) interactions in COAD patients. Third, we identified the co-regulated lncRNA–miRNA pairs by evaluating the significance of their shared regulated mRNAs (hypergeometric-test, p-adjusted < 0.01). Fourth, we used Pearson correlation to evaluate the synergistic direction and synergistic power of each co-regulated lncRNA–mRNA pair, and the pairs with $p < 0.05$ were considered the final synergistic lncRNA–mRNA pairs. Finally, after assembling the synergistic lncRNA–mRNA pairs, we obtained the lncRNA–miRNA synergistic interaction network. Two types of nodes were involved in the network (lncRNAs and miRNAs), with positive or negative synergistic regulations.

Survival Analysis According to the Risk Score Model

We evaluated the clinical outcomes for COAD patients by our risk score model based on the expression levels of MEG3 and ZEB1-AS1. We first divided the COAD samples into training (70% of the samples) and test (30% of the samples) sets. The risk score model was constructed by considering the individual power of MEG3 and ZEB1-AS1 evaluated by the univariable Cox regression analysis and their expression levels in training samples as follows:

$$\text{Risk score} = \sum_{i=1}^{2} coef_i \times exp_i$$

where $\exp_i$ is the expression level of MEG3 or ZEB1-AS1 and coef_i is the regression coefficient of MEG3 or ZEB1-AS1 estimated by univariate Cox regression analysis. As a result, the risk score for each COAD patient was computed by the formula:

$$\begin{aligned}\text{Risk score} = {} & (0.4933 \times \text{expression value of MEG3}) \\ & + (1.1077 \times \text{expression value of ZEB1} - \text{AS1}) \quad (1)\end{aligned}$$

The median risk score value in the training samples was chosen as the cut-off value to classify patients into high-risk and low-risk groups from the training and test sets, respectively. Survival analyses were performed to assess the difference in clinical outcome between the high-risk and low-risk groups, and statistical significance was evaluated by a log-rank test using the R package "survival."

Network Visualization

The networks were visualized by Cytoscape 3.3.0 (Shannon et al., 2003), including the synergistic lncRNA–miRNA network, the hub-related subnetwork and the MEG3/ZEB1-AS1-related hierarchical networks.

RESULTS

Construction and Characterization of the Synergistic lncRNA–miRNA Network

Based on the lncRNA/miRNA–mRNA regulation pairs downloaded from databases and expression profiles of lncRNA, miRNA, and mRNA, we constructed a synergistic lncRNA–miRNA network in the context of COAD (**Figure 1A**). As described in the 'Materials and Methods' section, we constructed this network in four steps. In the first step, we obtained 455 lncRNA–mRNA and 28,639 miRNA–mRNA COAD-specific regulation pairs (**Table 1**). Then, we identified 1,368 co-regulated lncRNA–miRNA pairs with significantly shared mRNAs by a hypergeometric test (**Table 1**). Finally, Pearson correlation analysis was used to filter the co-regulated lncRNA–miRNA pairs, and we obtained 305 positive and 294 negative synergistic lncRNA–miRNA interactions, covering 88 lncRNA and 161 miRNAs (**Table 1**, **Figure 1B** and **Supplementary Table S2**).

To explore the architecture and features of the synergistic network, its topological properties were analyzed. Through the analysis of node degree distribution, we found that the majority of nodes had few synergistic interactions, while a small portion had many interactions. This fits with the power-law distribution, suggesting that the synergistic lncRNA–miRNA network was scale-free and different from randomly generated networks ($R^2 = 0.82$, **Figure 1C**). Moreover, we compared the clustering coefficient between our synergistic network and random networks. The result showed that the lncRNAs and miRNAs in our network had tight synergistic interactions ($p < 0.001$, **Figure 1D**). The synergistic lncRNA–miRNA pairs share varying numbers of mRNAs. In order to explore the relationship between their synergistic intensity and the number of shared mRNAs, we compared the co-expression significance with different numbers of shared mRNAs. The result showed that those lncRNA–miRNA pairs with more shared regulated mRNAs tended to be more significantly co-expressed to achieve coordinated regulation ($R = 0.97$, $p = 0.026$, **Figure 1E**).

lncRNAs Are More Likely to Have Similar Synergistic Interactions With Its Neighbors

The lncRNAs and miRNAs in our network had positive and negative synergistic interactions. Therefore, we counted and computed the ratio of positive and negative miRNA neighbors

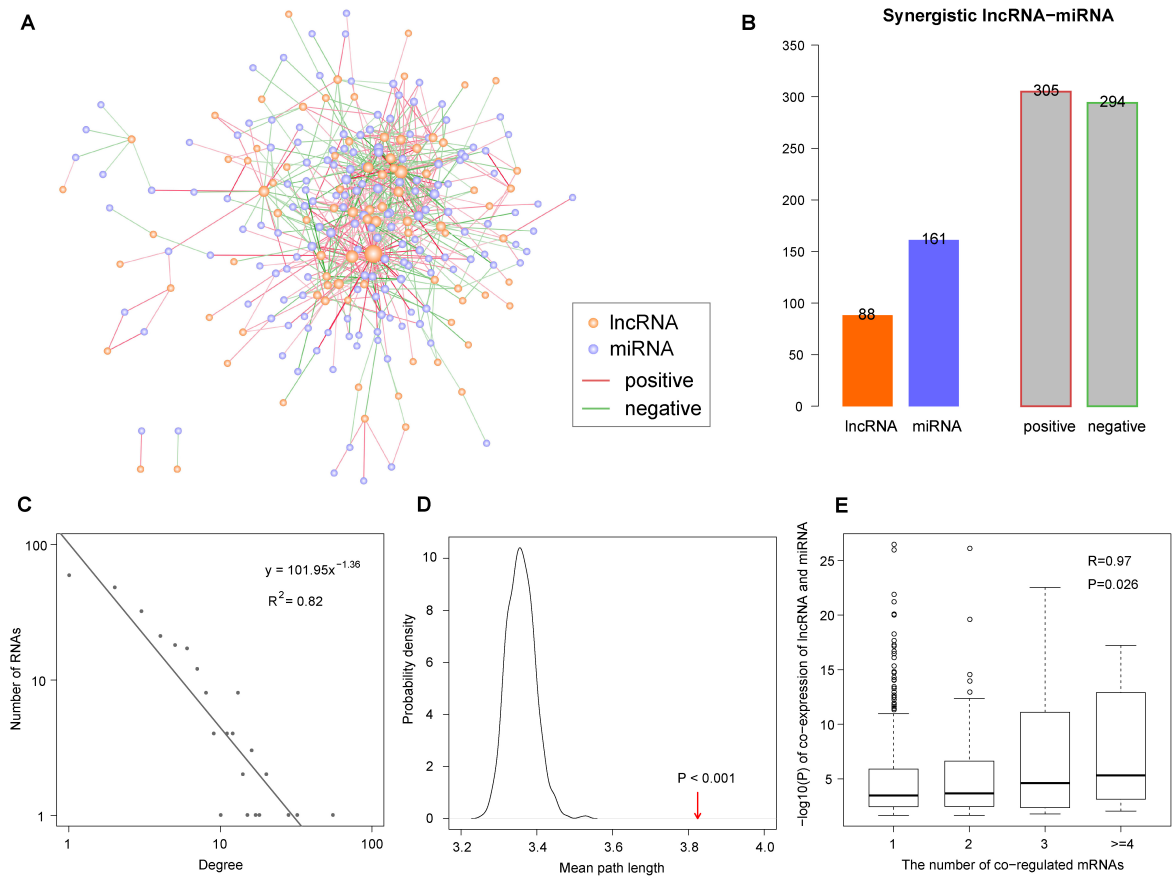

FIGURE 1 | The synergistic long non-coding RNAs (lncRNA)-microRNAs (miRNA) network in the context of colon adenocarcinoma (COAD). **(A)** Global view of the synergistic network. Orange and purple nodes represent lncRNAs and miRNAs, respectively. Red and green edges represent positive and negative synergistic relationships. Larger nodes represent higher degrees. **(B)** Statistics of nodes and edges in the synergistic network. **(C)** Degree distribution of the synergistic network. **(D)** The mean path length of the synergistic network is higher than randomization test. The arrow represents the mean path length in the real network. **(E)** The synergistic lncRNA–miRNA pairs with more co-regulated mRNAs show more significantly synergistic regulatory relationship. The white circles represents the outlier points in boxplots.

for each lncRNA. The results indicated that each lncRNA had 1∼55 lncRNA neighbors, including 0∼50 positive and 0∼26 negative neighbors (**Figure 2A**). After calculating the neighbor ratio of each lncRNA, we found that 82.95% of the lncRNAs tended to have the same synergistic direction as most (≥ 80%) of its miRNA neighbors. Among these lncRNAs, 48 and 52% were likely to have positive and negative synergistic interactions with miRNAs, respectively (**Figure 2B**). Examples of such lncRNAs are MALAT1, DANCR, and AGAP2-AS1 (**Figure 2C**).

Similarly, we counted and computed the neighbors of each miRNA, and discovered that each miRNA had 1∼14 lncRNA neighbors with 0∼10 positive and 0∼7 negative lncRNA neighbors (**Figure 2D**). As opposed to lncRNAs, only 43.48% of the miRNAs had the same synergistic direction as most (≥ 80%) of its lncRNA neighbors, while other miRNAs such as hsa-miR-106b-5p and hsa-miR-20a-5p had mixed synergistic relationships with their lncRNA neighbors (**Figures 2E,F**).

TABLE 1 | The statistic of COAD specific regulation pairs.

COAD specific pairs	Number of node 1	Number of node 2	Number of edges
lncRNA–mRNA	169 (lncRNA)	313 (mRNA)	455
miRNA–mRNA	289 (miRNA)	6,392 (mRNA)	28,639
Co-regulated lncRNA–miRNA	113 (lncRNA)	194 (miRNA)	1,368
Synergistic lncRNA–miRNA	88 (lncRNA)	161 (miRNA)	305 (positive) 294 (negative)

long non-coding RNAs (lncRNAs); microRNAs (miRNAs); colon adenocarcinoma (COAD).

The Cancer-Related lncRNAs in the Synergistic Network Have Centralized Roles

Many cancer-related genes have been discovered, and to explore the relationship between our synergistic network and COAD, we obtained COAD-related lncRNAs and miRNAs from databases as described in the 'Materials and Methods' section. In total, we

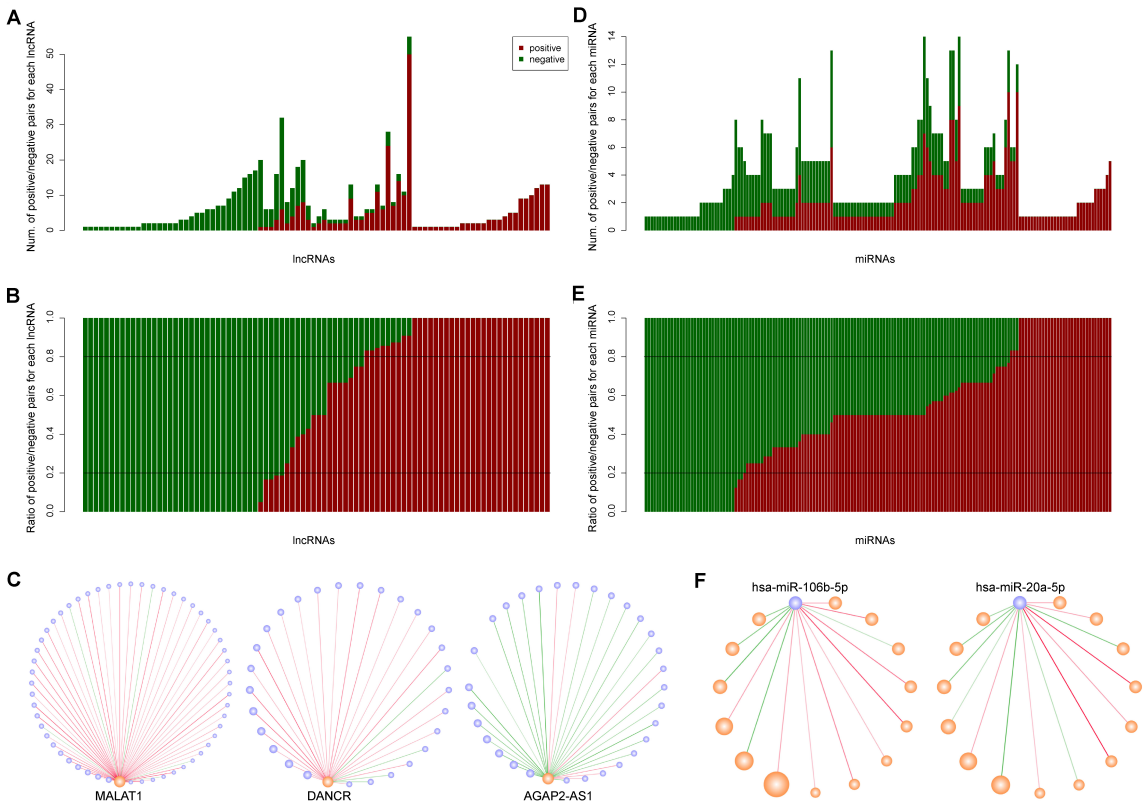

FIGURE 2 | The synergistic interactions between long non-coding RNAs (lncRNAs) and microRNAs (miRNAs). **(A)** The number of positive and negative miRNA neighbors for each lncRNA. Dark red and dark green represent positive and negative interactions, respectively. **(B)** The ratio of positive and negative miRNA neighbors for each lncRNA. **(C)** The synergistic interactions for three lncRNAs: MALAT1, DANCR, and AGAP2-AS1. **(D)** The number of positive and negative lncRNA neighbors for each miRNA. **(E)** The ratio of positive and negative lncRNA neighbors for each miRNA. **(F)** The synergistic interactions for two miRNAs: hsa-miR-106b-5p and hsa-miR-20a-5p.

obtained 116 COAD-related lncRNAs and 134 miRNAs. After mapping these known COAD-related RNAs to our synergistic network (**Figure 3A**), we found that the synergistic lncRNAs and miRNAs were significantly enriched with the cancer-related RNAs ($p < 2.2 \times 10^{-16}$ for lncRNAs and $p = 6.36 \times 10^{-10}$ for miRNAs, **Figure 3B**). Then, we compared the topological properties of cancer-related RNAs with other RNAs in the synergistic network. The results showed that COAD-related genes had significantly higher degrees, betweenness centrality, and closeness centrality than other nodes in the synergistic network ($p = 7.27 \times 10^{-10}$, 4.46×10^{-06}, and 3.77×10^{-05} for COAD-related lncRNAs; $p = 1.67 \times 10^{-03}$, 6.28×10^{-03}, and 3.65×10^{-04} for COAD-related miRNAs; **Figures 3C–E**). These results suggest that cancer-related lncRNAs and miRNAs have more centralized roles when compared to other RNAs. Moreover, COAD-related lncRNAs appear to be more important than COAD-related miRNAs.

Identification of Two Central and Prognostic lncRNAs

Numerous studies have reported hub genes as playing key roles in cancer (Chen et al., 2017; Das et al., 2017; Yin et al., 2019). To identify the hub lncRNAs and miRNAs in our synergistic network, we computed a degree for each node and sorted them in a descending order. Then, we chose 10% of the nodes with the highest degrees as the hub RNAs, including 18 lncRNAs and seven miRNAs (**Figure 4A**). When comparing the ratio of hub lncRNAs among all lncRNAs with the ratio of hub miRNAs among all miRNAs, we found that a greater ratio of lncRNAs were identified as the hub RNAs (20% *vs.* 4%, **Figure 4B**), suggesting important roles of lncRNAs in the synergistic network and in accordance with our previous results. Moreover, we found that the majority of the hub RNAs were known COAD-related RNAs, and this ratio was higher than that in non-hub RNAs (80% *vs.* 56%, **Figure 4C**). The COAD-related hub RNAs included 18 lncRNAs and seven miRNAs. Next, we extracted the edges that connected two hub RNAs and found that all hub RNAs were connected, except for one lncRNA: UCA1. The hub subnetwork is depicted in **Figure 4D**. In accordance with the synergistic network described above, we found that the lncRNAs and miRNAs in the hub subnetwork were likely to interact with their neighbors in similar and different directions, respectively (**Figure 4D**).

Considering that hub RNAs have key roles in the development of cancer, in our next step we evaluated the prognostic ability of each hub RNA. Through univariate Cox regression analysis, we identified two risk lncRNAs in COAD patients, MEG3 and

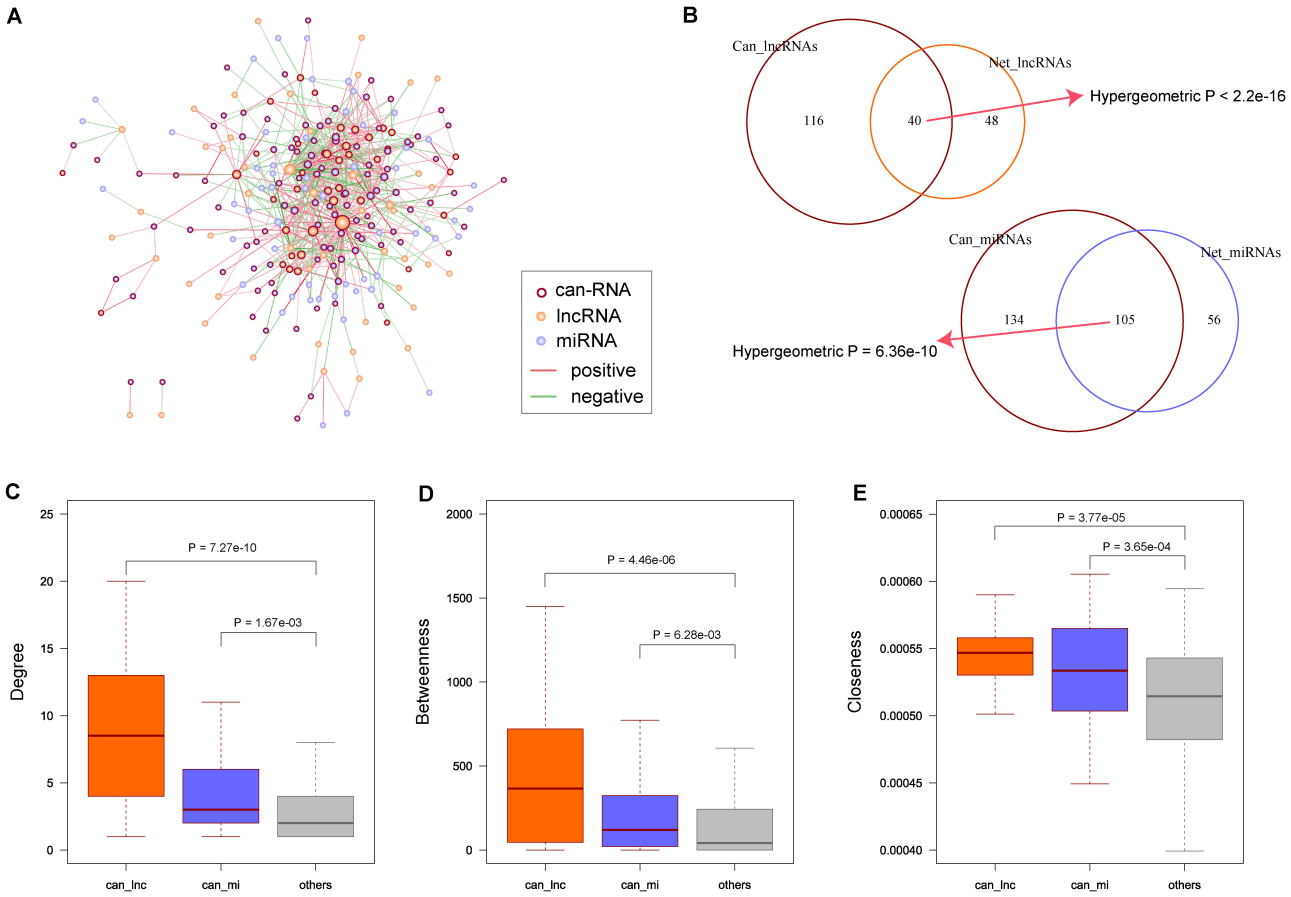

FIGURE 3 | Known colon adenocarcinoma (COAD) related RNAs in synergistic network. **(A)** Known COAD related RNAs were mapping to the synergistic network. Node color and edge color are same as **Figure 1A**. Nodes with dark red border represent the known COAD related RNAs. **(B)** Enrichment of known COAD related long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) in network. *P*-values are computed by hypergeometric test. **(C)** COAD related lncRNAs and miRNAs have higher degrees than other nodes. **(D)** COAD related lncRNAs and miRNAs have higher betweenness than other nodes. **(E)** COAD related lncRNAs and miRNAs have higher closeness than other nodes. *P*-values are calculated based on Wilcoxon rank sum test.

ZEB1-AS1 (HR = 1.52 and $p = 0.03$ for MEG3, HR = 3.19 and $p = 2.13 \times 10^{-05}$ for ZEB1-AS1, **Figures 4E,F**). Furthermore, we combined these two lncRNAs using the risk score model to predict the clinical outcome of COAD patients. The risk score model was constructed according to Equation 1, with 0.4933 as the regression coefficient for MEG3 estimated by the univariate Cox regression analysis, and 1.1077 as the regression coefficient for ZEB1-AS1. The median risk score of training samples was used as the cut-off value (1.22) to separate the high-risk and low-risk groups. Survival analysis revealed a significant difference in overall survival between these two groups (log-rank test $p = 0.00255$, **Figure 5A**). Furthermore, we computed the risk score of samples in the test set based on the risk score model and divided these samples into high-risk and low-risk groups. Comparing the clinical outcome of samples between the two groups, we found that low-risk samples in the test set also showed significantly better prognosis (log-rank test $p = 0.021$, **Figure 5B**). In addition, we randomly selected different sample sets (70∼90% of all COAD samples and the 70∼90% of the test samples), computed their risk scores, and compared the survival difference between high-risk and low-risk groups. The results showed that the risk score model could predict the clinical outcome of COAD patients (**Supplementary Figure S2**). This result illustrated the robust prognostic ability of the MEG3 and ZEB1-AS1 lncRNAs.

To explore the roles of MEG3 and ZEB1-AS1 during cancer progress, we compared the expression values of MEG3 and ZEB1-AS1 in cancer samples and normal samples. As a result, we found that MEG3 presented similar expression levels in both contexts ($p = 0.66$), while ZEB1-AS1 showed a significantly higher expression in cancer samples ($p = 3.983 \times 10^{-09}$, **Figure 5C**). In addition, we compared the expression levels of these two lncRNAs in high-risk and low-risk samples: results showed that MEG3 and ZEB1-AS1 had significantly higher expression values in high-risk samples (all $p < 2.2 \times 10^{-16}$, **Figure 5D**). We also downloaded two GEO lncRNA expression datasets associated with colon diseases, including GSE77013 and GSE67106 (Mirza et al., 2015; Padua et al., 2016). The expression of MEG3 and ZEB1-AS1 were found to be highly expressed in disease samples compared with control samples (*t*-test, $P = 0.09$, 0.016, and 1.23e-08, **Supplementary Figure S3**). Due to the lack of

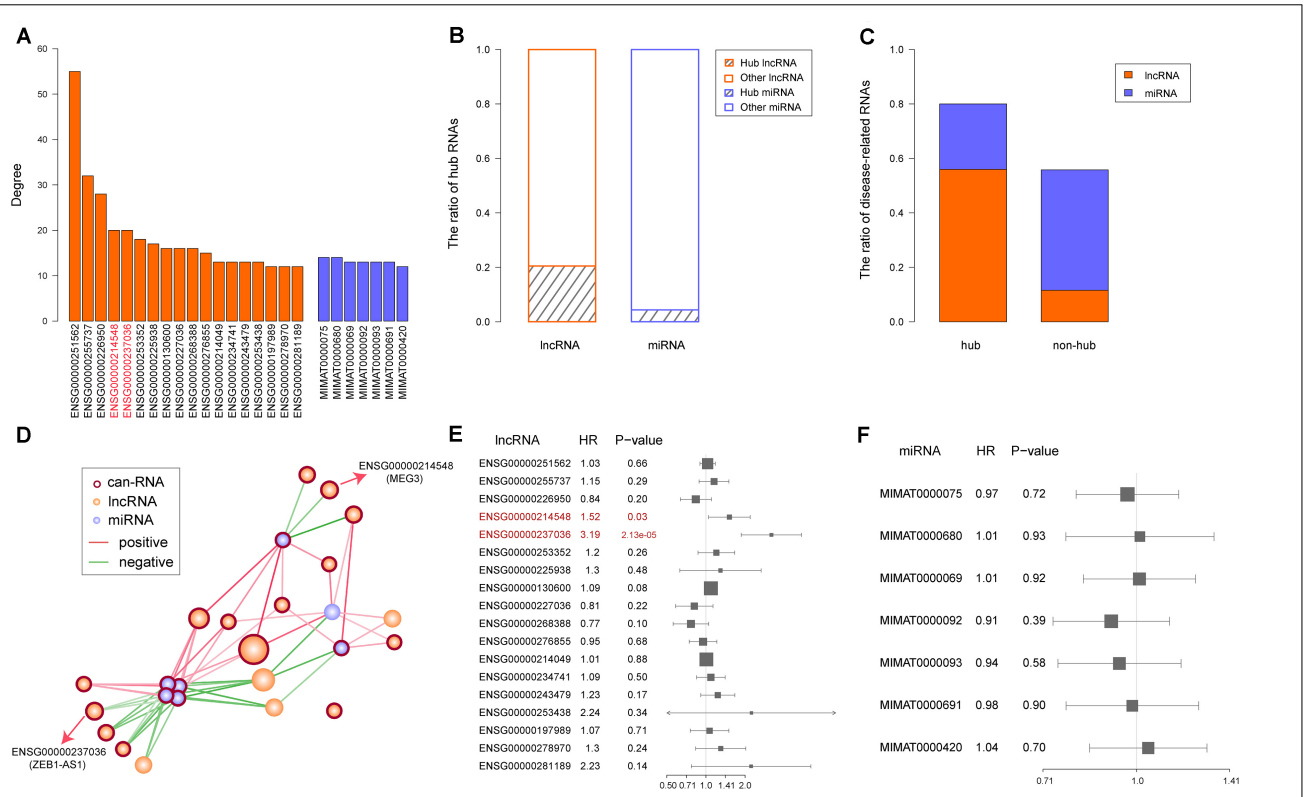

FIGURE 4 | Identification and characterization of hub RNAs. **(A)** Degree of hub RNAs. Orange bars and purple bars represent degree of long non-coding RNAs (lncRNAs) and microRNAs (miRNAs), respectively. **(B)** The ratio of hub and non-hub RNAs in lncRNAs and miRNAs, respectively. **(C)** The ratio of known colon adenocarcinoma (COAD) related lncRNAs and mRNAs in hub and non-hub RNAs. **(D)** Hub subnetwork extracted from the synergistic network. Color of node and edge are same as **Figure 3A**. **(E)** The result of univariable Cox regression for each hub lncRNA. The prognostic lncRNAs are marked in red. **(F)** The result of univariable Cox regression for each hub miRNA.

ZEB1-AS1 probes in GSE77013, we didn't compare its expression values between disease and control samples. These results demonstrate that the high expression of these two lncRNAs is associated with colon disease. Summarizing these results, we believe that MEG3 played a role in cancer development while ZEB1-AS1 could act in both carcinogenesis and cancer development. These results were consistent with previous studies (Fu and Cui, 2017; Gong et al., 2017; Liu et al., 2019; Hu et al., 2020; Ni et al., 2020).

Hierarchical Networks for Elucidating the Biological Mechanism of MEG3 and ZEB1-AS1

To contribute to the understanding of the synergistic interactions of MEG3 and ZEB1-AS1, we proposed hierarchical models to systematically illustrate the regulatory process. As shown in **Figure 6**, MEG3 or ZEB1-AS1 regulate mRNAs by synergistic interactions with miRNAs, and further participate in cancer biological processes. Among the miRNAs that have synergistic interactions with MEG3, 14 out of 20 miRNAs were known COAD-related miRNAs. These miRNAs, along with MEG3, further regulate 11 mRNAs, including CASP3, CASP8, and vascular endothelial growth factor (VEGFA). Previous studies have indicated that polymorphisms in CASP3 and CASP8 are related to colon cancer (Goodman et al., 2006; Choi et al., 2012), and the VEGFA was also significantly associated with rectal cancer (Slattery et al., 2014). In this study, we uncovered their upstream regulatory factors, which were MEG3 and its synergistic miRNAs. This result might be another indication of the roles of these mRNAs in carcinogenesis. Through the integration of annotation information on mRNAs, we found that MEG3 and its synergistic miRNAs were mainly associated with cancer-related processes such as immune system development, cell development, tissue development, cell differentiation, protein metabolism, and other processes.

Regarding the synergistic miRNAs related to ZEB1-AS1, 15 out of 20 were found to be known COAD-related miRNAs. ZEB1-AS1 and these miRNAs synergistically regulated six genes, such as CyclinD1 (CCND1) and zinc finger E-box binding homeobox 1 (ZEB1). CCND1 is a key cell cycle regulatory protein and its polymorphism has been found to be significantly associated with overall COAD risk (Xie et al., 2017). Li et al. (2012) reported that IL-1β may promote colon tumor growth and invasion through the activation of colon cancer stem cell self-renewal and epithelial-mesenchymal transition (EMT), and ZEB1 plays a critical role in these two processes. Moreover, we observed that the mRNAs regulated by ZEB1-AS1

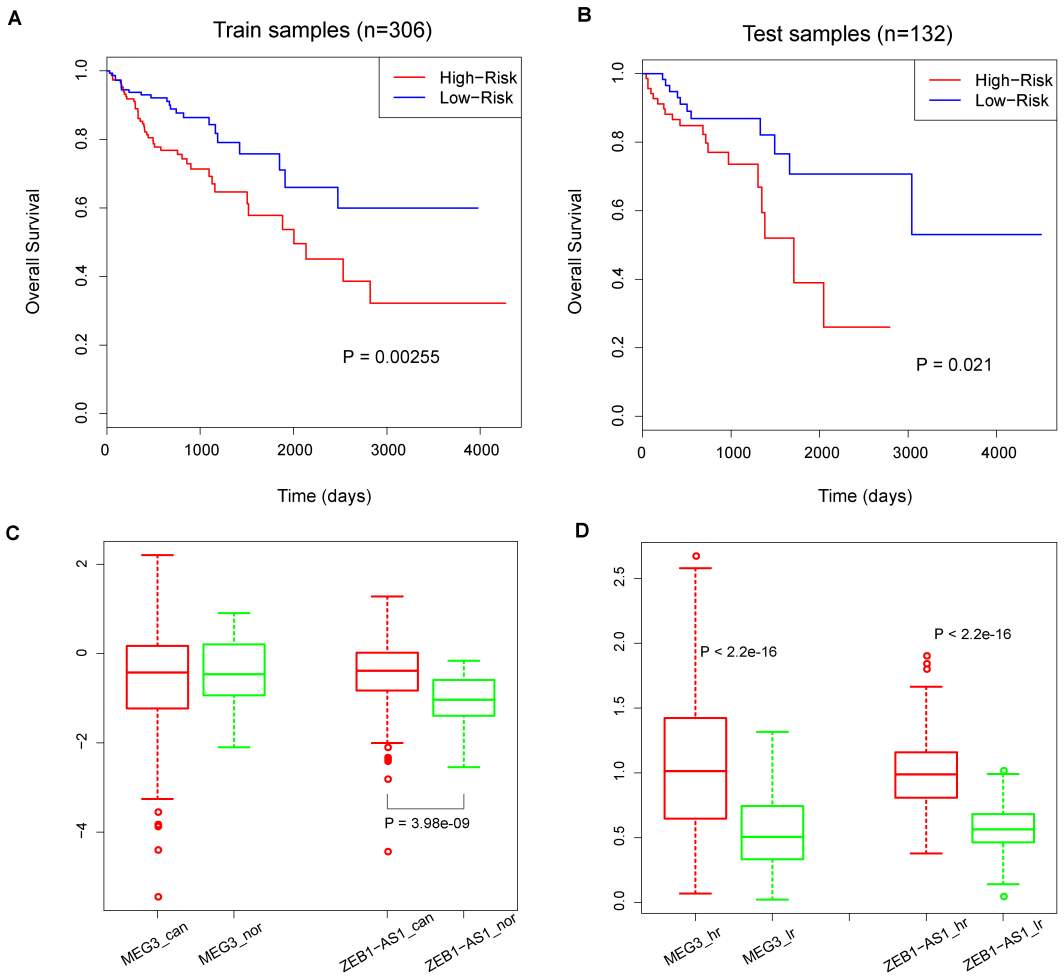

FIGURE 5 | (A) Kaplan–Meier curves of overall survival in high-risk and low-risk groups in training set. **(B)** Kaplan–Meier curves of overall survival in high-risk and low-risk groups in test set. *P*-values were calculated by the log-rank test. **(C)** The colored boxes and circles represents expression values of MEG3 and ZEB1-AS1 in cancer and normal samples. 'can' represent 'cancer' and 'nor' represent 'normal'. **(D)** The colored boxes and circles represents expression values of MEG3 and ZEB1-AS1 in high-risk and low-risk samples. 'hr' represent 'high-risk' and 'lr' represent 'low-risk'. *P*-values are calculated by Wilcoxon rank sum test.

and its synergistic miRNAs were annotated to cancer-related gene ontology (GO) terms such as cell differentiation, cell proliferation, and developmental processes.

In light of these results, we believe that MEG3 and ZEB1-AS1 play important roles in the initiation and progression of colon cancer, through their synergistic interactions with cancer-related miRNAs and finally regulating cancer-related mRNAs that were associated with cancer biological processes. Our results could contribute to the understanding of important roles of synergistic lncRNA–miRNA interactions in tumorigenesis, expand the complexity of the ncRNA–mRNA regulatory network, and provide potential therapeutic targets for colon cancer treatment.

DISCUSSION

Colon adenocarcinoma is the third most common cancer worldwide and has become a widespread health issue for its highly mortality and morbidity (Bertotti et al., 2015; Marmol et al., 2017). Recent studies suggested that interactions between lncRNAs and miRNAs in the regulation of mRNA expression played important regulatory roles in the initiation and progression of COAD (Yu et al., 2017). However, the regulatory mechanisms through which lncRNA–miRNA interactions are involved in the progression of this disease are stil unclear.

Long non-coding RNAs–miRNA synergistic interactions are critical for many biological functions and exploring these interactions contributes to a further understanding of the process of tumorigenesis and development of COAD (Guil and Esteller, 2015; Wang et al., 2019a). More importantly, increasing evidence shows that the lncRNA–miRNA interaction network is implicated in several pathophysiological processes, including human cancers (Lin et al., 2019). In this work, we constructed and characterized the lncRNA–miRNA synergistic network by integrating lncRNA–mRNA interactions, miRNA–mRNA

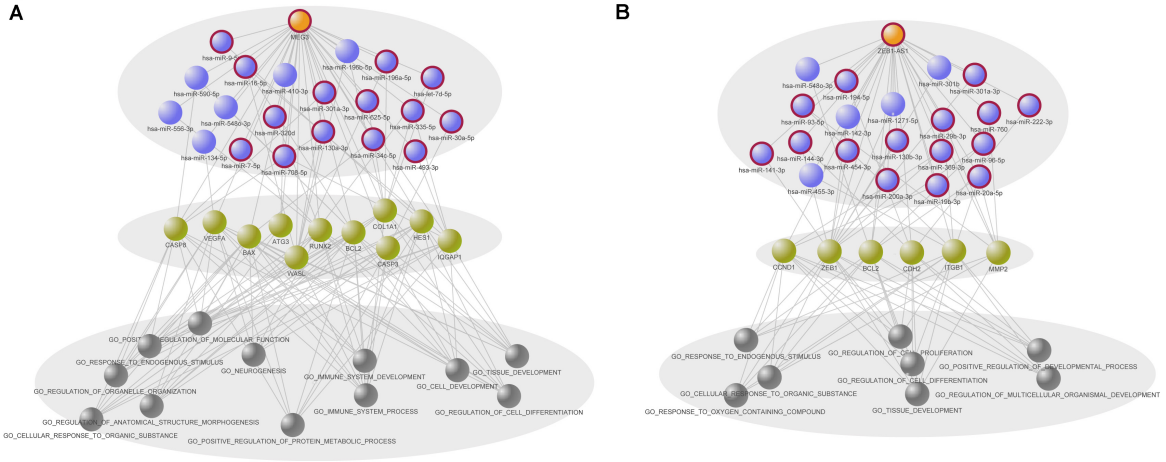

FIGURE 6 | Two hierarchical models of prognostic long non-coding RNAs (lncRNAs). **(A)** Hierarchical model of MEG3 related synergistic regulatory interactions. **(B)** Hierarchical model of ZEB1-AS1 related synergistic regulatory interactions. The models are exhibited hierarchically by three levels, including synergistic lncRNA–microRNAs (miRNA) interactions, regulatory mRNAs and annotated gene ontology (GO) biological processes of mRNAs. GO terms annotated to at least seven mRNAs and five mRNAs are showed in **(A,B)**, respectively. Orange, purple and green nodes represent lncRNAs, miRNAs and mRNAs, respectively. Gray nodes represent the GO terms.

interactions, and the expression profiles of lncRNA, miRNA, and mRNA in COAD samples. The analysis of this synergistic network allowed the detection of complicated features and functions of RNA regulatory interactions and how lncRNAs and miRNAs could play regulatory roles in the tumorigenesis and progression of COAD (Wang et al., 2020). Our results indicated that the synergistic lncRNAs and miRNAs were significantly enriched with cancer-related RNAs. In addition, COAD-related lncRNAs and miRNAs had significantly higher degrees, betweenness centrality, and closeness centrality than other nodes in the synergistic network. Further analysis showed that cancer-related miRNAs, especially lncRNAs, had more centralized roles when compared with other RNAs. Altogether, our study of lncRNA–miRNA interactions could contribute with crucial information in the understanding of the regulatory mechanisms through which ncRNAs act, as well as with the identification of molecular targets for the diagnosis and treatment of COAD.

Reliable prediction of RNA–RNA binding energies is crucial. RNAup is an effective method, which involved two energy contributions, including (1) the energy necessary to 'open' the binding site and (2) the energy gained from hybridization. To improve the medical effectiveness of our results, we performed RNAup to compute the potential binding possibility between lncRNAs and miRNAs. The sequence of lncRNA transcripts and miRNA were downloaded from Ensemble and miRBase database, respectively (**Supplementary Table S3**; Kozomara et al., 2019; Yates et al., 2020). For interactions in the hub-related subnetwork and two hierarchical networks, we provided the total free energy of binding for each lncRNA–miRNA interaction (**Supplementary Tables S4, S5**). Based on the total free energy of binding we provided, users can acquire both direct and indirect interactions by their own cutoffs. This way, we expect to provide results that have higher potential medical usefulness.

Accumulating evidence revealed that lncRNAs acted as prognostic biomarkers and regulated cell functions in colorectal cancer (Yin et al., 2015; Wang et al., 2019b). Through analyses of node degree and univariate Cox regression analysis, we identified two important lncRNAs: MEG3 and ZEB1-AS1. We further proposed two hierarchical models to systematically illustrate the regulatory process of these two lncRNAs. In the hierarchical models, most miRNAs which have synergistic interactions with MEG3 or ZEB1-AS1 were found to be known COAD-related miRNAs. Moreover, we found that some mRNAs regulated by the lncRNAs and miRNAs were reported to be associated with COAD. Our results proposed another indication of the roles of these mRNAs in carcinogenesis. We believe that other ncRNAs and mRNAs in the hierarchical models were also COAD-related RNAs. Our results provide potential therapeutic targets for colon cancer treatment. Finally, we proposed a risk score model to predict the clinical outcome of COAD patients and demonstrated the utility of lncRNAs as promising biomarkers.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

HY conceived and designed the experiments. BZ, XQ, XL, QW, DB, FY, XZ, ZJ, JN, GX, and YF collected and analyzed data. HY and BZ wrote this manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by scientific research project of Heilongjiang Health and Family Planning Commission [2014-242].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.572983/full#supplementary-material

REFERENCES

Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., and Dong, D. (2019). LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.* 47, D1034–D1037.

Bertotti, A., Papp, E., Jones, S., Adleff, V., Anagnostou, V., and Lupo, B. (2015). The genomic landscape of response to EGFR blockade in colorectal cancer. *Nature* 526, 263–267.

Chen, P., Wang, F., Feng, J., Zhou, R., Chang, Y., Liu, J., et al. (2017). Co-expression network analysis identified six hub genes in association with metastasis risk and prognosis in hepatocellular carcinoma. *Oncotarget* 8, 48948–48958. doi: 10.18632/oncotarget.16896

Cheng, L., Wang, P., Tian, R., Wang, S., Guo, Q., Luo, M., et al. (2019). LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.* 47, D140–D144.

Choi, J. Y., Kim, J. G., Lee, Y. J., Chae, Y. S., Sohn, S. K., Moon, J. H., et al. (2012). Prognostic impact of polymorphisms in the CASPASE genes on survival of patients with colorectal cancer. *Cancer Res. Treat.* 44, 32–36. doi: 10.4143/crt.2012.44.1.32

Das, S., Meher, P. K., Rai, A., Bhar, L. M., and Mandal, B. N. (2017). Statistical approaches for gene selection, hub gene identification and module interaction in gene co-expression network analysis: an application to aluminum stress in soybean (*Glycine max* L.). *PLoS One* 12:e0169605. doi: 10.1371/journal.pone.0169605 doi: 10.1371/journal.pone.0169605

Dienstmann, R., Salazar, R., and Tabernero, J. (2015). Personalizing colon cancer adjuvant therapy: selecting optimal treatments for individual patients. *J. Clin. Oncol.* 33, 1787–1796. doi: 10.1200/jco.2014.60.0213

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* 15, 7–21. doi: 10.1038/nrg3606

Fu, J., and Cui, Y. (2017). Long noncoding RNA ZEB1-AS1 expression predicts progression and poor prognosis of colorectal cancer. *Int. J. Biol. Markers* 32, e428–e433.

Gao, Y., Wang, P., Wang, Y., Ma, X., Zhi, H., Zhou, D., et al. (2019). Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Res.* 47, D1028–D1033.

Gong, H., Wen, H., Zhu, X., Lian, Y., Yang, X., Qian, Z., et al. (2017). High expression of long non-coding RNA ZEB1-AS1 promotes colorectal cancer cell proliferation partially by suppressing p15 expression. *Tumour Biol.* 39:1010428317705336.

Goodman, J. E., Mechanic, L. E., Luke, B. T., Ambs, S., Chanock, S., and Harris, C. C. (2006). Exploring SNP-SNP interactions and colon cancer risk using polymorphism interaction analysis. *Int. J. Cancer* 118, 1790–1797. doi: 10.1002/ijc.21523

Guil, S., and Esteller, M. (2015). RNA-RNA interactions in gene regulation: the coding and noncoding players. *Trends Biochem. Sci.* 40, 248–256. doi: 10.1016/j.tibs.2015.03.001

Hu, D., Zhang, B., Yu, M., Shi, W., and Zhang, L. (2020). Identification of prognostic biomarkers and drug target prediction for colon cancer according to a competitive endogenous RNA network. *Mol. Med. Rep.* 22, 620–632. doi: 10.3892/mmr.2020.11171

Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., et al. (2019). HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.* 47, D1013–D1017.

Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., et al. (2009). miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* 37, D98–D104.

Kallen, A. N., Zhou, X. B., Xu, J., Qiao, C., Ma, J., Yan, L., et al. (2013). The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol. Cell.* 52, 101–112. doi: 10.1016/j.molcel.2013.08.027

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47, D155–D162.

Lee, J. T. (2012). Epigenetic regulation by long noncoding RNAs. *Science* 338, 1435–1439. doi: 10.1126/science.1231776

Levy, S. E., and Myers, R. M. (2016). Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* 17, 95–115.

Li, J. H., Liu, S., Zhou, H., Qu, L. H., and Yang, J. H. (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 42, D92–D97.

Li, Y., Wang, L., Pappan, L., Galliher-Beckley, A., and Shi, J. (2012). IL-1beta promotes stemness and invasiveness of colon cancer cells through Zeb1 activation. *Mol. Cancer* 11:87. doi: 10.1186/1476-4598-11-87

Liao, J., Wang, J., Liu, Y., Li, J., and Duan, L. (2019). Transcriptome sequencing of lncRNA, miRNA, mRNA and interaction network constructing in coronary heart disease. *BMC Med. Genomics* 12:124. doi: 10.1186/s12920-019-0570-z

Lin, C., Yuan, G., Hu, Z., Zeng, Y., Qiu, X., Yu, H., et al. (2019). Bioinformatics analysis of the interactions among lncRNA, miRNA and mRNA expression, genetic mutations and epigenetic modifications in hepatocellular carcinoma. *Mol. Med. Rep.* 19, 1356–1364.

Liu, B., Li, J., Tsykin, A., Liu, L., Gaur, A. B., and Goodall, G. J. (2009). Exploring complex miRNA-mRNA interactions with Bayesian networks by splitting-averaging strategy. *BMC Bioinformatics* 10:408. doi: 10.1186/1471-2105-1-408

Liu, H., Ye, D., Chen, A., Tan, D., Zhang, W., Jiang, W., et al. (2019). A pilot study of new promising non-coding RNA diagnostic biomarkers for early-stage colorectal cancers. *Clin. Chem. Lab. Med.* 57, 1073–1083. doi: 10.1515/cclm-2019-0052

Liu, R., Zhang, W., Liu, Z. Q., and Zhou, H. H. (2017). Associating transcriptional modules with colon cancer survival through weighted gene co-expression network analysis. *BMC Genomics* 18:361. doi: 10.1186/s12864-017-3761-z

Liu, X., Wang, S., Meng, F., Wang, J., Zhang, Y., Dai, E., et al. (2013). SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics* 29, 409–411. doi: 10.1093/bioinformatics/bts698

Marmol, I., Sanchez-De-Diego, C., Pradilla Dieste, A., Cerrada, E., and Rodriguez Yoldi, M. J. (2017). Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. *Int. J. Mol. Sci.* 18:197. doi: 10.3390/ijms18010197

Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* 10, 155–159.

Mirza, A. H., Berthelsen, C. H., Seemann, S. E., Pan, X., Frederiksen, K. S., Vilien, M., et al. (2015). Transcriptomic landscape of lncRNAs in inflammatory bowel disease. *Genome Med.* 7:39.

Ni, X., Ding, Y., Yuan, H., Shao, J., Yan, Y., Guo, R., et al. (2020). Long non-coding RNA ZEB1-AS1 promotes colon adenocarcinoma malignant progression via miR-455-3p/PAK2 axis. *Cell Prolif.* 53:e12723.

Padua, D., Mahurkar-Joshi, S., Law, I. K., Polytarchou, C., Vu, J. P., Pisegna, J. R., et al. (2016). A long noncoding RNA signature for ulcerative colitis identifies IFNG-AS1 as an enhancer of inflammation. *Am. J. Physiol. Gastrointest. Liver Physiol.* 311, G446–G457.

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, 2498–2504. doi: 10.1101/gr.1239303

Slattery, M. L., Lundgreen, A., and Wolff, R. K. (2014). VEGFA, FLT1, KDR and colorectal cancer: assessment of disease risk, tumor molecular phenotype, and survival. *Mol. Carcinog.* 53(Suppl. 1), E140–E150.

Tay, Y., Rinn, J., and Pandolfi, P. P. (2014). The multilayered complexity of ceRNA crosstalk and competition. *Nature* 505, 344–352. doi: 10.1038/nature12986

Tomczak, K., Czerwinska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol.* 19, A68–A77.

Wang, D., Gu, J., Wang, T., and Ding, Z. (2014). OncomiRDB: a database for the experimentally verified oncogenic and tumor-suppressive microRNAs. *Bioinformatics* 30, 2237–2238. doi: 10.1093/bioinformatics/btu155

Wang, W., Lou, W., Ding, B., Yang, B., Lu, H., Kong, Q., et al. (2019a). A novel mRNA-miRNA-lncRNA competing endogenous RNA triple sub-network associated with prognosis of pancreatic cancer. *Aging* 11, 2610–2627. doi: 10.18632/aging.101933

Wang, W., Xie, Y., Chen, F., Liu, X., Zhong, L. L., Wang, H. Q., et al. (2019b). LncRNA MEG3 acts a biomarker and regulates cell functions by targeting ADAR1 in colorectal cancer. *World J. Gastroenterol.* 25, 3972–3984. doi: 10.3748/wjg.v25.i29.3972

Wang, Y., He, R., and Ma, L. (2020). Characterization of lncRNA-associated ceRNA network to reveal potential prognostic biomarkers in lung adenocarcinoma. *Front. Bioeng. Biotechnol.* 8:266. doi: 10.3389/fbioe.2020.00266

Xie, M., Zhao, F., Zou, X., Jin, S., and Xiong, S. (2017). The association between CCND1 G870A polymorphism and colorectal cancer risk: a meta-analysis. *Medicine* 96:e8269. doi: 10.1097/md.0000000000008269

Xu, J., Meng, Q., Li, X., Yang, H., Xu, J., Gao, N., et al. (2019). Long Noncoding RNA MIR17HG Promotes Colorectal Cancer Progression via miR-17-5p. *Cancer Res.* 79, 4882–4895. doi: 10.1158/0008-5472.can-18-3880

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., et al. (2020). Ensembl 2020. *Nucleic Acids Res.* 48, D682–D688.

Yin, D. D., Liu, Z. J., Zhang, E., Kong, R., Zhang, Z. H., and Guo, R. H. (2015). Decreased expression of long noncoding RNA MEG3 affects cell proliferation and predicts a poor prognosis in patients with colorectal cancer. *Tumour Biol.* 36, 4851–4859. doi: 10.1007/s13277-015-3139-2

Yin, K., Zhang, Y., Zhang, S., Bao, Y., Guo, J., Zhang, G., et al. (2019). Using weighted gene co-expression network analysis to identify key modules and hub genes in tongue squamous cell carcinoma. *Medicine* 98:e17100. doi: 10.1097/MD.0000000000017100

Yu, Y., Nangia-Makker, P., Farhana, L., and Majumdar, A. P. N. (2017). A novel mechanism of lncRNA and miRNA interaction: CCAT2 regulates miR-145 expression by suppressing its maturation process in colon cancer cells. *Mol. Cancer* 16:155. doi: 10.1186/s12943-017-0725-5

Zhang, W., Tang, G., Zhou, S., and Niu, Y. (2019). LncRNA-miRNA interaction prediction through sequence-derived linear neighborhood propagation method with information combination. *BMC Genomics* 20:946. doi: 10.1186/s12864-019-6284-y

Zhou, Z., Shen, Y., Khan, M. R., and Li, A. (2015). LncReg: a reference resource for lncRNA-associated regulatory networks. *Database* 2015:bav083. doi: 10.1093/database/bav083

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 September 2020
doi: 10.3389/fgene.2020.572663

Lysine Acetylome Study of Human Hepatocellular Carcinoma Tissues for Biomarkers and Therapeutic Targets Discovery

*Qianwei Zhao[1,2], Zhendong Zhang[1], Jinxia Li[3], Fang Xu[1], Bingxia Zhang[4], Mengduan Liu[1], Yixian Liu[1], Huiping Chen[1], Junxia Yang[1] and Jintao Zhang[1]**

[1] BGI College and Henan Institute of Medical and Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, China, [2] Henan Key Laboratory for Pharmacology of Liver Diseases, Zhengzhou University, Zhengzhou, China, [3] School of Basic Medical Sciences, Zhengzhou University, Zhengzhou, China, [4] School of Life Sciences, Zhengzhou University, Zhengzhou, China

Lysine acetylation is a vital post-translational modification (PTM) of proteins, which plays an important role in cancer development. In healthy human liver tissues, multiple non-histone proteins were identified with acetylation modification, however, the role of acetylated proteins in hepatocellular carcinoma (HCC) development remains largely unknown. Here we performed a quantitative acetylome study of tumor and normal liver tissues from HCC patients. Overall, 598 lysine acetylation sites in 325 proteins were quantified, and almost 59% of their acetylation levels were significantly changed. The differentially acetylated proteins mainly consisted of non-histone proteins located in mitochondria and cytoplasm, which accounted for 42% and 24%, respectively. Bioinformatics analysis showed that differentially acetylated proteins were enriched in metabolism, oxidative stress, and signal transduction processes. In tumor tissues, 278 lysine sites in 189 proteins showed decreased acetylation levels, which occupied 98% of differentially acetylated proteins. Moreover, we collected twenty pairs of tumor and normal liver tissues from HCC male patients, and found that expression levels of SIRT1 (p = 0.002), SIRT2 (p = 0.01), and SIRT4 (p = 0.045) were significantly up-regulated in tumor tissues. Over-expression of possibly accounted for the widespread deacetylation of non-histone proteins identified in HCC tumor tissues, which could serve as promising predictors of HCC. Taken together, our work illustrates abundant differentially acetylated proteins in HCC tumor tissues, and offered insights into the role of lysine acetylation in HCC development. It provided potential biomarker and drug target candidates for clinical HCC diagnosis and treatment.

Keywords: hepatocellular carcinoma, lysine acetylation, non-histone proteins, sirtuins, biomarker candidates, drug discovery

OPEN ACCESS

Edited by:
Xiaofeng Dai,
Jiangnan University, China

Reviewed by:
Fengfeng Zhou,
Jilin University, China
Gary David Lopaschuk,
University of Alberta, Canada

***Correspondence:**
Jintao Zhang
zhangjt_66@126.com

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *15 June 2020*
Accepted: *26 August 2020*
Published: *17 September 2020*

Citation:
Zhao Q, Zhang Z, Li J, Xu F, Zhang B, Liu M, Liu Y, Chen H, Yang J and Zhang J (2020) Lysine Acetylome Study of Human Hepatocellular Carcinoma Tissues for Biomarkers and Therapeutic Targets Discovery. Front. Genet. 11:572663. doi: 10.3389/fgene.2020.572663

INTRODUCTION

HCC is one of the most common human cancer types and accounts for about 90% of primary liver cancer (El-Serag and Rudolph, 2007). Because of its high degree of malignancy, recurrence, metastases, and poor prognosis, HCC is the second deadliest human cancer in the world. The 5-year survival rate of HCC patients at advanced stage is less than 10% while patients at early

stage can experience significantly improved survival (Hiwatashi et al., 2015). To explore effective biomarkers for HCC diagnosis at early stage, multiple kinds of omics studies have been performed with samples from HCC patients or HCC cell lines, such as genomics, transcriptomics, proteomics, metabolomics, and so on. These omics studies provided huge number of molecules including circulating tumor DNA (ctDNA), circulating miRNAs, proteins, and metabolites as biomarker candidates for clinical HCC diagnosis (Tang et al., 2016; Zhou et al., 2011; Dittharot et al., 2018; Gao et al., 2019). Based on the data from omics studies, several effective biomarkers were identified and validated in clinical samples from HCC patients, among which AFP was the most widely used protein biomarker for early diagnosis of HCC (Marrero et al., 2009). However, clinical studies showed that the sensitivity and specificity of one protein marker was limited in detecting HCC at early stage (Sun et al., 2008).

In recent years, with the development of biotechnology and its application in omics study, different types of protein PTM were identified and researches indicated that protein PTM was closely associated with disease development. Protein PTM was usually evolutionarily conserved and regulated protein stability, localization, function, protein-protein interaction, and protein-nucleic acid interaction. Omics studies of protein PTM including ubiquitination, methylation, glycosylation, acetylation, crotonylation, and lactylation showed significant difference between tumor and normal samples, which indicated that protein PTM played an important role in cancer development (Zhang et al., 2019; Wan et al., 2019; Zhong and Huang, 2016; Sun et al., 2006; Chen et al., 2018). For example, studies showed that the glycosylation levels of proteins significantly changed in different types of tumor tissues, and some glycosylated proteins were promising biomarkers for diagnosis of HCC at early stage (Wang et al., 2017; Block et al., 2005). Therefore, omics study of protein PTM was of great importance because it not only provided new insight into mechanism study of proteins in cancer development, but also served as new candidates for biomarker and therapy target discovery.

Lysine acetylation was a kind of protein PTM that is involved in various cellular processes, such as metabolic pathways, signal transduction, cell proliferation, migration and apoptosis. Increasing evidence showed that lysine acetylation participated in development of multiple kinds of cancers (Martile et al., 2016; Gil et al., 2017). For example, Lysine acetylation on histones could change local chromatin structure for transcription factors to bind and initiate gene transcription (Jin et al., 2011). Lysine acetylation on non-histone proteins including enzymes and transcription factors induced metabolic rewiring and gene transcription, and affected cell proliferation, apoptosis, and metastasis (Leo et al., 2019; Thangjam et al., 2016; Wang B.et al., 2020; He et al., 2019). These findings indicated that lysine acetylation was an important factor in cancer development. In human liver tissues, more than 1000 acetylation sites in proteins were identified, among which metabolic enzymes accounted for a large amount (Zhao et al., 2010). Increasing evidence showed that lysine acetylation modification on enzymes played a vital role in metabolic processes during liver disease development (Weems and Olson, 2011; Cao et al., 2017; Hu et al., 2017; Gu et al., 2020). Besides, most members of histone acetylase (HAT) and histone deacetylase (HDAC) families were reported to be aberrantly expressed in HCC tumor tissues, which were associated with clinical stage, prognosis, and survival rate, and some HDACs inhibitors were taken as candidates for clinical HCC treatment (Li et al., 2011; Quint et al., 2011; Freese et al., 2019; Hu et al., 2019; Inagaki et al., 2016). Taken together, it showed that lysine acetylation was playing an important role in HCC initiation and development. Therefore, it is necessary to study lysine acetylation in liver tissues of HCC patients thoroughly. Here, we performed lysine acetylome study of HCC tumor and normal liver tissues, and identified multiple proteins with differential acetylation levels. The differentially acetylated proteins mainly consisted of non-histone proteins located in mitochondria and cytoplasm. Besides, SIRT1, 2, and 4 showed increased expression levels in tumor tissues, which probably accounted for widespread deacetylation of the non-histone proteins, and may serve as diagnostic predictors of HCC.

MATERIALS AND METHODS

Experimental Design and Statistical Rationale

The tissue samples were selected from HCC male patients (stage II) with an average age of 43 (41–47), who had not been treated before. The para-carcinoma normal liver tissues around the tumor tissues from the same HCC patients were selected as control samples. Principal Component Analysis (PCA) method was used to represent the correlation between tumor and normal liver tissues (**Supplementary Figure S1**). The acetylation intensity of tissue samples was quantified by four full-quantitative quantitative experiments. First step, we calculated differential abundance of the acetylation between the cancer and normal tissue samples. In brief, we firstly calculated the average value of the quantitative values of each sample in four biological replicates, and then we calculated the ratio of the average values between the cancer and normal samples. The ratio was taken as the final quantitation. Next step, the significant p value of differential expression between two samples was calculated. The relative quantitative values of each sample were taken as log2 transform (so that the data conforms to the normal distribution), and p value was calculated by the two-sample two-tailed T-test method. Among the differentially expressed proteins (with a corrected p value < 0.05), it was considered as significant up-regulation or down-regulation when the fold change was greater than 1.5 or less than 0.67. Raw abundance ratios of acetylation sites were normalized based on the corresponding protein ratio. Negative logarithm (-log10) transformation was carried out for p value (obtained from Fisher's exact test) to indicate the significance of GO, KEGG, and protein domain enrichment.

Sample Preparation

Five pairs of tumor and para-carcinoma normal liver tissues from HCC patients were surgically resected in the First Affiliated Hospital of Zhengzhou University, with approval from the Research Ethics Committee of Zhengzhou University and consent from the patients. The tumor tissues were washed in PBS and surrounding tissues were removed. The para-carcinoma normal liver tissues were managed with the same method. The whole process took place on ice, and it was finished within 30 min after the surgery. The collected tissues were divided into small pieces and stored in 1.5 ml tubes, which were kept in liquid nitrogen. The selected samples were renamed with codes (A1, A2, C1, C2, D1, D2, E1, E2, F1, F2, the number 1 represented tumor tissues while 2 represented normal tissues) instead of the patient′s name in this study.

Protein Extraction and Digestion

Samples were homogenized by a high intensity ultrasonic processor (Scientz, Germany) in 1 mL lysis buffer (8 M urea, 3 μM TSA, 50 mM NAM, 2 mM EDTA, and 1% Protease Inhibitor Cocktail). Then the lysate was centrifuged at 12000 g for 10 min at 4°C, and the supernatant was transferred to new tubes. The concentration of the protein was determined by BCA Protein Assay Kit (Sangon Biotech, China) before use. For digestion, the protein was first reduced with DTT (5 mM) for 30 min at 56°C. Then iodoacetamide was added to the protein to make its final concentration of 11 mM, and the mixture was incubated for 15 min at room temperature in the dark. Finally, the concentration of urea in the protein solution was diluted to less than 2 M by adding triethyl ammonium bicarbonate (TEAB, 100 mM). The protein was digested by trypsin with a trypsin/protein mass ratio of 1/50 at 37°C overnight, and then additional trypsin was added to the protein with a trypsin/protein mass ratio of 1/100 and incubated at 37°C for 4 h.

TMT Labeling and Affinity Enrichment of Acetylated Peptides

Tryptic peptides were freeze-dried after being desalted by using Strata X C18 (Phenomenex, United States). Then the peptides were dissolved in TEAB (0.5 M) to be labeled by TMT tag according to the protocol of the TMT labeling kit (Thermo, United States). TMT reagents were thawed and reconstituted in acetonitrile. The peptide mixtures were incubated for 2 h at room temperature, and then desalted and dried by vacuum centrifugation. Next, TMT labeled peptides were dissolved in NETN buffer (100 mM NaCl, 1 mM EDTA, 50 mM Tris–HCl, 0.5% NP-40, pH = 8.0), and the supernatant was collected and incubated with agarose beads coupled to anti-acetyl antibody PTM-104 (PTM Bio, Jingjie, China) overnight at 4°C with gentle shaking. After, the beads were washed with NETN buffer four times, and then twice with ddH_2O. Finally, the bound peptides were eluted with 0.1% trifluoroacetic acid. The eluent was vacuum-dried and then desalted by C18 ZipTips (Millipore, United States) according to the manufacturer's instruction. The TMT labeled peptides were freeze-dried for Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

Protein Identification and Quantification by LC-MS/MS Analysis

LC-MS/MS analysis was performed according to the literature procedure. The EASY-nLC 1000 UPLC system was used to separate peptides dissolved in 0.1% formic acid (FA). The mobile phase consisted of buffer A (0.1% FA and 2% acetonitrile in ddH_2O) and buffer B (0.1% FA and 90% acetonitrile in ddH_2O); it was eluted with a linear gradient of eluent, starting with 10% buffer B which was increased to 25% in 40 min, and then from 25% to 38% for 12 min, and later from 38% to 80% for 4 min, finally, it was kept at 80% buffer B for 4 min. The flow rate was 700 nL/min. The peptides were subjected to the NIS ion source and then performed MS/MS analysis by Orbitrap Fusion™ (Thermo, United States) system. The electrospray voltage applied was 2.0 kV. For MS1 scan, it was set at a resolution of 60,000 with a scan ranging from 350 to 1550 m/z. For MS2 scan, it was set at a resolution of 30,000 (100 m/z). Data were generated by using the Data-Dependent Acquisition (DDA) strategy.

Database Search

The MS/MS data was performed by Maxquant search engine (v1.5.2.8), and tandem mass spectra was searched against SwissProt Human (20130 sequences) database. FurtheMore, additional libraries were added for false discovery rate (FDR) calculation and elimination of contaminated protein′s influence. Trypsin/P was specified as the cleavage enzyme allowing up to 4 missing cleavages, 7 modifications per peptide and 5 charges. Mass error was set to 10 ppm and 5 ppm for precursor ions of first and main searches, respectively, and mass error for fragment ions was set to 0.02 Da. Carbamidomethyl on Cys was specified as fix modification, while oxidation on Met and acetylation on Lys and N-terminal of protein were specified as variable modifications. Quantitative method was set to TMT-10plex. FDR thresholds for protein, peptide and modification sites were specified at 1%.

Bioinformatics Analysis

Gene Ontology (GO) annotation proteome was performed from the UniProt-GOA database[1], and proteins were classified into three categories: biological process, cellular component and molecular function. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was based on database[2]. Subcellular location was derived from the software named WoLF PSORT. Protein domain was annotated by soft InterProScan based on InterPro domain database[3]. Motif-x software was employed to analyze the model of sequences constituted with amino acids in specific positions of acetyl-21-mers (10 amino acids upstream and downstream of the site) in all the protein sequences. Protein-protein interaction (PPI) network analysis was performed by STRING 10.5[4].

[1]https://www.ebi.ac.uk/GOA/
[2]http://www.genome.jp/kegg/
[3]http://www.ebi.ac.uk/interpro/
[4]https://string-db.org/cgi/input.pl

Western Blot Analysis

About 100 mg tumor or normal liver tissues were acquired and digested in 1 mL RIPA lysis buffer on ice for 30 min. The lysate was centrifuged for 20 min, 12000 rpm at 4°C, protein extraction samples were collected and concentration were detected with a microplate reader. Samples were boiled at 95°C for 5 min and directly loaded onto a 10% SDS-PAGE gel. Then proteins were transferred to a PVDF membrane. The membrane was incubated within 5% non-fat dry milk in TBS with 0.1% Twee-20 (TBS-T) to block the protein spots, and then it was incubated with primary antibodies (SIRT1, 2, and 4 from BBI, China) at 4°C overnight. The membrane was washed with TBS-T for three times and incubated in secondary antibody. Signals were detected with the Amersham Imager 600 System (General Electric Company, United States). GAPDH was used as internal control. The expression levels were quantified by measuring the intensity of each band with Image J software.

RESULTS

Identification and Analysis of Acetylated Proteins in HCC Tumor and Normal Liver Tissues

To profile the protein lysine acetylation of liver tissues during HCC development, five pairs of primary tumor and para-carcinoma normal liver tissues were surgically resected from HCC patients. Four pairs of qualified tissues were selected for further analysis after quality test. Proteins were extracted from the collected liver tissues and then digested into peptides by trypsin. After being labeled with TMT tag, the digested peptides with lysine acetylation sites were enriched and analyzed by LC-MS/MS. Based on the human Uniprot database, we identified 792 acetylation sites in 415 proteins, among which 598 acetylation sites in 325 proteins were quantified. Compared to the normal liver tissues, 278 acetylation sites in 189 proteins were down-regulated while three acetylation sites in three proteins (EP300, GRHPR, and GLDC) were up-regulated with significant differences in tumor tissues (**Supplementary Table S1**; **Figure 1A**). In a previous acetylation proteomics study, more than 1000 acetylated proteins were identified in human liver tissues (Zhao et al., 2010). By comparison of these two studies, we found that about 31% of proteins with differential acetylation levels identified in our study were present in the previous study (**Figure 1B**). It indicated that the spectrum of acetylated proteins was highly conserved in human liver tissues, and the proteins with aberrant acetylation levels in tumor tissues probably participated in HCC development.

Then we investigated the distribution of identified lysine acetylation sites in proteins by calculating the number of acetylation sites per protein. As shown in **Figure 1C**, most of the proteins contained no more than three lysine acetylation sites, which accounted for 87%, while 13% of them contained three more sites. Next, distribution of acetylated lysine sites and proteins were analyzed (**Figure 1D**), it showed that acetylated proteins were mainly associated with mitochondria (31%), nucleus (24%), and cytoplasm (22%). Besides, among the remaining 23% proteins, most of them were located in organelles including endoplasmic reticulum and peroxisome. To explore the relationship between lysine acetylation sites and protein secondary structures, a structure analysis of the acetylated lysine sites and all lysine sites was performed. As shown in **Figure 1E**, the α-helix and coil probabilities of acetylated lysine sites were significantly different from that of all lysine sites, while there was no obvious difference in β-strand probabilities between the acetylated lysine sites and all lysine sites. Additionally, surface accessibility of lysine acetylation sites analysis showed that about 38.55% of acetylated lysine sites were exposed to protein surfaces, compared to 38.88% of all lysine sites (**Figure 1F**). Therefore, lysine acetylation seemed to have little effect on surface property of identified proteins in HCC liver tissues.

We also analyzed the amino acids from the -10 to + 10 positions surrounding the identified lysine acetylation sites and searched for occurrences of amino acid motifs. Of all the acetylated peptides, 417 were matched to six conserved motifs, including KK.,.KH., ...K. . .KA. . . . . .., ...KR. . . . . .., ...KN., and K.K (K indicates the acetylated lysine, and H, A, R, N represent histidine, alanine, arginine, and asparagine, respectively) (**Figure 2A**). In **Figure 2B**, local sequence context around the lysine acetylation sites was analyzed and it indicated that positively charged amino acids (K, H, and R) were almost completely excluded from the -1 position, while they were enriched in the + 1 position. This data suggested that according to the heat map of the amino acid compositions surrounding the acetylation sites, the frequency of G (G indicates glycine) in position -1 was the highest, and the frequency of K and A from the -10 to + 10 positions was mainly kept at a higher level in the motifs. Acetylated peptides containing motif. KK. accounted for 35% of all the identified peptides, while peptides containing motifs. KR.,.K.K., ...KH. . . . . .., ...KN. . . . . .., and . . . . ..K.KA. accounted for 17.3%, 14.6%, 13.7%, 13.7%, and 5.8%, respectively (**Figure 2C**). In addition, the distribution of different motifs in cellular compartments were assessed, and results showed that acetylated peptides were predominantly associated with cytoplasm and mitochondria, where the motifs were similar, but different from nuclear motifs (**Figures 2D–F**).

Functional Enrichment and Subcellular Location of Differentially Acetylated Proteins

To characterize the role of lysine acetylome in HCC development, we performed GO analysis on the differentially acetylated proteins based on biological process, cellular component, and molecular function (**Figures 3A–C**). Biological processes analysis indicated that acetylated proteins were enriched in cellular process (17%), single-organism process (16%), metabolic process (16%), biological regulation (9%) and response to stimulus (9%). Besides, according to cellular component analysis, there were 24% and 23% acetylated proteins located in cell and organelle, respectively, 15% in membrane-enclosed lumen,

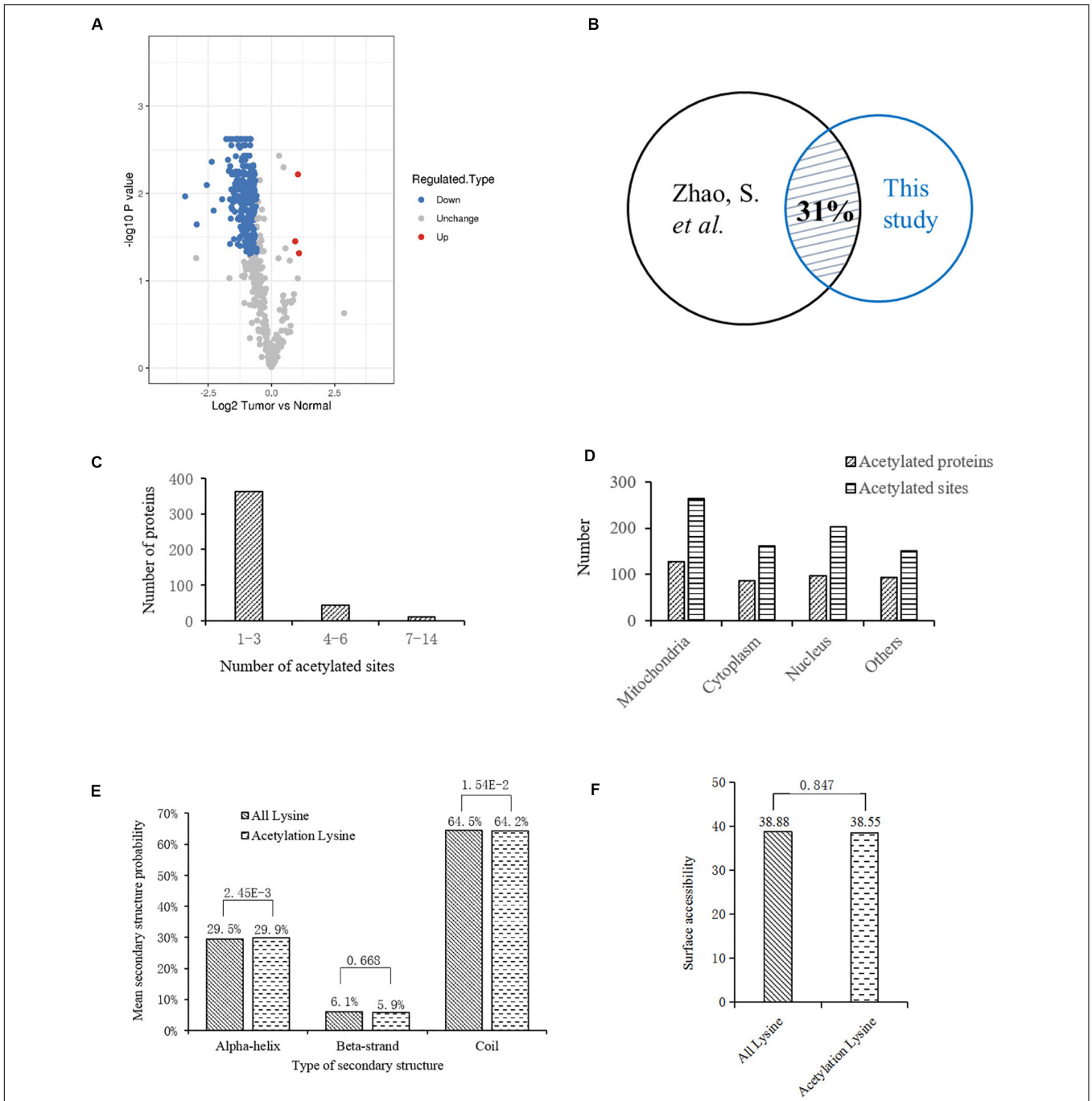

FIGURE 1 | Bioinformatics analysis of lysine acetylation sites identified in liver tissues of HCC patients. **(A)** Volcano plot showing the distribution of acetylation sites rations (Tumor vs. Normal). **(B)** Comparison of two acetylation proteomic studies in human liver tissues: this study and reference (Zhao et al., 2010). **(C)** Statistic analysis of acetylated sites in the acetylated proteins. **(D)** Distribution of acetylated sites and proteins in different cellular components. **(E)** Distribution of lysine acetylation sites in protein secondary structures. **(F)** Predicted surface accessibility of acetylation sites.

15% in extracellular region, 11% in membrane, and 8% in macromolecular complex. Molecular function analysis showed that 45% and 40% acetylated proteins were involved in binding and catalytic activities, respectively, which accounted for most of the acetylated proteins. Taken together, GO analysis revealed that differentially acetylated proteins were enriched in metabolic, catabolic, and signal transduction processes. Consistently, subcellular location showed that differentially acetylated proteins were primarily associated with mitochondria and cytoplasm, which accounted for 42% and 24%, respectively.

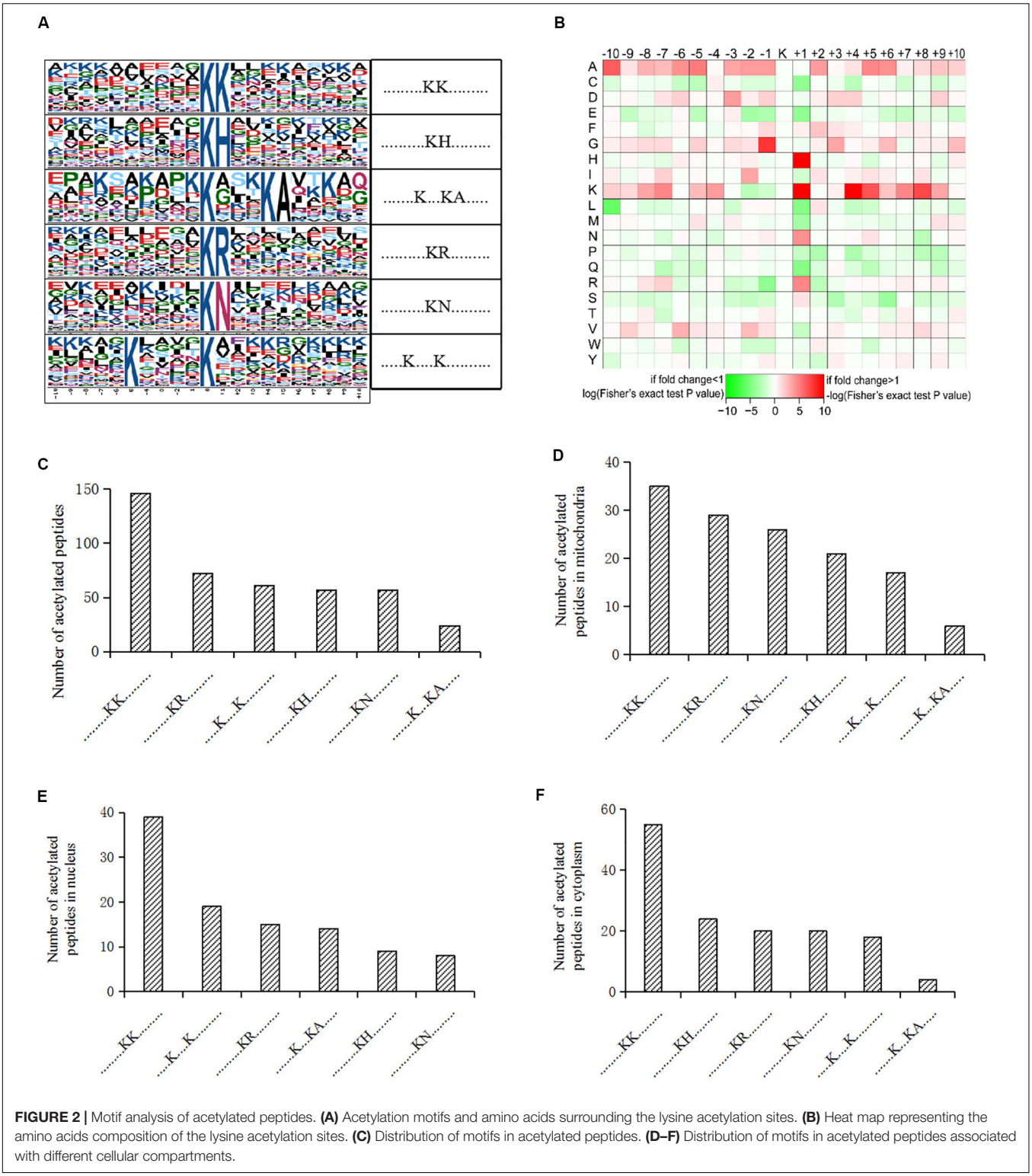

FIGURE 2 | Motif analysis of acetylated peptides. **(A)** Acetylation motifs and amino acids surrounding the lysine acetylation sites. **(B)** Heat map representing the amino acids composition of the lysine acetylation sites. **(C)** Distribution of motifs in acetylated peptides. **(D–F)** Distribution of motifs in acetylated peptides associated with different cellular compartments.

Less acetylated proteins were in located in endoplasmic reticulum and peroxisome (11%), nucleus (7%), membrane and cytoskeleton (4%), and extracellular (3%) (**Figure 3D**).

To better understand the cellular processes regulated by differentially acetylated proteins in HCC tumor tissues, GO and KEGG pathway enrichment analysis were carried out. As shown in **Figure 4A**, it showed that differentially acetylated proteins were markedly involved in epithelium development, mitochondria function, and intracellular metabolic processes. In agreement with this finding, KEGG pathway enrichment analysis

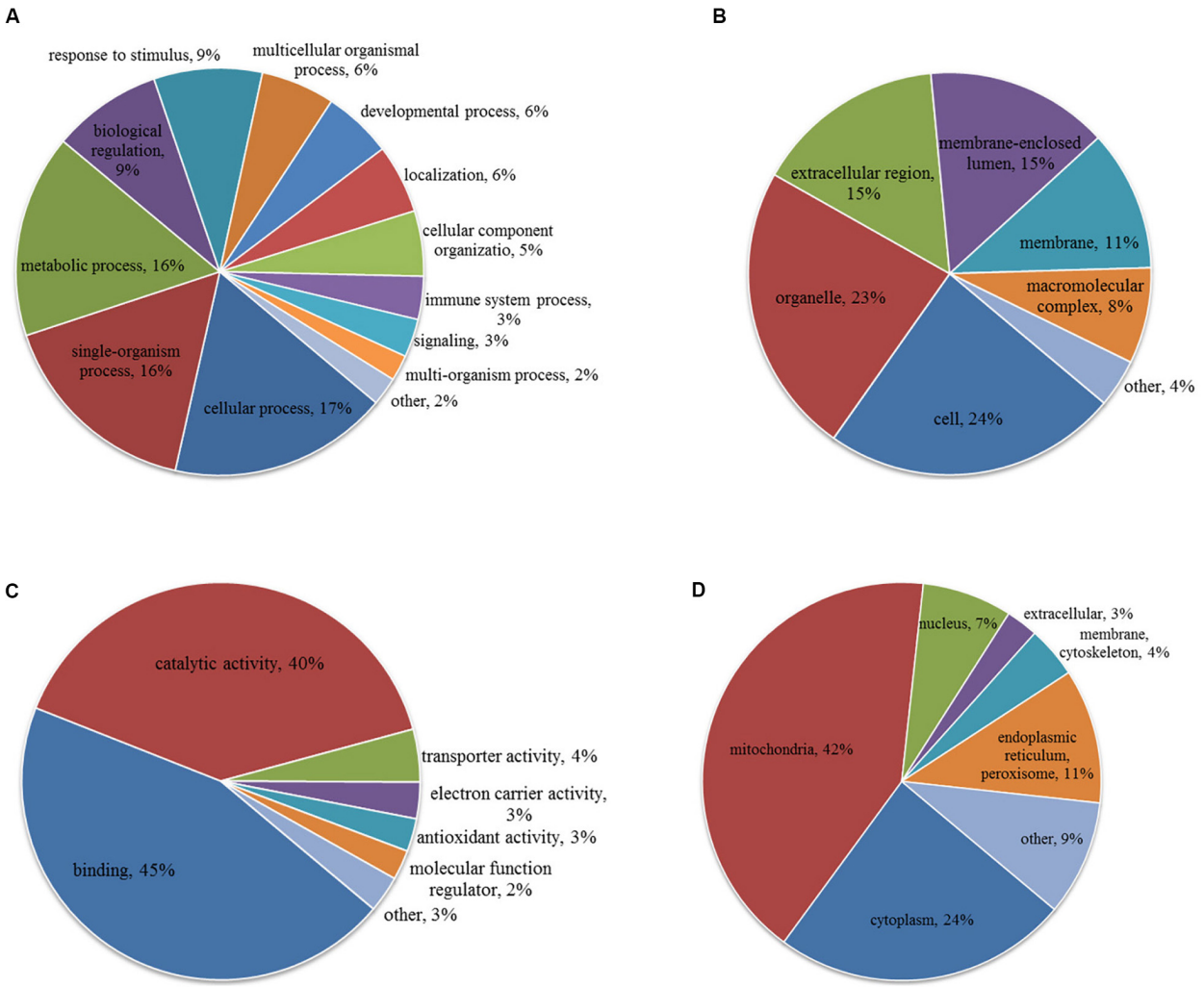

FIGURE 3 | GO **(A–C)** and subcellular localization **(D)** analysis of differentially acetylated proteins. **(A)** Biological process. **(B)** Cellular component. **(C)** Molecular function.

demonstrated that differentially acetylated proteins participating in glucose, fatty acid, and amino acids metabolism pathways were mostly enriched (**Figure 4B**). Besides, protein domain enrichment analysis showed that CIpP/crotonase-like, Thiolase-like, Thiolase, N-terminal, and Thiolase, C-terminal domains were mainly enriched (**Figure 4C**). All these domains were involved in enzymatic activities, especially lipid metabolism. Additionally, the differentially acetylated proteins were subjected to PPI network analysis based on the STRING database, and it showed that proteins were clustered into three categories including metabolism, oxidative stress and DNA repair, and the ubiquitin proteasome system (UPS) (**Figure 5**).

Disorder of Lysine Acetylation Modification on Metabolic Enzymes in HCC Tumor Tissues

Among the differentially acetylated proteins identified in this research, metabolic enzymes represented a significant proportion, and they mainly participated in processes such as glycolysis, gluconeogenesis, the TCA cycle, fatty acid oxidation, glutamine metabolism, and the urea cycle (**Figure 5**). According to the statistical analysis, proteins participating in glucose, fatty acid and amino acid metabolism accounted for 46%, 39%, and 15%, respectively. In the glucose metabolic process, most differentially acetylated enzymes directly participated in glycolysis (GAPDH, PGK1, PGAM2, LDHA, DLAT, PDHA1), TCA cycle (IDH1/2, DHTKD1, SDHA, FH, MDH1/2), and oxidative phosphorylation (NDUFB3, SDHA, UQCRC1, ATPase, COX5B) processes (**Figure 6A**). Additionally, in fatty acid metabolic process (**Figure 6B**), about 60% of enzymes contained more than one lysine sites with differential acetylation levels. For example, we identified five lysine acetylation sites in HADHA, four lysine acetylation sites in HADHB, and three lysine acetylation sites in both ACADVL and PCCA, all of which showed lower acetylation levels in tumor tissues. Besides, several acetylated proteins

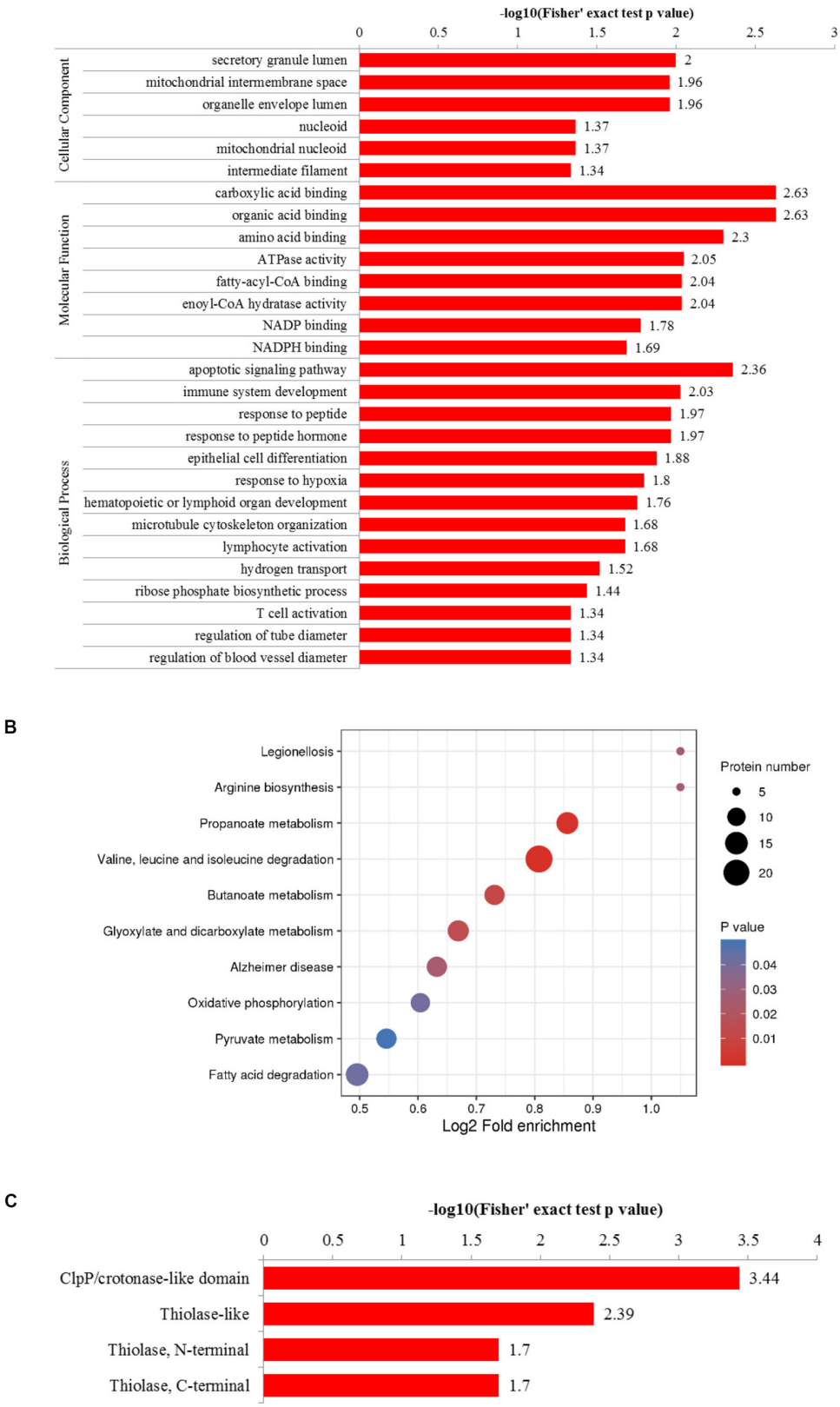

FIGURE 4 | GO enrichment **(A)**, KEGG pathway enrichment **(B)**, and protein domain enrichment **(C)** analysis of differentially acetylated proteins.

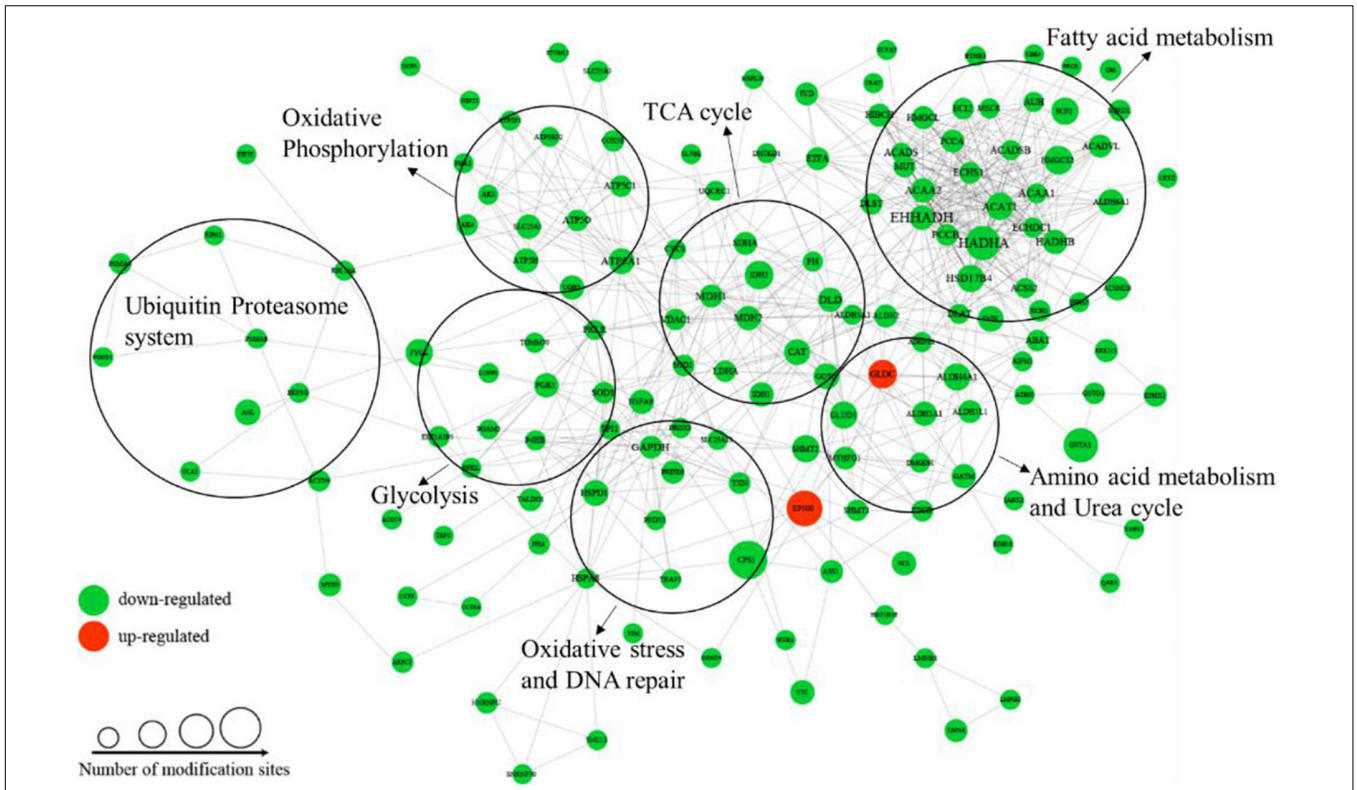

FIGURE 5 | PPI network analysis of differentially acetylated proteins in HCC tumor tissues.

with decreased lysine acetylation levels were located in peroxisome, where the main function was to catalyze the oxidation of fatty acid. Moreover, several pivotal enzymes involved in amino acid metabolism process and urea cycle were identified with decreased lysine acetylation levels in HCC tumor tissues, such as GLUD1, ASL, GOT2 and so on (**Figure 6C**). GLUD1 plays an important role in glutamine consumption, which is a common feature of HCC liver tissues and provides energy for tumor growth (Dang, 2013). Here we identified five lysine acetylation sites in GLUD1, and all the acetylation levels were down-regulated in tumor tissues. In general, most of the acetylation levels of lysine sites in enzymes decreased in tumor tissues (**Figure 6D**). Additionally, some of the lysine-acetylated peptides were confirmed by MS/MS spectra, with their relative intensities in HCC tumor and normal liver tissues (**Supplementary Figures 2A–C**).

Involvement of Acetylated Proteins in Signal Transduction Process and Oxidative Stress in HCC Tumor Tissues

Besides metabolic enzymes, transcription factors accounted for another large part of non-histone substrates of HATs and HDACs. Transcription factors usually function on genomic DNA and play important roles in genes transcription. In this research, nearly 10% of differentially acetylated proteins were located in nuclear, and most of them were involved in gene transcription process, such as chromatin organization, telomere dynamics, DNA/RNA binding, and mRNA splicing (**Supplementary Figure S3**). To evaluate the importance of these acetylated nuclear factors during HCC progression, here we performed PPI network analysis. Results in **Figure 7A** showed that most of the acetylated nuclear factors had direct and close interaction with important factors in HCC signal pathways, such as p53, PI3K, and c-Myc. Among the nuclear factors, EP300 is a transcriptional co-activator of which the K1550 site showed increased acetylation level in tumor tissues. Besides, as a member of HATs, EP300 had direct association with multiple non-histone proteins, among which several proteins identified in this study were included, such as HSD17B4, EHHADH, LDHA, SLC25A5, NCL, SMAD4, HSPA8, and so on.

HCC development usually accompanies oxidative stress, and antioxidant stress system deficiency is a common feature in tumor cells. In this study, most proteins involved in reactive oxygen species (ROS) and oxidative stress contained lysine acetylation sites, such as SOD1, SOD2, PRDX1, PRDX3, PRDX6, HSD17B4, PECR, GSTA1, TXN, and CAT. Compared to the normal liver tissues, the acetylation levels of lysine sites in these proteins were generally down-regulated in tumor tissues. As a response to oxidative stress and ROS, the damaged proteins are mainly degraded by the UPS, which is to protect cells from ROS damage (Marques et al., 2006). Consistently, several lysine sites of proteins

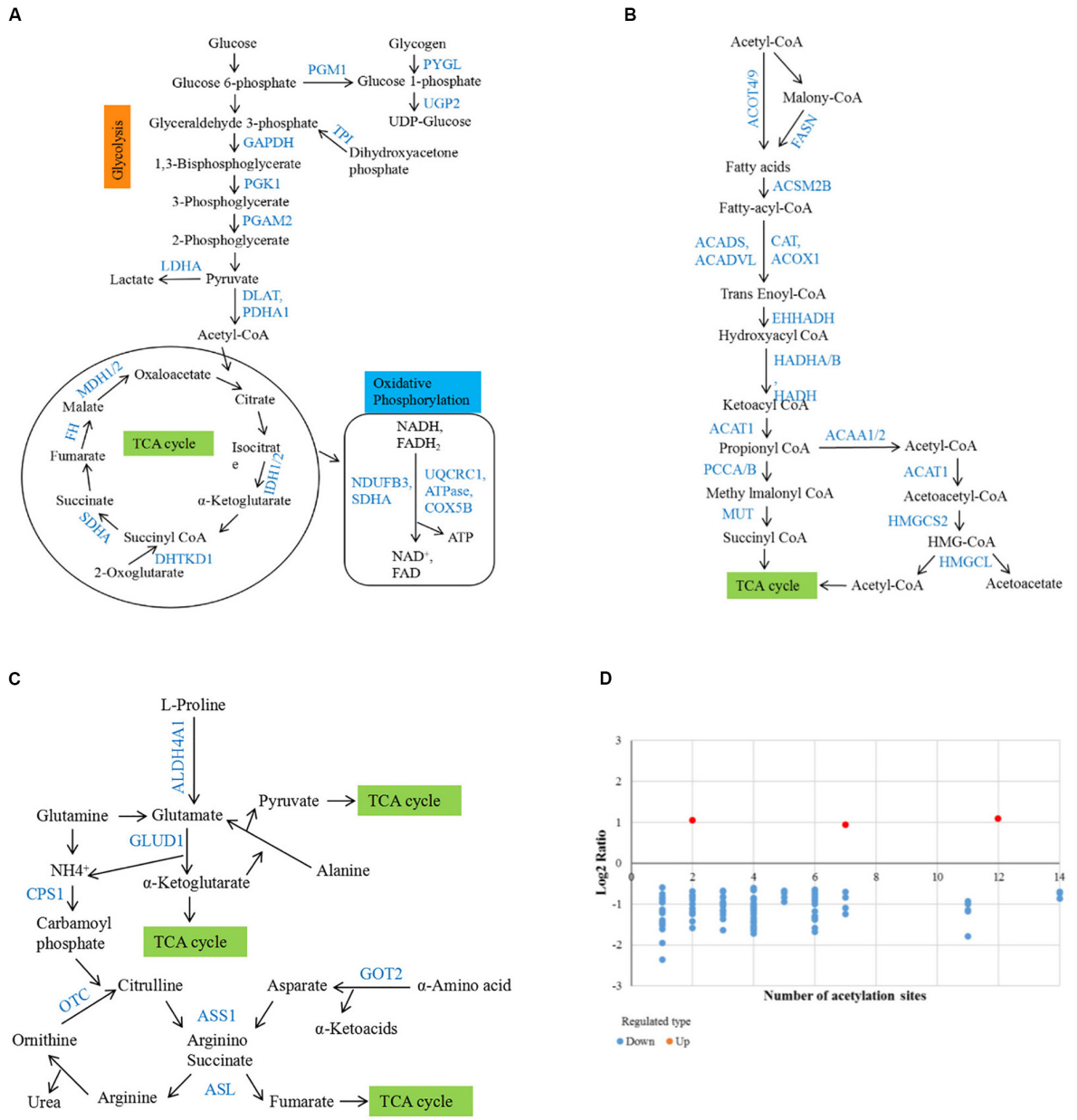

FIGURE 6 | Involvement of differentially acetylated proteins in metabolic pathways. **(A–C)** Schematic representation of differentially acetylated enzymes (shown in blue) in glucose, fatty acid, and amino acid metabolism. **(D)** Distribution of lysine acetylation sites rations in metabolic enzymes (Tumor vs. Normal).

involved in UPS showed decreased acetylation levels in tumor tissues (**Supplementary Table 2**). Moreover, heat shock proteins (HSPs) including HSPA8, HSPA9, and HSPD1 in tumor tissues displayed aberrant lysine acetylation modification. HSPs are molecular chaperones that play key roles in refolding denatured proteins. HSPs promoted tumorigenesis, tumor growth, and metastasis, and blockade of HSPs and activation of UPS induced protein degradation and prevented tumor development (Calderwood, 2018). Finally, twenty acetylated proteins were analyzed in a PPI network (**Figure 7B**), and it showed close interaction among proteins associated with oxidative stress in HCC tumor tissues.

SIRT1, 2 and 4 Were Up-Regulated in HCC Tumor Tissues

According to the bioinformatics analysis, differentially acetylated proteins mainly consisted of non-histone proteins located in mitochondria and cytoplasm, most of which displayed decreased lysine acetylation levels in tumor tissues. Among HDACs, SIRT1-5 accounted for deacetylation of non-histone proteins and several acetylated proteins identified in tumor tissues were substrates

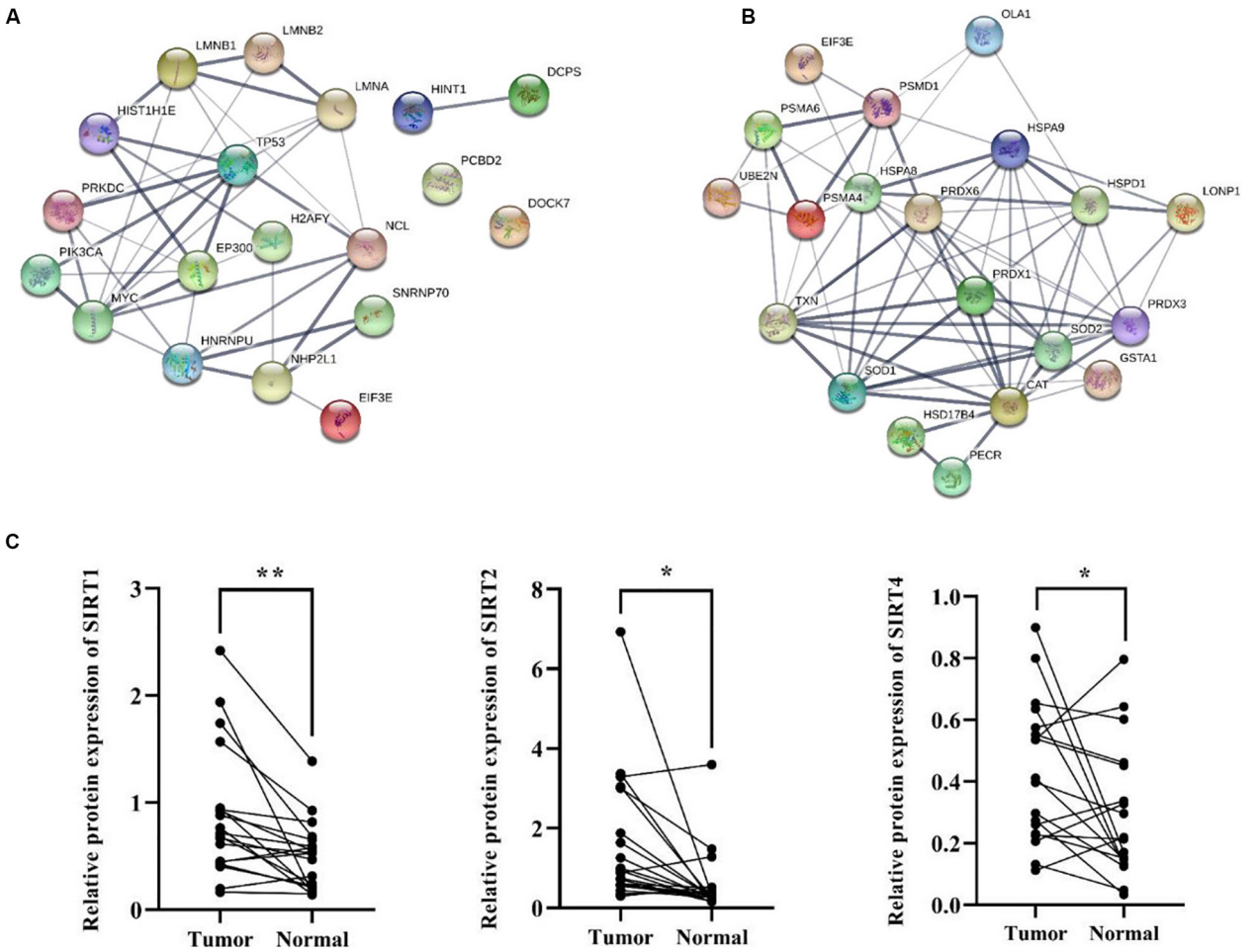

FIGURE 7 | PPI network analysis of differentially acetylated nuclear factors **(A)** and oxidative stress associated proteins **(B)** in HCC tumor tissues. **(C)** Protein expression levels of SIRT1, 2, and 4 in tumor and normal liver tissues from HCC patients. $^{*}p < 0.05$ and $^{**}p < 0.01$ indicate significant differences.

of SIRT1-5. Therefore, we put forward the hypothesis that the widespread deacetylation of lysine sites in non-histone proteins identified in HCC tumor tissues was possibly associated with the function of SIRT1-5. Then we detected expression levels of SIRT1-5 in four pairs of tumor and normal liver tissues from HCC patients. Results showed that SIRT1, 2, and 4 were up-regulated in tumor tissues while SIRT3 and 5 displayed decreased expression levels (**Supplementary Figure S4**). Furthermore, we collected another twenty pairs of tumor and normal liver tissues from HCC male patients. The age of patients ranged from 35 to 67, and about 57% of them were HBV positive. We detected the expression levels of SIRT1, 2, and 4 by western blot and results in **Figure 7C** showed that the frequencies of over-expression of SIRT1, 2, and 4 in tumor tissues were 85%, 80%, and 72%, respectively. Compared to normal liver tissues, SIRT1 ($p = 0.002$), SIRT2 ($p = 0.01$), and SIRT4 ($p = 0.045$) were significantly up-regulated in tumor tissues. These results indicated that up-regulation of SIRT 1, 2, and 4 in HCC tumor tissues possibly accounted for the deacetylation of non-histone proteins located in mitochondria and cytoplasm. Besides, it provided sirtuins as promising biomarker and therapeutic target candidates for HCC diagnosis and cure.

DISCUSSION

In recent years, with the development of LC-MS/MS technique and its application in proteomics study, an increasing number of lysine acetylation sites in non-histone proteins were identified in liver tissues (Zhao et al., 2010; Lundby et al., 2012). More and more evidence showed that lysine acetylation played an important role in metabolic function and signal transduction during HCC development (Ding et al., 2017; Guo et al., 2018). To better understand the function of lysine acetylation during HCC development, we performed lysine acetylome study and identified a large number of lysine sites in proteins with differential acetylation levels in HCC tumor and normal liver tissues.

In this study, the differentially acetylated proteins mainly consisted of non-histone proteins that were normally located in mitochondria and cytoplasm, where metabolic enzymes accounted for a large amount. Lysine acetylation was vital to an

enzyme's function, and disordered lysine acetylation level lead to enzyme dysfunction and metabolic syndrome (Lin et al., 2016). Metabolic rewiring including Warburg effect, enhanced fatty acid metabolism and glutamine consumption was a common feature of HCC, which provided energy for tumor growth and metastasis (Kitamura et al., 2011; Mashek et al., 2015; Infantino et al., 2019; Daye and Wellen, 2012). Consistently, the GO and KEGG pathway analysis showed that differentially modified proteins mainly participated in metabolic processes. For example, pivotal enzymes participating in glycolysis, glutaminolysis, and fatty acid oxidation processes such as PGAM2, GLUD1, HADHA, and so on, displayed aberrant acetylation levels. Besides, some of them contained more than one lysine sites with different acetylation levels in HCC tumor and normal liver tissues. Lysine acetylation modification was an important PTM that was associated with active sites and advanced structure formation. These findings provided us a new way for mechanism study of metabolic enzymes during HCC development. Besides, the differentially acetylated enzymes could serve as potential biomarkers for clinical HCC diagnosis at early stage.

In addition, we found that almost all of the proteins involved in oxidative stress showed decreased acetylation levels in HCC tumor tissues. Moreover, as the main manners for degradation and refolding of damaged and denatured proteins, proteins from the UPS and HSPs displaying disturbed lysine acetylation levels were closely connected with oxidative stress associated proteins. Oxidative stress was induced by excessive amounts of ROS produced from metabolic processes, and the antioxidant stress system was usually damaged or deficient in tumor tissues, which were vital factors for HCC initiation and progression (Birben et al., 2012; Ma-On et al., 2017). These finding not only proved the involvement of metabolic disorder and oxidative stress in promoting HCC development, but also implied the participation of acetylated proteins in ROS production and elimination processes, which might be a pivotal way for oxidative stress regulation in HCC tumor tissues.

Taken together, the lysine acetylome study provided a large amount of non-histone proteins with decreased acetylation levels in tumor tissues. Meanwhile, we found that the expression levels of SIRT 1, 2, and 4 increased in tumor tissues. Among HDACs, SIRT1, 2, and 4 were mainly located in mitochondria and cytoplasm, and they were closely related to metabolic process and oxidative stress during cancer development (Min et al., 2018; Watanabe et al., 2018; Lee et al., 2018). Besides histones, increasing numbers of non-histone proteins were identified as substrates of sirutins, which included several metabolic enzymes identified in this study (Wang L.T.et al., 2020; Xu et al., 2014). Therefore, up-regulation of SIRT 1, 2, and 4 in tumor tissues was probably responsible for the widespread deacetylation of non-histone proteins. Furthermore, researches showed that over-expression of HDACs was closely associated with cancer development, some of which were potential predictors for diagnosis and prognosis of HCC (Quint et al., 2011). HDACs inhibitors could inhibit HCC cells growth in various states including apoptosis, cell cycle arrest, and inhibition of cell migration/invasion, which made HDACs as hopeful therapeutic targets for HCC treatment (Freese et al., 2019). Therefore, over-expression of SIRT1, 2, and 4 in tumor tissues provided new promising biomarkers and drug targets for clinical HCC treatment.

In conclusion, our findings served as an important resource for the functional study of acetylated proteins in HCC development. It made insights into the association of sirtuins with widespread deacetylation of non-histone proteins in HCC tumor tissues, and provided promising diagnostic biomarkers and therapeutic targets for clinical HCC treatment.

DATA AVAILABILITY STATEMENT

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://www.ebi.ac.uk/pride/archive/) via the PRIDE partner repository with the dataset identifier PXD014915.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee of Zhengzhou University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ML, BZ, and YL performed information verification of patients and samples collection. HC and JY contributed to introduction and discussion of this study. QZ, ZZ, JL, and FX contributed to bioinformatics analysis. QZ and ZZ performed western blot analysis. QZ wrote the manuscript. JZ supervised all phases of this study and proofread the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by grants from Natural Science Foundation of Henan Province (No. 182300410361); the Major Project of Science and Technology in Henan Province (No. 161100311400); and the National Science and Technology Major Project of China (2018ZX10302205).

ACKNOWLEDGMENTS

We thank Dr. Jianxiang Shi for assistance in bioinformatic analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.572663/full#supplementary-material

REFERENCES

Birben, E., Sahiner, U. M., Sackesen, C., Erzurum, S., and Kalayci, O. (2012). Oxidative stress and antioxidant defense. *World Allergy Organ J.* 5, 9–19. doi: 10.1097/WOX.0b013e3182439613

Block, T. M., Comunale, M. A., Lowman, M., Steel, L. F., Romano, P. R., Fimmel, C., et al. (2005). Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. *Proc. Natl. Acad. Sci. U.S.A.* 102, 779–784. doi: 10.1073/pnas.0408928102

Calderwood, S. K. (2018). Heat shock proteins and cancer: intracellular chaperones or extracellular signalling ligands? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373:20160524. doi: 10.1098/rstb.2016.0524

Cao, J., Peng, J., An, H., He, Q., Boronina, T., Guo, S., et al. (2017). Endotoxemia-mediated activation of acetyltransferase P300 impairs insulin signaling in obesity. *Nat. Commun.* 8:131. doi: 10.1038/s41467-017-00163-w

Chen, L., Miao, Y., Liu, M., Zeng, Y., Gao, Z., Peng, D., et al. (2018). Pan-cancer analysis reveals the functional importance of protein lysine modification in cancer development. *Front. Genet.* 9:254. doi: 10.3389/fgene.2018.00254

Dang, C. V. (2013). MYC, metabolism, cell growth, and tumorigenesis. *Cold Spring Harb. Perspect. Med.* 3:a014217. doi: 10.1101/cshperspect.a014217

Daye, D., and Wellen, K. E. (2012). Metabolic reprogramming in cancer: unraveling the role of glutamine in tumorigenesis. *Semin. Cell Dev. Biol.* 23, 362–369. doi: 10.1016/j.semcdb.2012.02.002

Ding, G., Li, W., Liu, J., Zeng, Y., Mao, C., Kang, Y., et al. (2017). LncRNA GHET1 activated by H3K27 acetylation promotes cell tumorigenesis through regulating ATF1 in hepatocellular carcinoma. *Biomed. Pharmacother.* 94, 326–331. doi: 10.1016/j.biopha.2017.07.046

Dittharot, K., Jittorntam, P., Wilairat, P., and Sobhonslidsuk, A. (2018). Urinary metabolomic profiling in chronic hepatitis b viral infection using gas Chromatography/Mass Spectrometry. *Asian. Pac. J. Cancer Prev.* 19, 741–748. doi: 10.22034/APJCP.2018.19.3.741

El-Serag, H. B., and Rudolph, K. L. (2007). Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* 132, 2557–2576. doi: 10.1053/j.gastro.2007.04.061

Freese, K., Seitz, T., Dietrich, P., Lee, S. M. L., Thasler, W. E., Bosserhoff, A., et al. (2019). Histone deacetylase expressions in hepatocellular carcinoma and functional effects of histone deacetylase inhibitors on liver cancer cells in vitro. *Cancers* 11:1587. doi: 10.3390/cancers11101587

Gao, Q., Zhu, H., Dong, H., Shi, W., Chen, R., Song, Z., et al. (2019). Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell* 179, 561–577. doi: 10.1016/j.cell.2019.08.052

Gil, J., Ramirez-Torres, A., and Encarnacion-Guevara, S. (2017). Lysine acetylation and cancer: a proteomics perspective. *J. Proteomics* 150, 297–309. doi: 10.1016/j.jprot.2016.10.003

Gu, L., Zhu, Y., Lin, X., Tan, X., Lu, B., and Li, Y. (2020). Stabilization of FASN by ACAT1-mediated GNPAT acetylation promotes lipid metabolism and hepatocarcinogenesis. *Oncogene* 39, 2437–2449. doi: 10.1038/s41388-020-1156-0

Guo, D., Song, X. H., Guo, T. F., Gu, S. G., Chang, X. L., Su, T., et al. (2018). Vimentin acetylation is involved in SIRT5-mediated hepatocellular carcinoma migration. *Am. J. Cancer Res.* 8, 2453–2466.

He, C., Danes, J. M., Hart, P. C., Zhu, Y., Huang, Y., de Abreu, A. L., et al. (2019). SOD2 acetylation on lysine 68 promotes stem cell reprogramming in breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* 116, 23534–23541. doi: 10.1073/pnas.1902308116

Hiwatashi, K., Ueno, S., Sakoda, M., Iino, S., Minami, K., Yamasaki, Y., et al. (2015). Problems of long survival following surgery in patients with NonBNonC-HCC: comparison with HBV and HCV related-HCC. *J. Cancer* 6, 438–447. doi: 10.7150/jca.10398

Hu, H., Zhu, W., Qin, J., Chen, M., Gong, L., Li, L., et al. (2017). Acetylation of PGK1 promotes liver cancer cell proliferation and tumorigenesis. *Hepatology* 65, 515–528. doi: 10.1002/hep.28887

Hu, Y., Zheng, Y., Dai, M., Wang, X., Wu, J., Yu, B., et al. (2019). G9a and histone deacetylases are crucial for Snail2-mediated E-cadherin repression and metastasis in hepatocellular carcinoma. *Cancer Sci.* 110, 3442–3452. doi: 10.1111/cas.14173

Inagaki, Y., Shiraki, K., Sugimoto, K., Yada, T., Tameda, M., Ogura, S., et al. (2016). Epigenetic regulation of proliferation and invasion in hepatocellular carcinoma cells by CBP/p300 histone acetyltransferase activity. *Int. J. Oncol.* 48, 533–540. doi: 10.3892/ijo.2015.3288

Infantino, V., Dituri, F., Convertini, P., Santarsiero, A., Palmieri, F., Todisco, S., et al. (2019). Epigenetic upregulation and functional role of the mitochondrial aspartate/glutamate carrier isoform 1 in hepatocellular carcinoma. *Biochim. Biophys. Acta Mol. Basis Dis.* 1865, 38–47. doi: 10.1016/j.bbadis.2018.10.018

Jin, Q., Yu, L. R., Wang, L., Zhang, Z., Kasper, L. H., Lee, J. E., et al. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J.* 30, 249–262. doi: 10.1038/emboj.2010.318

Kitamura, K., Hatano, E., Higashi, T., Narita, M., Seo, S., Nakamoto, Y., et al. (2011). Proliferative activity in hepatocellular carcinoma is closely correlated with glucose metabolism but not angiogenesis. *J. Hepatol.* 55, 846–857. doi: 10.1016/j.jhep.2011.01.038

Lee, I. C., Ho, X. Y., George, S. E., Goh, C. W., Sundaram, J. R., Pang, K. K. L., et al. (2018). Oxidative stress promotes SIRT1 recruitment to the GADD34/PP1alpha complex to activate its deacetylase function. *Cell Death Differ.* 25, 255–267. doi: 10.1038/cdd.2017.152

Leo, L. D., Vegliante, R., Ciccarone, F., Salvatori, I., Scimeca, M., Bonanno, E., et al. (2019). Forcing ATGL expression in hepatocarcinoma cells imposes glycolytic rewiring through PPAR-alpha/p300-mediated acetylation of p53. *Oncogene* 38, 1860–1875. doi: 10.1038/s41388-018-0545-0

Li, M., Luo, R. Z., Chen, J. W., Cao, Y., Lu, J. B., He, J. H., et al. (2011). High expression of transcriptional coactivator p300 correlates with aggressive features and poor prognosis of hepatocellular carcinoma. *J. Transl. Med.* 9:5. doi: 10.1186/1479-5876-9-5

Lin, H. P., Cheng, Z. L., He, R. Y., Song, L., Tian, M. X., Zhou, L. S., et al. (2016). Destabilization of fatty acid synthase by acetylation inhibits de novo lipogenesis and tumor cell growth. *Cancer Res.* 76, 6924–6936. doi: 10.1158/0008-5472.CAN-16-1597

Lundby, A., Lage, K., Weinert, B. T., Bekker-Jensen, D. B., Secher, A., Skovgaard, T., et al. (2012). Proteomic analysis of lysine acetylation sites in rat tissues reveals organ specificity and subcellular patterns. *Cell Rep.* 2, 419–431. doi: 10.1016/j.celrep.2012.07.006

Ma-On, C., Sanpavat, A., Whongsiri, P., Suwannasin, S., Hirankarn, N., Tangkijvanich, P., et al. (2017). Oxidative stress indicated by elevated expression of Nrf2 and 8-OHdG promotes hepatocellular carcinoma progression. *Med. Oncol.* 34:57. doi: 10.1007/s12032-017-0914-5

Marques, C., Guo, W. M., Pereira, P., Taylor, A., Patterson, C., Evans, P. C., et al. (2006). The triage of damaged proteins: degradation by the ubiquitin-proteasome pathway or repair by molecular chaperones. *FASEB J.* 20, 741–743. doi: 10.1096/fj.05-5080fje

Marrero, J. A., Feng, Z., Wang, Y., Nguyen, M. H., Befeler, A. S., Roberts, L. R., et al. (2009). Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology* 137, 110–118. doi: 10.1053/j.gastro.2009.04.005

Martile, M. D., Bufalo, D. D., and Trisciuoglio, D. (2016). The multifaceted role of lysine acetylation in cancer: prognostic biomarker and therapeutic target. *Oncotarget* 7, 55789–55810. doi: 10.18632/oncotarget.10048

Mashek, D. G., Khan, S. A., Sathyanarayan, A., Ploeger, J. M., and Franklin, M. P. (2015). Hepatic lipid droplet biology: getting to the root of fatty liver. *Hepatology* 62, 964–967. doi: 10.1002/hep.27839

Min, Z. Y., Gao, J. M., and Yu, Y. (2018). The roles of mitochondrial SIRT4 in cellular metabolism. *Front. Endocrinol.* 9:783. doi: 10.3389/fendo.2018.00783

Quint, K., Agaimy, A., Di Fazio, P., Montalbano, R., Steindorf, C., Jung, R., et al. (2011). Clinical significance of histone deacetylases 1, 2, 3, and 7: HDAC2 is an independent predictor of survival in HCC. *Virchows Arch.* 459, 129–139. doi: 10.1007/s00428-011-1103-0

Sun, B. Y., Ranish, J. A., Utleg, A. G., White, J. T., Yan, X. W., Lin, B. Y., et al. (2006). Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics. *Mol. Cell Proteomics* 6, 141–149. doi: 10.1074/mcp.T600046-MCP200

Sun, Y. L., Mi, W., Cai, J. Q., Ying, W. T., Liu, F., Lu, H. Z., et al. (2008). Quantitative proteomic signature of liver cancer cells: tissue transglutaminase 2 could be a novel protein candidate of human hepatocellular carcinoma. *J. Proteome Res.* 7, 3847–3859. doi: 10.1021/pr800153s

Tang, J. C., Feng, Y. L., Guo, T., Xie, A. Y., and Cai, X. J. (2016). Circulating tumor DNA in hepatocellular carcinoma: trends and challenges. *Cell Biosci.* 6:32. doi: 10.1186/s13578-016-0100-z

Thangjam, G. S., Birmpas, C., Barabutis, N., Gregory, B. W., Clemens, M. A., Newton, J. R., et al. (2016). Hsp90 inhibition suppresses NF-κB transcriptional activation via Sirt-2 in human lung microvascular endothelial cells. *Am. J. Physiol. Lung. Cell Mol. Physiol.* 310, L964–L974. doi: 10.1152/ajplung.00054.2016

Wan, J., Liu, H., and Ming, L. (2019). Lysine crotonylation is involved in hepatocellular carcinoma progression. *Biomed. Pharmacother.* 111, 976–982. doi: 10.1016/j.biopha.2018.12.148

Wang, B., Ye, Y., Yang, X., Liu, B., Wang, Z., Chen, S. Y., et al. (2020). SIRT2-dependent IDH1 deacetylation inhibits colorectal cancer and liver metastases. *EMBO Rep.* 21:e48183. doi: 10.15252/embr.201948183

Wang, L. T., Liu, K. Y., Jeng, W. Y., Chiang, C. M., Chai, C. Y., Chiou, S. S., et al. (2020). PCAF-mediated acetylation of ISX recruits BRD4 to promote epithelial-mesenchymal transition. *EMBO Rep.* 21:e48795. doi: 10.15252/embr.201948795

Wang, M., Sanda, M., Comunale, M. A., Herrera, H., Swindell, C., Kono, Y., et al. (2017). Changes in the glycosylation of kininogen and the development of a kininogen-based algorithm for the early detection of HCC. *Cancer Epidemiol. Biomarkers Prev.* 26, 795–803. doi: 10.1158/1055-9965.EPI-16-0974

Watanabe, H., Inaba, Y., Kimura, K., Matsumoto, M., Kaneko, S., Kasuga, M., et al. (2018). Sirt2 facilitates hepatic glucose uptake by deacetylating glucokinase regulatory protein. *Nat. Commun.* 9:30. doi: 10.1038/s41467-017-02537-6

Weems, J., and Olson, A. L. (2011). Class II histone deacetylases limit GLUT4 gene expression during adipocyte differentiation. *J. Biol. Chem.* 286, 460–468. doi: 10.1074/jbc.M110.157107

Xu, Y. P., Li, F., Lv, L., Li, T., Zhou, X., Deng, C. X., et al. (2014). Oxidative stress activates SIRT2 to deacetylate and stimulate phosphoglycerate mutase. *Cancer Res.* 74, 3630–3642. doi: 10.1158/0008-5472.CAN-13-3615

Zhang, D., Tang, Z., Huang, H., Zhou, G., Cui, C., Weng, Y., et al. (2019). Metabolic regulation of gene expression by histone lactylation. *Nature* 574, 575–580. doi: 10.1038/s41586-019-1678-1

Zhao, S., Xu, W., Jiang, W., Yu, W., Lin, Y., Zhang, T., et al. (2010). Regulation of cellular metabolism by protein lysine acetylation. *Science* 327, 1000–1004. doi: 10.1126/science.1179689

Zhong, J. L., and Huang, C. Z. (2016). Ubiquitin proteasome system research in gastrointestinal cancer. *World J. Gastrointest. Oncol.* 8, 198–206. doi: 10.4251/wjgo.v8.i2.198

Zhou, J., Yu, L., Gao, X., Hu, J., Wang, J., Dai, Z., et al. (2011). Plasma microRNA panel to diagnose hepatitis B virus-related hepatocellular carcinoma. *J. Clin. Oncol.* 29, 4781–4788. doi: 10.1200/JCO.2011.38.2697

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 26 October 2020
doi: 10.3389/fgene.2020.593273

KPNA2-Associated Immune Analyses Highlight the Dysregulation and Prognostic Effects of GRB2, NRAS, and Their RNA-Binding Proteins in Hepatocellular Carcinoma

Xiuzhi Zhang[1], Jialing Zhang[1], Fenglan Gao[1], Shasha Fan[2,3], Liping Dai[4]* and Jinzhong Zhang[1]**

[1]*Department of Pathology, Henan Medical College, Zhengzhou, China,* [2]*Oncology Department, The First Affiliated Hospital of Hunan Normal University, Hunan Provincial People's Hospital, Changsha, China,* [3]*Key Laboratory of Study and Discovery of Small Targeted Molecules of Hunan Province, Hunan Normal University, Changsha, China,* [4]*Henan Institute of Medical and Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, China*

OPEN ACCESS

Edited by:
Xiaofeng Dai,
Jiangnan University, China

Reviewed by:
Xiangqian Guo,
Henan University, China
Pingping Chen,
University of Miami, United States
Jing Xie,
The University of Melbourne, Australia

***Correspondence:**
Shasha Fan
shashafan0225@hunnu.edu.cn
Liping Dai
lpdai@zzu.edu.cn
Jinzhong Zhang
jzzhanghnyz@126.com

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *10 August 2020*
Accepted: *21 September 2020*
Published: *26 October 2020*

Citation:
Zhang X, Zhang J, Gao F, Fan S, Dai L and Zhang J (2020) KPNA2-Associated Immune Analyses Highlight the Dysregulation and Prognostic Effects of GRB2, NRAS, and Their RNA-Binding Proteins in Hepatocellular Carcinoma. Front. Genet. 11:593273. doi: 10.3389/fgene.2020.593273

Karyopherin α2 (KPNA2) was reported to be overexpressed and have unfavorable prognostic effects in many malignancies including hepatocellular carcinoma (HCC). Although its contributions to inflammatory response were reported in many studies, its specific associations with immune infiltrations and immune pathways during cancer progression were unclear. Here, we aimed to identify new markers for HCC diagnosis and prognosis through KPNA2-associated immune analyses. RNA-seq expression data of HCC datasets were downloaded from The Cancer Genome Atlas and International Cancer Genome Consortium. The gene expressions were counts per million normalized. The infiltrations of 24 kinds of immune cells in the samples were evaluated with ImmuCellAI (Immune Cell Abundance Identifier). The Spearman correlations of the immune infiltrations with KPNA2 expression were investigated, and the specific positive correlation of B-cell infiltration with KPNA2 expression in HCC tumors was identified. Fifteen genes in KEGG (Kyoto Encyclopedia of Genes and Genomes) B-cell receptor signaling pathway presented significant correlations with KPNA2 expression in HCC. Among them, GRB2 and NRAS were indicated to be independent unfavorable prognostic factors for HCC overall survival. Clinical Proteomic Tumor Analysis Consortium HCC dataset was investigated to validate the results at protein level. The upregulation and unfavorable prognostic effects of KPNA2 and GRB2 were confirmed, whereas, unlike its mRNA form, NRAS protein was presented to be downregulated and have favorable prognostic effects. Through receiver operating characteristic curve analysis, the diagnostic potential of the three proteins was shown. The RNA-binding proteins (RBPs) of KPNA2, NRAS, and GRB2, downloaded *via* The Encyclopedia of RNA Interactomes, were investigated for their clinical significance in HCC at protein level. An eight-RBP signature with independent prognostic value and dysregulations in HCC was identified. All the RBPs were significantly correlated with MKI67 expression and at least one of KPNA2, GRB2, and NRAS at protein level in HCC, indicating

their roles in HCC progression and the regulation of the three proteins. We concluded that KPNA2, GRB2, NRAS, and their RBPs might have coordinating roles in HCC immunoregulation and progression. They might be new markers for HCC diagnosis and prognosis predication and new targets for HCC immunotherapy.

Keywords: hepatocellular carcinoma, immune infiltration, KPNA2, GRB2, NRAS, prognosis, RNA-binding protein

INTRODUCTION

Liver cancer is the seventh most common cancer and the third leading cause of cancer death worldwide (Bray et al., 2018). Hepatocellular carcinoma (HCC) is the predominant type of primary liver malignancies, with a 5-year overall survival (OS) rate of less than 20%, mainly due to its late diagnosis. Although great advancement has been made in HCC screening and treatment during the last few decades, the effectiveness still remains unsatisfactory. It was reported that for patients with early-stage HCC, the 5-year survival rate was greater than 70% (Siegel et al., 2018), much better than that of late-stage patients, calling for effective markers of its early diagnosis and prognostic predication.

Karyopherin α2 (KPNA2), also named importin α1, is a member of karyopherin α family and plays crucial roles in nucleocytoplasmic transport (Goldfarb et al., 2004). In rheumatoid arthritis, KPNA2 was shown to be a trigger of interleukin 6 (IL-6) secretion, colocalized with T cells and neutrophils, and could be upregulated *via* tumor necrosis factor *α* stimulation (Liu et al., 2015). It was also demonstrated to play crucial roles in the negative regulation of regulatory T cells (Tregs) differentiation through its interaction and translocation of proinflammatory molecule NLRP3 (Park et al., 2019). In a study of vaccinia virus (VACV), the positive effects of KPNA2 on immunoregulation was shown in the induction of VACV specific $CD8^+$ T-cell memory *via* its interaction with p65 (Pallett et al., 2019). In another study, KPNA2 downregulation was reported to be associated with enterovirus 71-induced innate immune response (Peng et al., 2018), indicating its negative immunoregulatory roles.

In recent years, The dysregulation and/or prognostic roles of KPNA2 were reported in many tumors including esophageal squamous cell carcinoma (Song et al., 2019), breast cancer (Groheux et al., 2018; Wang et al., 2019a), myeloma (Kriegsmann et al., 2019; Tachita et al., 2020), clear cell renal cell carcinoma (Müller et al., 2019), oral cancer (Wang et al., 2018), bladder cancer (Jeong et al., 2017), and colorectal cancer (Ostasiewicz et al., 2016; Jeong et al., 2017). The immune-related roles of KPNA2 were also shown in tumors. In colorectal cancer cells, upregulated KPNA2 was demonstrated to be associated with immunogenic cell death (Song et al., 2016). In breast cancer, KPNA2 knockdown could suppress the inflammatory responses and malignant progression of the tumor cells induced by IL-6 (Duan et al., 2020). In HCC, KPNA2 overexpression and its prognostic effects were also reported in several studies (Jiang et al., 2014; Yang et al., 2017b; Chen et al., 2019; Guo et al., 2019; Liu et al., 2019; Yue et al., 2019; Yu et al., 2020). In fact, its tumor-promoting activities were also shown (Gao et al., 2018; Lin et al., 2018; Zan et al., 2019) in HCC. However, although the correlations between KPNA2 expression and immune infiltrations were investigated in a recent study (Hua et al., 2020), its associations with immune response in HCC were not clearly illustrated. Furthermore, with only one dataset and only six kinds of immune cells included, the reliability of the results was limited.

Considering the close relationship between inflammation and HCC (El-Serag et al., 2008), we speculated that KPNA2 upregulation might participate in or associate with specific immune pathways during HCC progression. In this study, we evaluated the correlations between KPNA2 expression and infiltrations of 24 kinds of immune cells in HCC tumors and normal liver tissues from The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) HCC datasets to identify the specific correlations between KPNA2 expression and immune infiltration in HCC tumors in contrast to those in normal liver tissues. The Kyoto Encyclopedia of Genes and Genomes (KEGG) B-cell receptor (BCR) signaling pathway genes were analyzed and validated for their correlations with KPA2 expression and their independent prognostic effects in HCC. Considering the crucial roles of RNA-binding proteins (RBPs) in RNA splicing, RNA translation, and RNA degradation (Mohibi et al., 2019), the RBPs of KPNA2 and its correlated BCR signaling pathway genes with prognostic values were further investigated to find their roles in the dysregulations of the identified genes/proteins and their diagnostic value and prognostic effects in HCC. The results here might provide new clues for the roles of KPNA2 in immunoregulation, new markers for HCC early diagnosis and prognostic predication, and new targets for HCC immunotherapy.

MATERIALS AND METHODS

Data Collection

RNA-seq data of HCC tumors and normal liver tissues in HCC datasets with the clinical information were downloaded from Genomic Data Commons data portal (TCGA-LIHC dataset, called TCGA-HCC in this study) and ICGC database (ICGC-HCC dataset). There were 371 primary tumors and 50 normal liver tissues from 371 HCC patients in the TCGA-HCC dataset and 223 primary tumors and 202 normal tissues from 223 HCC patients in the ICGC-HCC dataset. The clinical features of the patients are shown in **Table 1**. The gene expression read count data of the two datasets were counts per million (CPM) transformed for normalization for further analyses.

TABLE 1 | Clinical features of the patients in the TCGA-HCC dataset (n = 371) and ICGC-HCC (n = 223) dataset.

Variables	TCGA-HCC (n = 371), n (%)	ICGC-HCC (n = 223), n (%)
Gender		
Male	250 (67.4%)	164 (73.5%)
Female	121 (32.6%)	59 (26.5%)
Age		
≤60	177 (47.7%)	47 (21.1%)
>60	193 (52.0%)	176 (78.9%)
NA	1 (0.3%)	0 (0%)
TNM stage		
I	171 (46.1%)	36 (15.7%)
II	86 (23.2%)	102 (45.7%)
III	85 (22.9%)	67 (30.5%)
IV	5 (1.3%)	18 (8.1%)
NA	24 (6.5%)	0 (0%)
Grade		
G1	55 (14.8%)	20 (9.0%)
G2	177 (47.7%)	128 (57.4%)
G3	122 (32.9%)	56 (45.5%)
G4	12 (3.2%)	1 (0.4%)
NA	5 (1.3%)	18 (8.1%)
Survival status		
Alive	240 (64.7%)	181 (81.2%)
Dead	130 (35.0%)	42 (18.8%)
NA	1 (0.3%)	0 (0%)

HCC, hepatocellular carcinoma; ICGC, International Cancer Genome Consortium; NA, not available; TCGA, The Cancer Genome Atlas.

Correlation Analyses Between KPNA2 Expression and Immune Cell Infiltrations and the Investigation of Prognostic Roles of B-Cell Infiltration in HCC

With their gene expression data, the infiltrations of 24 kinds of immune cells including 18 T-cell subsets [$CD4^+$, $CD8^+$, $CD4^+$ naive, $CD8^+$ naive, central memory T (Tcm), effector memory T, Tr1, iTreg, nTreg, T_H1, T_H2, T_H17, Tfh, cytotoxic T, MAIT, exhausted T (Tex), gamma delta T (γδ T), and natural killer T (NKT) cells] and six other important immune cells [B cells, macrophages, monocytes, neutrophils, dendritic cell (DC), and NK cells] in the HCC tumors and normal liver tissues from the TCGA-HCC and ICGC-HCC datasets were evaluated with ImmuCellAI (Immune Cell Abundance Identifier; Miao et al., 2020). The correlations between KPNA2 expression and the immune cell infiltrations of the HCC tumors and normal liver tissues were evaluated individually. Spearman correlation analysis was used, and a |correlation coefficient| > 0.15 with $p < 0.01$ was considered as statistically significant.

Kaplan-Meier survival analysis was used to evaluate the prognostic effects of B-cell infiltration on HCC OS with a cutoff value of B-cell infiltration identified from survminer package in R.[1] To investigate the independent prognostic effects of B-cell infiltration HCC OS, multivariable Cox regression analysis with ezcox package (Wang et al., 2019b) in R was performed and the gender-, age-, and stage-corrected prognostic effects of B-cell infiltration were evaluated in the TCGA-HCC and ICGC-HCC datasets individually. For survival analyses, only the patients with survival time >0 were included.

Identification of KPNA2-Correlated BCR Signaling Pathway Genes

KEGG BCR signaling pathway (hsa04662) genes (n = 75) were investigated through KEGG database.[2] Their expressions were analyzed for their correlations with KPNA2 expression in HCC samples form TCGA-HCC and ICGC-HCC datasets. Spearman correlation analysis was used, and a |correlation coefficient| > 0.15 with $p < 0.01$ was considered as statistically significant.

Differential Expression Analyses and Prognostic Effects Evaluation of the KPNA2-Correlated BCR Signaling Pathway Genes in HCC

The BCR signaling pathway genes with significant correlations with KPNA2 expression were selected, and their expression profiles and prognostic roles were investigated in the TCGA-HCC and ICGC-HCC datasets. Gene expression comparisons between HCC tumors and normal liver tissues were performed with Wilcoxon test. Univariable Cox regression analysis was performed to evaluate the prognostic effects of the genes. The independent prognostic values were investigated through multivariable Cox regulation analysis with ezcox package (Wang et al., 2019b) in R software with the gene expressions as covariates and gender, age, and stage as controls. Only the BCR signaling pathway genes with prognostic effects independent of gender, age, and stage were selected for further analyses. For the analyses, $p < 0.05$ was considered significant.

Validation of the Dysregulation and Prognostic Effects of KPNA2, GRB2, and NRAS at Protein Level

To validate the results above, the HCC proteomic data of a Chinese cohort from Clinical Proteomic Tumor Analysis Consortium (CPTAC; Gao et al., 2019) was investigated. The tumor samples (n = 159) and their paired liver tissues (n = 159) from 159 HCC patients were included. The clinical information of the patients is shown in **Supplementary Table S1**. The expressional differences of the proteins were evaluated with paired-samples Wilcoxon test. The diagnostic power of the proteins was evaluated through receiver operating characteristic (ROC) curve analysis with pROC package (Robin et al., 2011) in R. The area under the curve (AUC) and the sensitivity and specificity at the cutoff point with the biggest Youden index (sensitivity + specificity − 1) were evaluated. The correlations of the protein expressions were investigated through Spearman correlation analysis. The prognostic effects of KPNA2, GRB2, and NRAS on HCC OS were investigated through Kaplan-Meier survival analysis for which survminer package in R was used to get the cutoff value for grouping the patients.[3]

[1]https://CRAN.R-project.org/package=survminer

[2]https://www.genome.jp/kegg/

[3]https://CRAN.R-project.org/package=survminer

Furthermore, ezcox package (Wang et al., 2019b) in R was used to evaluate the age-, gender-, and tumor size-corrected prognostic effects of the proteins. For all the analyses, $p < 0.05$ as statistically significant.

Exploration of the Potential Roles of KPNA2, GRB2, and NRAS in HCC Proliferation

To explore the associations of KPNA2, GRB2, and NRAS with HCC proliferation, these correlations with cell proliferation MKI67 (Gerdes et al., 1983; Scholzen and Gerdes, 2000; Booth et al., 2014) were evaluated in HCC tumor samples through Spearman correlation analysis, and $p < 0.05$ was considered statistically significant.

Further Investigation of the Prognostic Effects and Dysregulation of RBPs of KPNA2, GRB2, and NRAS in HCC

RBP binding at specific target sites could impact the expression of functionally coordinated sets of mRNAs and regulate their function in the cell (Sternburg and Karginov, 2020). Here, through The Encyclopedia of RNA Interactomes (ENCORI),[4] the RBPs of KPNA2 and its correlated BCR signaling pathway genes (GRB2 and NRAS) were also investigated. The gender-, age-, and tumor size-corrected prognostic effects of the RBPs were analyzed through multivariable Cox regression analysis with ezcox package (Wang et al., 2019b) in R software. The RBPs with prognostic effects independent of gender, age, and tumor size, as well as KPNA2, GRB2, and NRAS, were then applied to the least absolute shrinkage and selection operator (LASSO) Cox regression to get an effective risk model for the prognostic predication of HCC. The risk score of each patient was evaluated with the formula as follows:

$$\text{risk score} = \sum_{i=1}^{n} \text{coefficient}(i) * \text{expression}(i)$$

where n is the number of selected proteins, coefficient(i) is the coefficient of the protein i, and expression(i) is the expression level of protein i. The accuracy of the model in the survival status predication of HCC patients was investigated through ROC curve analysis, and the AUC was evaluated. Then, the HCC patients were than divided into high- and low-risk groups with the median risk score as the threshold. Through Kaplan-Meier survival analysis, the survival differences between high- and low-risk patients were visualized. The independent prognostic value of the model was also evaluated through Kaplan-Meier survival analysis in the high- and low-risk patients of different tumor size groups, different gender groups, and different age groups individually.

For the RBPs in the model above, their expressional differences between HCC tumors and their paired normal liver tissues were investigated through paired-samples Wilcoxon test, and their diagnostic value was investigated through ROC curve analysis with pROC package (Robin et al., 2011) in R. To further investigate their potential roles in the dysregulation of KPNA2, GRB2, and NRAS, their Spearman correlations in HCC were also evaluated. Furthermore, to investigate their associations with HCC proliferation, their correlations with proliferation marker MKI67 were evaluated through spearman correlation analyses. For these analyses, $p < 0.05$ was considered significant.

[4]http://starbase.sysu.edu.cn/index.php

RESULTS

Positive Correlation of B-Cell Infiltration With KPNA2 Expression and Its Prognostic Roles in HCC

For the TCGA-HCC dataset, as shown in **Supplementary Figure S1**, 10 of the immune cell infiltrations were shown to be negatively ($R < -0.15$ and $p < 0.01$: CD4$^+$ naive, T_H17, Tfh, MAIT, NK, and CD4$^+$ T infiltrations) or positively correlated ($R > 0.15$ and $p < 0.01$: Tex, nTreg, DC, and B-cell infiltrations) with KPNA2 expression in HCC tumors in the TCGA-HCC datasets. In the normal liver tissues of the TCGA-HCC dataset, the negative correlations ($R < -0.15$ and $p < 0.01$) of CD4$^+$ naive and B-cell and the positive correlation ($R > 0.15$ and $p < 0.01$) of DC infiltration with KPNA2 expression were shown. Interestingly, B-cell infiltration was shown to be positively correlated ($R = 0.397$, $p < 0.01$) with KPNA2 expression in HCC tumors while negatively correlated ($R = -0.416$, $p < 0.01$) with KPNA2 expression in the normal liver tissues (**Figure 1A**).

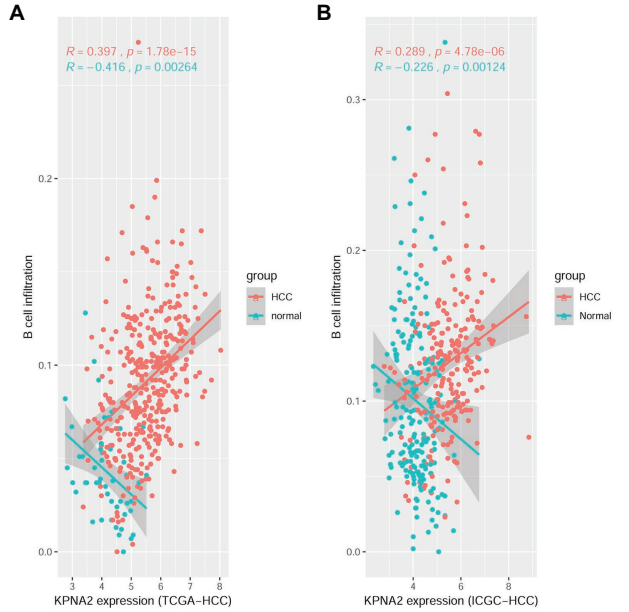

FIGURE 1 | Spearman correlation analysis between KPNA2 expression and B-cell immune infiltrations in HCC. **(A)** KPNA2 expression was positively correlated with B-cell immune infiltration in HCC tumors while negatively correlated with B-cell infiltration in normal liver tissues in the TCGA-HCC dataset. **(B)** KPNA2 expression was positively correlated with B-cell immune infiltration in HCC tumors while negatively correlated with B-cell infiltration in normal liver tissues in the ICGC-HCC dataset. Spearman correlation analysis was used. $|R| > 0.15$ with $p < 0.01$ was considered statistically significant. R, correlation coefficient.

For the ICGC-HCC dataset, as shown in **Supplementary Figure S2**, six kinds of immune cells including $CD4^+$ naive, Tfh, MAIT, monocyte, NK cell, and $CD4^+$ T cell were shown to be significantly negatively correlated ($R < -0.15$ and $p < 0.01$), while DC and B-cell infiltrations were positively correlated ($R > 0.15$ and $p < 0.05$) with KPNA2 expression in HCC tumors. In the normal liver tissues, negative correlations ($R < -0.15$ and $p < 0.01$) of $CD4^+$ naive, B cell, NK, γδ T, and $CD4^+$ T infiltrations, whereas positive correlations ($R > 0.15$ and $p < 0.01$) of Tcm, DC, and macrophage infiltrations with KPNA2, were shown. Noticeably, opposite correlations of B-cell infiltration with KPNA2 expression in HCC tumors ($R = 0.289$, $p < 0.01$) and the normal liver tissues ($R = -0.226$, $p < 0.01$) were also obvious here (**Figure 1B**), consistent with the results in the TCGA-HCC dataset, indicating the specificity of the positive correlation between KPNA2 expression and B-cell infiltration in the HCC tumors in contrast to the normal controls.

Through Kaplan-Meier survival analysis (**Figures 2A,B**) and multivariable Cox regression analysis (**Figures 2C,D**), the unfavorable prognostic effects of B-cell infiltration on HCC OS were shown both in the TCGA-HCC and ICGC-HCC datasets, indicating its potential in HCC prognosis predication. BCR is a master regulator of B cells. Through BCR, the B cells recognize foreign antigen, leading to maturation of the B-cell into either a memory B-cell or an effector (plasma) B-cell (Puri et al., 2013; Skånland et al., 2020). We speculated that there might be also associations between BCR signaling pathway genes and KPNA2, and they might play important roles during HCC progression.

Correlations Between KPNA2 Expression and BCR Signaling Pathway Genes in HCC

There were 27 BCR signaling pathway genes negatively or positively correlated with KPNA2 at mRNA level in HCC tumors of the TCGA-HCC and ICGC-HCC datasets individually (**Supplementary Table S2**). Among them, seven genes including GRB2, NRAS, NFKBIE, MAPK3, BCL10, NFATC2, and PIK3R2 were positively correlated ($R > 0.15$ and $p < 0.01$), whereas eight genes including BLNK, IFITM1, AKT3, FOS, AKT1, NFKBIA, CD81, and PLCG2 were negatively correlated ($R < -0.15$ and $p < 0.01$) with KPNA2 expression in HCC in both TCGA-HCC dataset (**Figure 3A**) and ICGC-HCC dataset (**Figure 3B**).

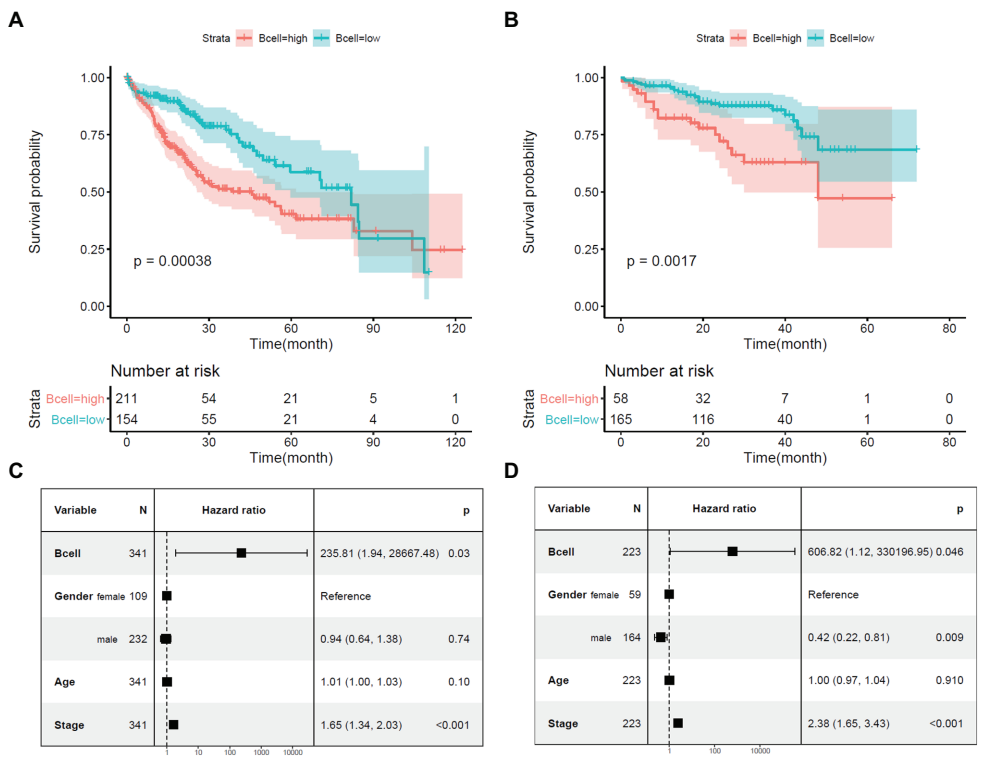

FIGURE 2 | Prognostic effects of B-cell infiltration in HCC. **(A)** Significant shorter survival in the HCC patients with high B-cell infiltration than the patients with low B-cell infiltration in the TCGA-HCC dataset. **(B)** Significant shorter survival in the HCC patients with high B-cell infiltration than the patients with low B-cell infiltration in the ICGC-HCC dataset. **(C)** Prognostic effects of B-cell infiltration on HCC overall survival when adjusted for gender, age, and stage in the TCGA-HCC dataset. **(D)** Prognostic effects of B-cell infiltration on HCC overall survival when adjusted for gender, age, and stage in the ICGC-HCC dataset. Kaplan-Meier survival analysis with log-rank test and multivariable Cox regression analysis were used for evaluation of the prognostic effects of B-cell infiltration. For all the analyses, $p < 0.05$ was considered statistically significant.

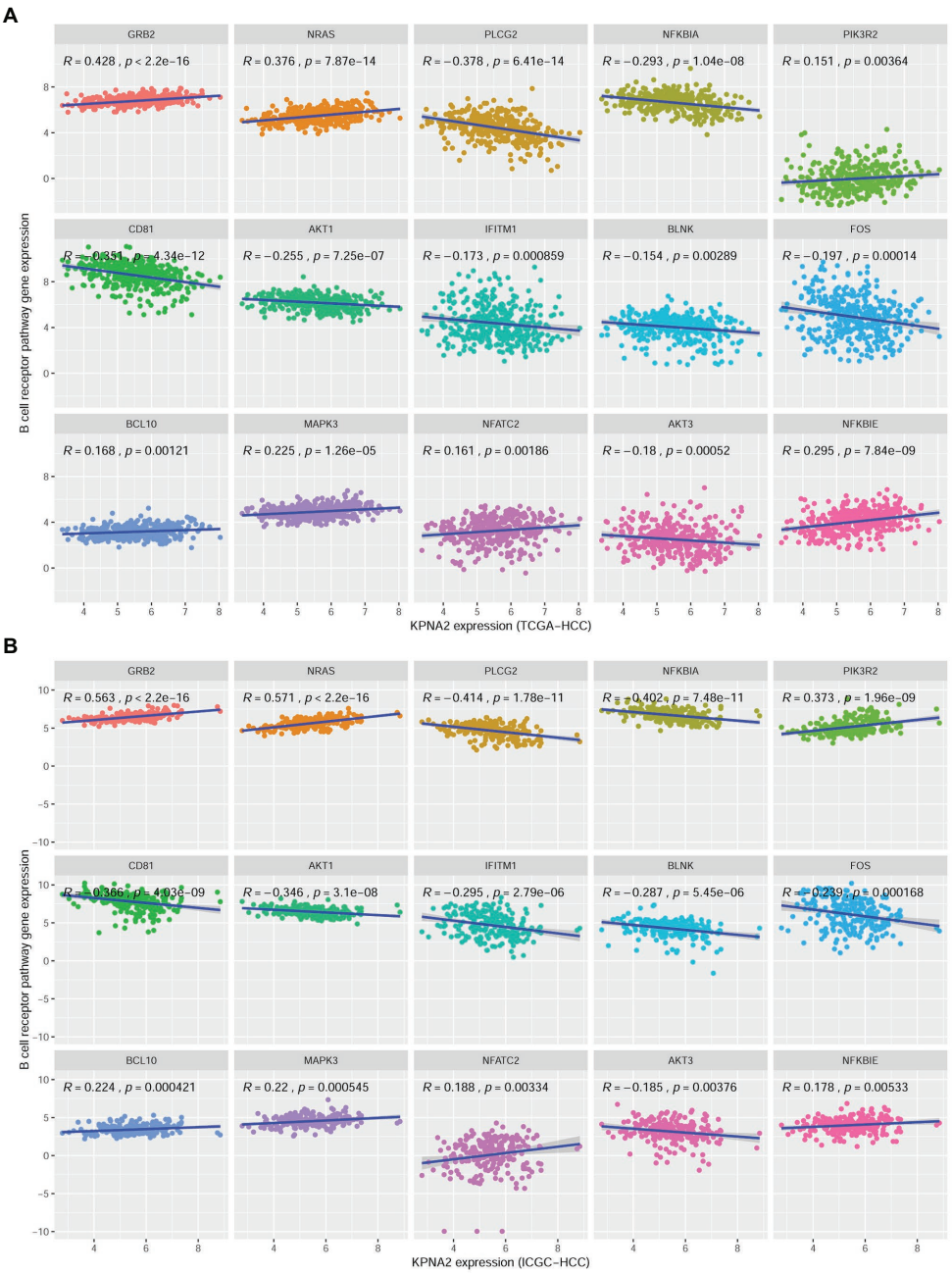

FIGURE 3 | Significant correlations between BCR signaling pathway gene expressions and KPNA2 expression in HCC. **(A)** Significant positive correlations of GRB2, NRAS, NFKBIE, MAPK3, BCL10, NFATC2, and PIK3R2 and negative correlations of BLNK, IFITM1, AKT3, FOS, AKT1, NFKBIA, CD81, and PLCG2 with KPNA2 expression in HCC in the TCGA-HCC dataset. **(B)** Significant positive correlations of GRB2, NRAS, NFKBIE, MAPK3, BCL10, NFATC2, and PIK3R2 and negative correlations of BLNK, IFITM1, AKT3, FOS, AKT1, NFKBIA, CD81, and PLCG2 with KPNA2 expression in HCC in the ICGC-HCC dataset. The *x*-axis and *y*-axis represented the relative expression of KPNA2 and BCR signaling pathway genes, respectively. The gene expressions were CPM normalized and $\log_2(x + 0.001)$ transformed. BCR, B-cell receptor; *R*, correlation coefficient; CPM, count per million. Spearman correlation was used, and $|R| \geq 0.15$ with $p < 0.01$ was considered significant.

Differential Expression and Prognostic Effects of KPNA2 and Its Correlated BCR Signaling Pathway Genes in HCC

As shown in **Figure 4**, besides the overexpression of KPNA2 in HCC tumors, among the seven-KPNA2 positively correlated BCR signaling pathway genes, six genes including GRB2, NRAS, NFKBIE, MAPK3, and NFATC2 were shown to be higher expressed, whereas NFATC2 was lower expressed in the HCC tumors than the normal liver tissues ($p < 0.05$) in both TCGA-HCC dataset (**Figure 4A**) and ICGC-HCC dataset (**Figure 4B**).

However, BCL10 was shown to be downregulated in the TCGA-HCC tumors ($p < 0.05$) but not statistically significant in the ICGC-HCC tumors ($p > 0.05$). For the eight-KPNA2 negatively correlated BCR signaling pathway genes, all of them were shown to be downregulated in HCC tumors comparing with the normal liver tissues in the two datasets ($p < 0.05$).

Through univariable Cox regression analysis (**Figures 5A,B**), besides the unfavorable prognostic effects of KPNA2 ($HR_{TCGA\text{-}HCC} = 1.823$, $HR_{ICGC\text{-}HCC} = 2.234$, $p < 0.05$), two of its positively correlated BCR signaling pathway genes (NRAS: $HR_{TCGA\text{-}HCC} = 1.775$, $HR_{ICGC\text{-}HCC} = 2.664$, $p < 0.05$; GRB2: $HR_{TCGA\text{-}HCC} = 1.901$, $HR_{ICGC\text{-}HCC} = 2.470$, $p < 0.05$) and one of its negatively correlated genes (BLNK: $HR_{TCGA\text{-}HCC} = 0.771$, $HR_{ICGC\text{-}HCC} = 0.786$, $p < 0.05$) were shown to be prognostic factors both in the TCGA-HCC patients and ICGC-HCC patients. For the other 12 genes, their prognostic effects were shown in only one of or neither of the two datasets. Then, KPNA2 and the three BCR signaling pathway genes (GRB2, NRAS, and BLNK) with consistent prognostic effects in the two datasets were selected for further investigation of their independent prognostic values. When their prognostic effects were adjusted for gender, age, and tumor stage, the unfavorable effects of KPNA2 ($HR_{TCGA\text{-}HCC} = 1.78$, $HR_{ICGC\text{-}HCC} = 2.45$, $p < 0.05$; **Figures 5C,G**), GRB2 ($HR_{TCGA\text{-}HCC} = 1.67$, $HR_{ICGC\text{-}HCC} = 1.97$, $p < 0.05$; **Figures 5D,H**), and NRAS ($HR_{TCGA\text{-}HCC} = 1.59$, $HR_{ICGC\text{-}HCC} = 2.39$, $p < 0.05$; **Figures 5E,I**) on HCC OS also existed in both TCGA-HCC and ICGC-HCC datasets, indicating their potential in the predication of HCC prognosis. However, although the independent favorable prognostic effect of BLNK in the TCGA-HCC (HR = 0.80, $p < 0.05$; **Figure 5F**) was shown, it was not so significant in the ICGC-HCC dataset (HR = 0.82, $p > 0.05$; **Figure 5J**).

Validation of Dysregulations, Correlations, and Prognostic Effects of KPNA2, GRB2, and NRAS in HCC at Protein Level

At protein level, through CPTAC-HCC analysis, KPNA2 ($p < 0.001$; **Figure 6A**) and GRB2 ($p < 0.001$; **Figure 6B**) were found to be higher expressed in HCC tumors than their normal liver tissues, consistent with their upregulation in HCC tumors at mRNA level. However, in contrast to its overexpression in HCC tumors at mRNA level, NRAS protein was shown to be lower expressed in HCC tumors than their normal liver controls ($p < 0.001$; **Figure 6C**). Upon ROC curve analysis

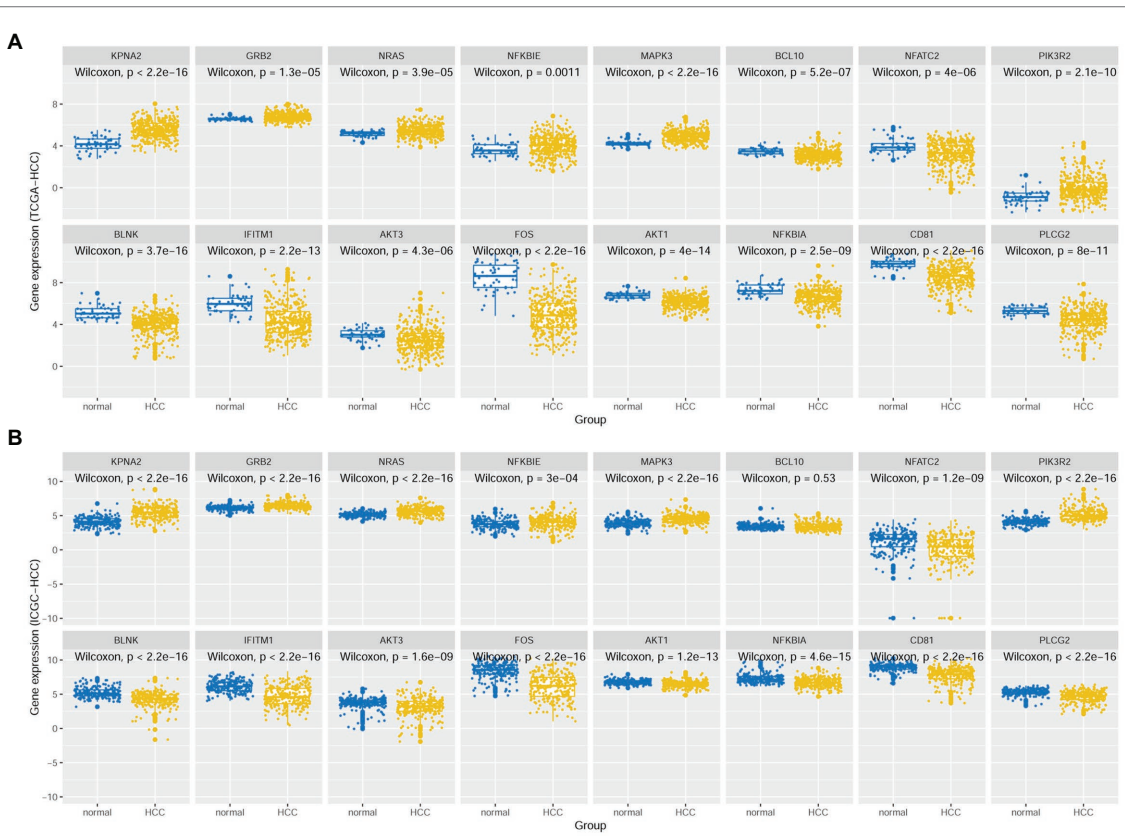

FIGURE 4 | Expressional comparisons of KPNA2 and its correlated BCR signaling pathway genes between HCC tumors and normal livers. **(A)** Expressional differences of GRB2, NRAS, NFKBIE, MAPK3, BCL10, NFATC2, PIK3R2, BLNK, IFITM1, AKT3, FOS, AKT1, NFKBIA, CD81, and PLCG2 between HCC tumors and normal livers in the TCGA-HCC dataset. **(B)** Expressional differences of GRB2, NRAS, NFKBIE, MAPK3, BCL10, NFATC2, PIK3R2, BLNK, IFITM1, AKT3, FOS, AKT1, NFKBIA, CD81, and PLCG2 between HCC tumors and normal livers in the ICGC-HCC dataset. The *x*-axis and *y*-axis represented the sample groups and the relative expressions of the genes, respectively. The gene expressions were CPM normalized and $\log_2(x + 0.001)$ transformed. BCR, B-cell receptor; CPM, count per million. Wilcoxon test was used for comparisons, and $p < 0.05$ was considered significant.

(**Figures 6D–F**), the diagnostic power of the three proteins in discriminating HCC from normal controls was shown with AUCs of 0.896, 0.740, and 0.719 for KPNA2, GRB2, and NRAS, respectively. At the optimal cutoff points, the sensitivity was 0.761, 0.553, and 0.579, and the specificity was 0.899, 0.843, and 0.868 for KPNA2, GRB2, and NRAS, respectively.

For their correlations, KPNA2 expression was presented to be positively correlated with GRB2 expression (R = 0.36, p < 0.001; **Figure 6G**) while negatively correlated with NRAS expression (R = −0.21, p < 0.01; **Figure 6H**). However, no significant expressional correlation between GRB2 and NRAS was shown (p > 0.05; **Figure 6I**).

For their prognostic effects, through Kaplan-Meier analysis, at protein level, KPNA2 (p < 0.001; **Figure 7A**) and GRB2 (p < 0.05; **Figure 7B**) were shown to be unfavorable prognostic factors for HCC OS, consistent with their unfavorable prognostic effects at mRNA level. Interestingly, for NRAS (p < 0.05; **Figure 7C**), at protein level, it was shown to have favorable effects on HCC OS, inconsistent with its unfavorable effects on HCC OS. When adjusted for gender, age, and tumor size, the prognostic effects of KPNA2 (p < 0.001; **Figure 7D**) and NRAS (p < 0.05; **Figure 7F**) also existed, whereas the independent prognostic value of GRB2 was not so significant (p = 0.074, p > 0.05; **Figure 7E**).

Associations of KPNA2, GRB2, and NRAS With HCC Proliferation

As shown in **Figure 6**, KPNA2 (R = 0.7, p < 0.001; **Figure 6J**) and GRB2 (R = 0.32, p < 0.001; **Figure 6K**) were shown to be significantly positively correlated with MKI67 expression, indicating their associations with HCC proliferation. Whereas no significant correlation was shown between NRAS expression and MKI67 expression in HCC (p > 0.05; **Figure 6L**).

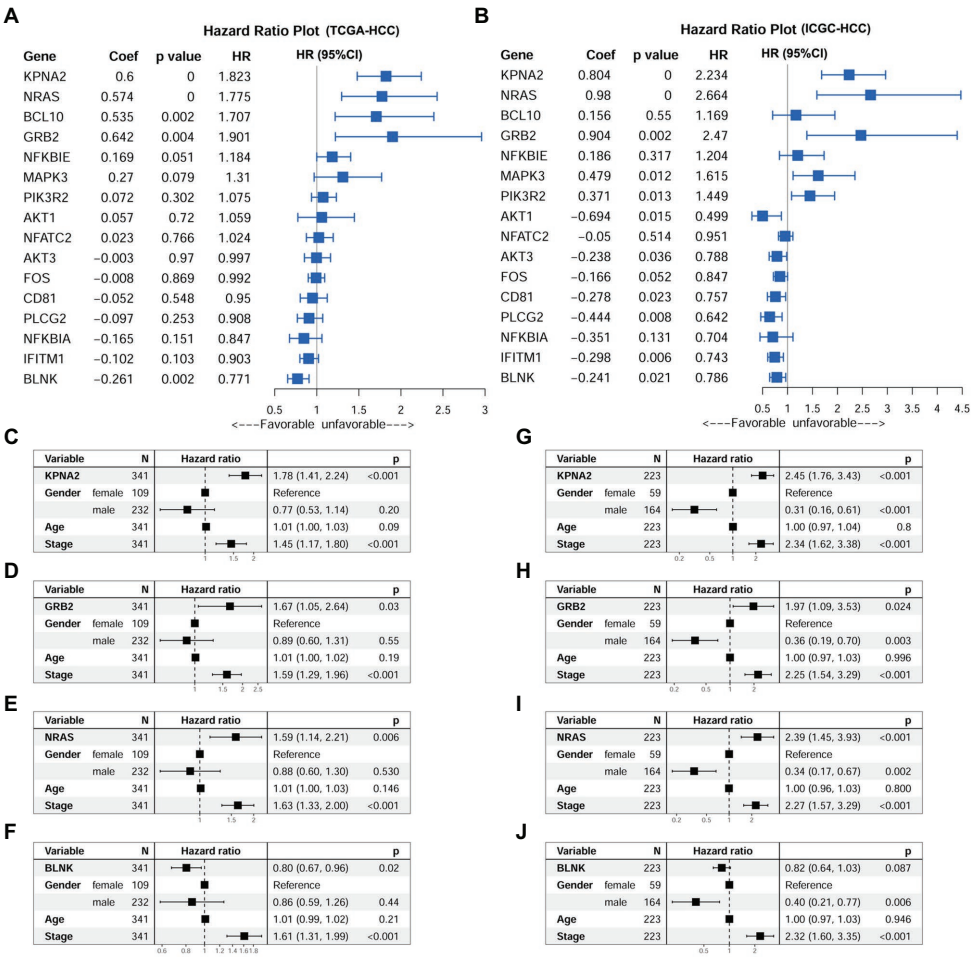

FIGURE 5 | Prognostic effects of KPNA2 and its correlated BCR signaling pathway genes in HCC. **(A)** Prognostic effects of KPNA2 and its correlated BCR signaling pathway genes on overall survival of the TCGA-HCC patients through univariable Cox regression analysis. **(B)** Prognostic effects of KPNA2 and its correlated BCR signaling pathway genes on overall survival of the ICGC-HCC patients through univariable Cox regression analysis. **(C–F)** The gender-age-stage-corrected prognostic effects of KPNA2, GRB2, NRAS, and BLNK on OS of the TCGA-HCC patients, respectively. **(G–J)** The gender-age-stage–corrected prognostic effects of KPNA2, GRB2, NRAS, and BLNK on overall survival of the ICGC-HCC patients, respectively. For **(A,B)**, univariable Cox regression analysis was used. For **(C–J)**, multivariable Cox regression analysis was performed. The analyses were performed with R software, and p < 0.05 was considered statistically significant.

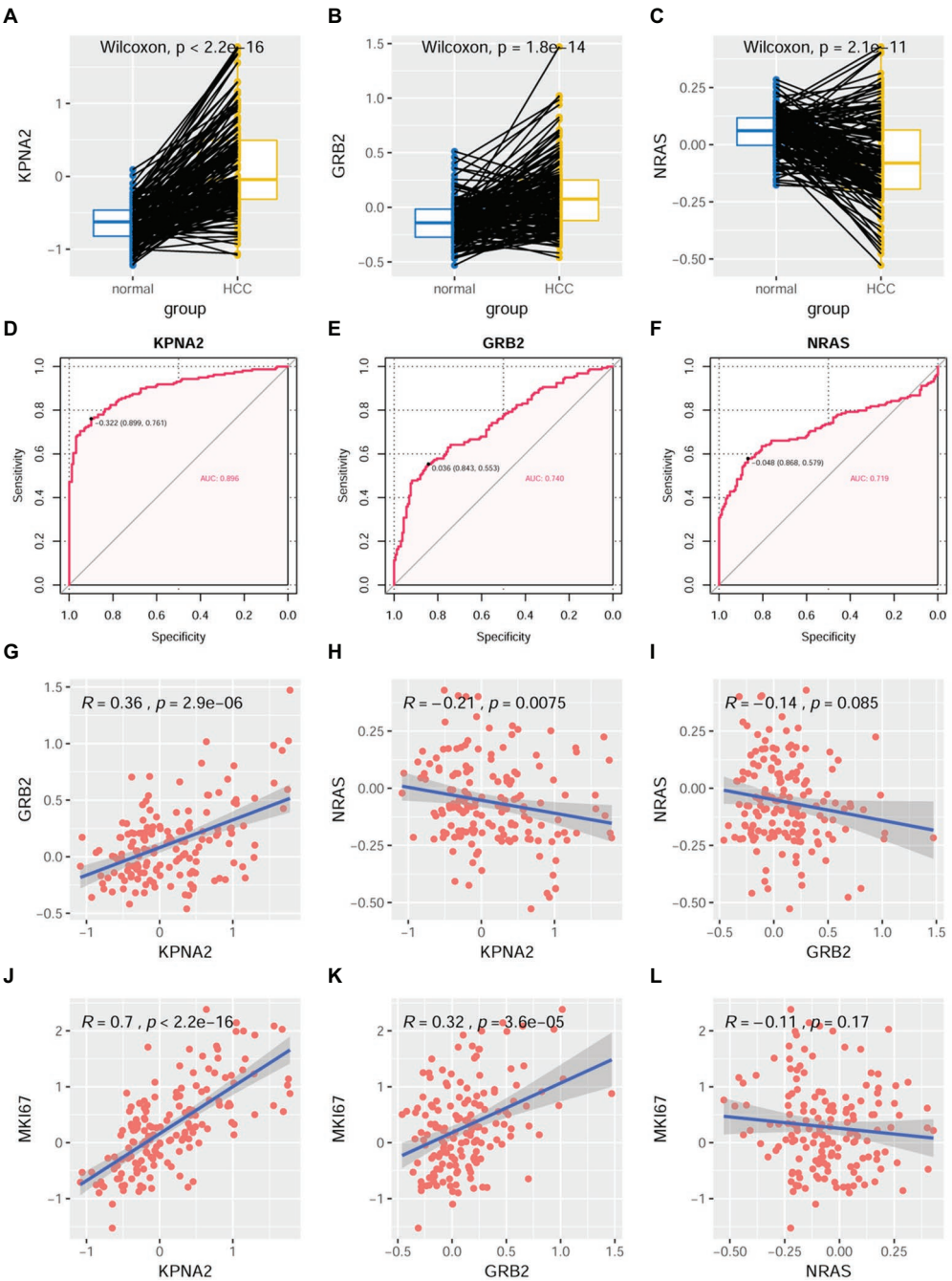

FIGURE 6 | Dysregulations of KPNA2, GRB2, and NRAS proteins and their correlations with MKI67 in HCC. **(A)** Higher expression of KPNA2 in HCC than the normal liver tissues. **(B)** Higher expression of GRB2 in HCC than the normal liver tissues. **(C)** Lower expression of NRAS in HCC than the normal liver tissues. **(D–F)** ROC curve analysis of KPNA2, GRB2, and NRAS in discriminating HCC tumors from normal liver tissues, the cutoff point with biggest Youden index was labeled, and the corresponding specificity and sensitivity were shown. **(G)** Significant positive correlation between KPNA2 expression and GRB2 expression in HCC. **(H)** Significant negative correlation between KPNA2 expression and GRB2 expression in HCC. **(I)** There was no significant correlation between GRB2 expression and NRAS expression. **(J)** Significant positive correlation between KPNA2 expression and MKI67 expression. **(K)** Significant positive correlation between GRB2 expression and MKI67 expression. **(L)** There was no significant correlation between NRAS expression and MKI67 expression. AUC, area under the curve; ROC, receiver operating characteristic. For **(A–C)**, the *x*-axis and *y*-axis represented the sample groups and the relative expressions of the proteins (the protein abundance of the samples, with respect to the pooled reference sample, as $\log_2$ ratios), respectively. For **(D–F)**, the *x*-axis and *y*-axis represented the specificity and sensitivity of the variables in discriminating HCC tumors from the normal liver tissues. For **(G–L)**, the *x*-axis and *y*-axis represented the relative expressions of the variables (the protein abundance of the samples, with respect to the pooled reference sample, as $\log_2$ ratios). Paired-samples Wilcoxon test and Spearman correlation analysis were used for expressional comparison and correlation evaluation, respectively. For the analyses, $p < 0.05$ was considered significant.

Prognostic Effects of RBPs of KPNA2, GRB2, and NRAS and Their Potential Roles in Dysregulations of the Three Proteins in HCC

As shown in **Supplementary Table S3**, there were 105, 112, and 114 RBPs that could bind to KPNA2, GRB2, and NRAS, respectively. As 94 of the RBPs were common for the three genes, there were 123 unique RBPs that could bind at least one of them. Because six RBPs were not available from the CPTAC-HCC dataset, only 117 RBPs (**Supplementary Table S3**) were investigated. Through multivariable Cox regression analysis (**Supplementary Table S4**), 16 RBPs were shown to have prognostic effects on HCC OS independent of gender, age, and tumor size in the CPTAC-HCC dataset (**Figure 8A**). Then, the 16 RBPs and KPNA2, GRB2, and NRAS were further applied to LASSO regression analysis, and an eight-RBP signature including AU RNA-binding methylglutaconyl-CoA hydratase (AUH), la ribonucleoprotein 4B (LARP4B), splicing factor 3b subunit 4 (SF3B4), YTH domain family 1 (YTHDF1), DEAD-box helicase 3 X-linked (DDX3X), eukaryotic translation initiation factor 4 gamma 2 (EIF4G2), pumilio RNA-binding family member 2 (PUM2), and TAR DNA-binding protein (TARDBP; **Figures 8B,C**) was identified. Through ROC curve analysis, the signature was shown to be able to predicate HCC survival status with an AUC of 0.752 (**Figure 8D**). With the model [risk score = $(-0.107)^{*}\ AUH_{expression} + 0.047^{*}\ LARP4B_{expression} + 0.131^{*}\ SF3B4_{expression} + 0.004^{*}\ YTHDF1_{expression} + 0.011^{*}\ DDX3X_{expression} + 0.132^{*}$

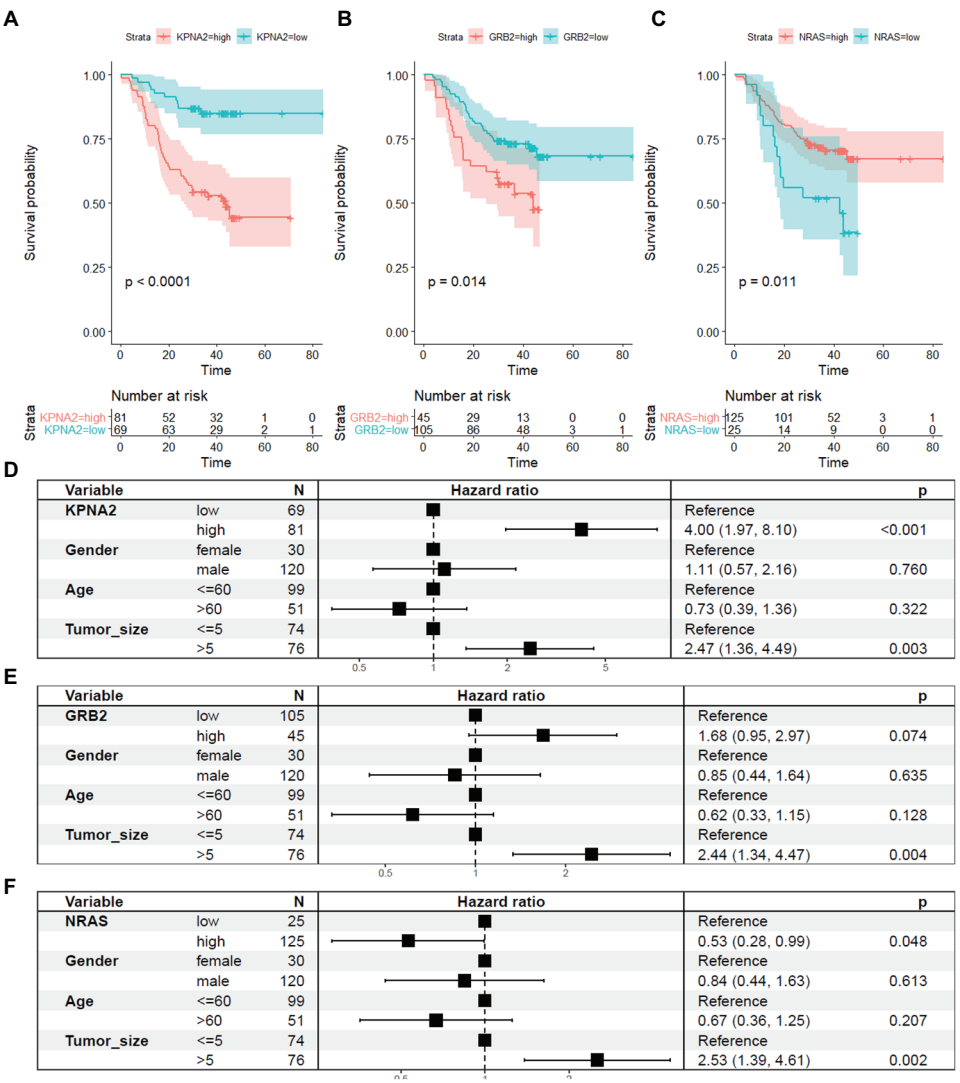

FIGURE 7 | Prognostic roles of KPNA2, GRB2, and NRAS in HCC at protein level. **(A)** Higher expression of KPNA2 indicated shorter survival in HCC patients. **(B)** Higher expression of GRB2 indicated shorter survival in HCC patients. **(C)** Higher expression of NRAS indicated longer survival in HCC patients. **(D–F)** Prognostic effects of KPNA2, GRB2, and NRAS on HCC overall survival when adjusted for gender, age, and tumor size. Paired-samples Wilcoxon test was used for protein level comparisons. Kaplan-Meier survival analysis with log-rank test and multivariable Cox regression analysis were used for evaluation of the prognostic effects of the proteins. For all the analyses, $p < 0.05$ was considered statistically significant.

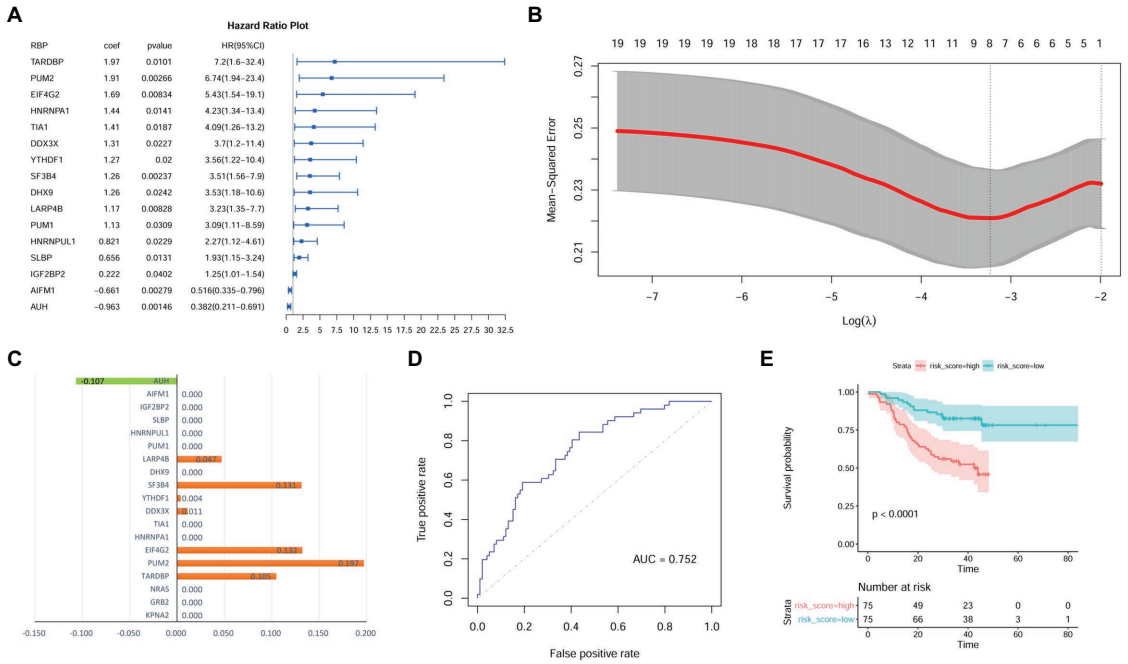

FIGURE 8 | Identification of prognostic signature with RBPs of KPNA2, GRB2, and NRAS in HCC at protein level. **(A)** Gender, age, and tumor size corrected prognostic effects of RBPs of KPNA2, GRB2, and NRAS on HCC overall survival, and the right part was the visualization of the HR (95% CI). **(B)** Tuning parameter lambda (λ) selection using 10-fold cross validation. **(C)** Eight RBPs with absolute value of coefficient >0 were included in the LASSO regression model. The *x*-axis represented the coefficients of the RBPs, and the *y*-axis indicated the RBP variables. **(D)** The LASSO regression model could predicate the survival status with an AUC of 0.752 through ROC curve analysis. **(E)** HCC patients with higher risk score were indicated to have a shorter survival than those with lower risk score. RBPs, RNA-binding proteins; HR, hazard ratio; AUC, area under the curve; ROC, receiver operating characteristic. Ezcox package in R were used for Cox regression analysis of the RBPs **(A)** with the RBP expressions as covariates and age, gender, and tumor size as controls. Kaplan-Meier survival analysis with log-rank test **(E)** was used for the overall survival comparison between patients with high and low risk score, and the median value of the risk score was used as threshold for grouping the patients. For the analysis, $p < 0.05$ was considered significant.

$EIF4G2_{expression}$ + 0.197 * $PUM2_{expression}$ + 0.105 * $TARDBP_{expression}$], the risk scores of the HCC patients were evaluated, and through Kaplan-Meier survival analysis (**Figure 8E**), the patients with high risk score were shown to have a shorter OS than those with low risk score ($p < 0.0001$).

The independent prognostic value of the model was also investigated. As shown in **Figures 9A,B**, the prognostic effects of the model were shown in both HCC patients with smaller tumors (tumor size ≤ 5 cm; $p = 0.004$; **Figure 9A**) and those with larger ones (tumor size > 5 cm; $p = 0.012$; **Figure 9B**). In male patients, higher risk score also indicated shorter survival ($p < 0.01$; **Figure 9C**). Although the survival difference between female patients with high and low risk score was not so significant ($p = 0.24$), there was an obvious shorter survival trend in the patients with high risk score than those with low one (**Figure 9D**). In addition, comparing with the HCC patients with low risk score, the patients with high risk score were shown to survive shorter both in the younger group ($p < 0.01$; **Figure 9E**) and older group ($p < 0.01$; **Figure 9F**).

As shown in **Figure 10**, through Wilcoxon test, the eight RBPs in the LASSO regression model above were also presented to be significantly differentially expressed between HCC tumors and their paired liver tissue controls. AUH, which was indicated to be a favorable prognostic factor for HCC OS, was shown to be downregulated ($p < 0.01$; **Figure 10A**), whereas the other seven unfavorable prognostic factors including LARP4B ($p < 0.01$; **Figure 10B**), SF3B4 ($p < 0.01$; **Figure 10C**), YTHDF1 ($p < 0.01$; **Figure 10D**), DDX3X ($p < 0.01$; **Figure 10E**), EIF4G2 ($p < 0.01$; **Figure 10F**), PUM2 ($p < 0.01$; **Figure 10G**), and TARDBP ($p < 0.01$; **Figure 10H**) were shown to be upregulated in HCC tumors. According to the ROC curve analysis (**Figures 10I–P**), all the eight RBPs could discriminate HCC tumors form normal livers with an AUC ranging from 0.787 (**Figure 10I**) to 0.971 (**Figure 10L**) individually. At the optimal cutoff points, their sensitivity ranged from 0.717 to 0.925, and the specificity ranged from 0.855 to 0.987, indicating their diagnostic potential in HCC diagnosis. Furthermore, all the eight RBPs were shown to be negatively or positively correlated with MKI67 expression in HCC tumors ($p < 0.05$, **Supplementary Figure S3**). Interestingly, the significant correlations of AUH (**Supplementary Figure S3A**), SF3B4 (**Supplementary Figure S3C**), DDX3X (**Supplementary Figure S3E**), PUM2 (**Supplementary Figure S3G**), and TARDBP (**Supplementary Figure S3H**) with MKI67 were only shown in the HCC tumors, while not in their paired normal liver tissues ($p > 0.05$), indicating the specific associations with HCC progression.

Through Spearman correlation analysis (**Supplementary Figure S4** and **Figure 11**), the eight RBPs above were all indicated to be correlated with KPNA2 expression at protein level ($p < 0.05$; **Figures 11A–H**), and five (AUH, YTHDF1, DDX3X, EIF4G2, and LARP4B) and four (AUH, SF3B4, TARDBP, and LARP4B)

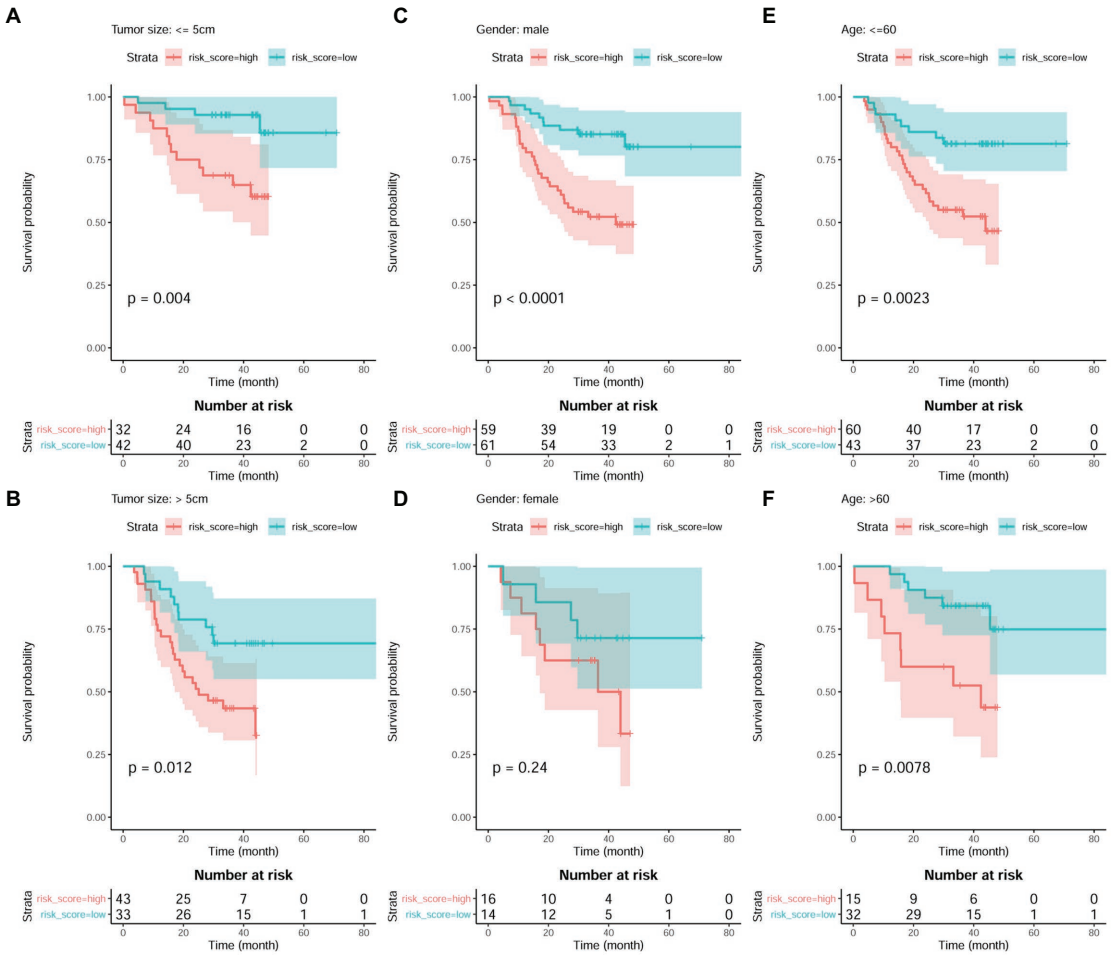

FIGURE 9 | Prognostic effects of risk score in HCC patient of different gender, age, and tumor size groups. **(A,B)** Kaplan-Meier survival analysis of risk score in HCC patients with small tumors (tumor size ≤5 cm) group and large tumors (tumor size > 5 cm), respectively. **(C,D)** Kaplan-Meier survival analysis of risk score in male HCC patients and female ones, respectively. **(E,F)** Kaplan-Meier survival analysis of risk score in young HCC patients (≤60 years) and old ones (>60 years), respectively. Kaplan-Meier survival analysis with log-rank test was used, and $p < 0.05$ was considered as significant.

of them were presented to be significantly correlated with GRB2 expression ($p < 0.05$; **Figures 11I,K–M,O**) and NRAS expression ($p < 0.05$; **Figures 11Q,R,V,W**), respectively. In addition, there were significant correlations between every two of the RBPs (**Supplementary Figure S3**), indicating their complicated associations. RBPs were implicated in RNA splicing, polyadenylation, mRNA stability, mRNA localization, and translation (Qin et al., 2020). Here, the dysregulations of RBPs might also contribute to the overexpression of KPNA2 and GRB2 and the underexpression of NRAS at protein level.

DISCUSSION

As a central immunomodulator of body, in liver, there are elements that can promote both immune tolerance and antitumor immunity, and the deregulation of the immunological network was demonstrated to be associated with chronic inflammation and tumor development (Protzer et al., 2012; Keenan et al., 2019). In a recent study, it was reported that in contrast to lung cancer and melanoma, HCC presented a low response rate to immune checkpoint inhibitors (Feng et al., 2020), indicating the complexity in the application of HCC immunotherapy. Because HCC occurs almost exclusively in the context of chronic inflammation (Ringelhan et al., 2018), further understanding of the associations of immune response would provide new clues for HCC diagnosis and therapy.

Although KPNA2 was implicated in many inflammatory processes (Liu et al., 2015; Pallett et al., 2019; Park et al., 2019), and its tumor-promoting activities in HCC (Gao et al., 2018; Lin et al., 2018; Zan et al., 2019) were shown in many studies, its associations with immune response in HCC and their clinical significance were not clearly illustrated. Here, we not only confirmed the upregulation and unfavorable prognostic effects of KPNA2 in HCC, and through further systemic analyses of its correlations with immune cell infiltrations in HCC tumors and normal liver tissues individually, we presented the specific positive correlation between KPNA2 expression and B-cell

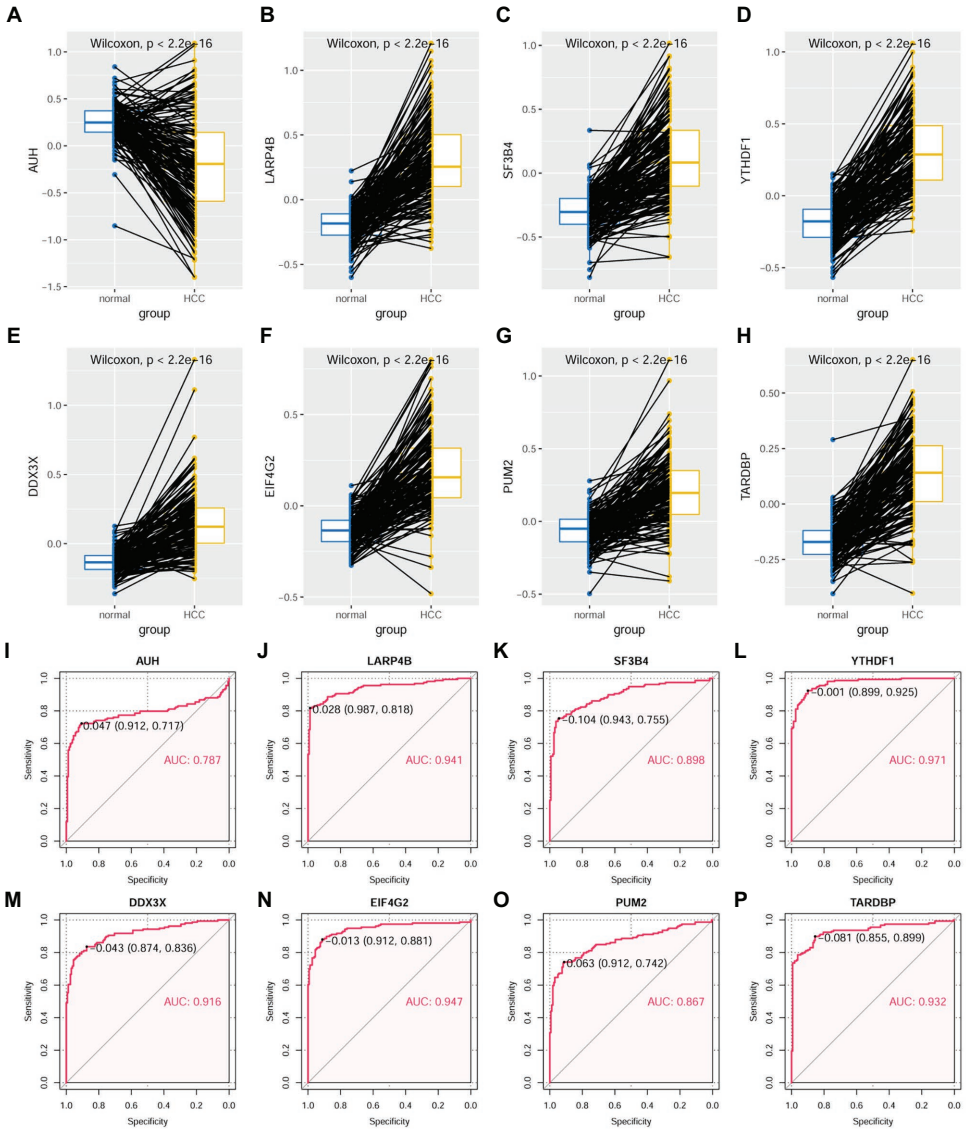

FIGURE 10 | Expressional comparisons of RBPs of KPNA2, GRB2, and NRAS between HCC and normal liver tissues. **(A)** Lower expression of AUH in HCC tumors than their paired normal liver controls. **(B–H)** Higher expression of LARP4B, SF3B4, YTHDF1, DDX3X, EIF4G2, PUM2, and TARDBP in the HCC tumors than their paired normal liver controls. **(I–P)** ROC curve analysis of AUH, LARP4B, SF3B4, YTHDF1, DDX3X, EIF4G2, PUM2, and TARDBP in discriminating HCC tumors from normal liver tissues; the cutoff point with biggest Youden index was labeled, and the corresponding specificity and sensitivity were shown. AUC, area under the curve; ROC, receiver operating characteristic. For **(A–H)**, the *x*-axis and *y*-axis represented the sample groups and the protein abundance of the proteins in the samples, with respect to the pooled reference sample, as $\log_2$ ratios. For **(I–P)**, the *x*-axis and *y*-axis represented the specificity and sensitivity of the RBPs in discriminating HCC tumors from the normal liver tissues. Paired-samples Wilcoxon test and ROC analysis was used for expression comparisons of the protein levels and evaluation of the diagnostic power of the proteins, respectively. For all the analyses, $p < 0.05$ was considered significant.

infiltration in the tumors. In addition, similar to KPNA2, in this study, B-cell infiltration was also shown to be an unfavorable prognostic factor for HCC OS, indicating the associations between B-cell infiltration and HCC progression.

BCR signaling pathway is crucial for B-cell response to the antigens, governing the processes of B-cell activation and fate decisions (Kwak et al., 2019). Here, we found the significant correlations of KPNA2 with many BCR signaling pathway genes including GRB2 and NRAS, which were shown to be dysregulated both at mRNA and protein levels. As a ubiquitously expressed adapter protein, GRB2 is crucial for normal development (Cheng et al., 1998) and implicated in cell proliferation (Downward, 1994). With its Src homology 2 (SH2) domain, GRB2 can also interact with several proteins in oncogenic signaling pathways including epidermal growth factor receptor, hepatocyte growth factor receptor, platelet-derived growth factor receptor, Bcr/Abl, and focal adhesion kinase (Lowenstein et al., 1992; Schlaepfer et al., 1994; Verbeek et al., 1997; Liu and McGlade, 1998;

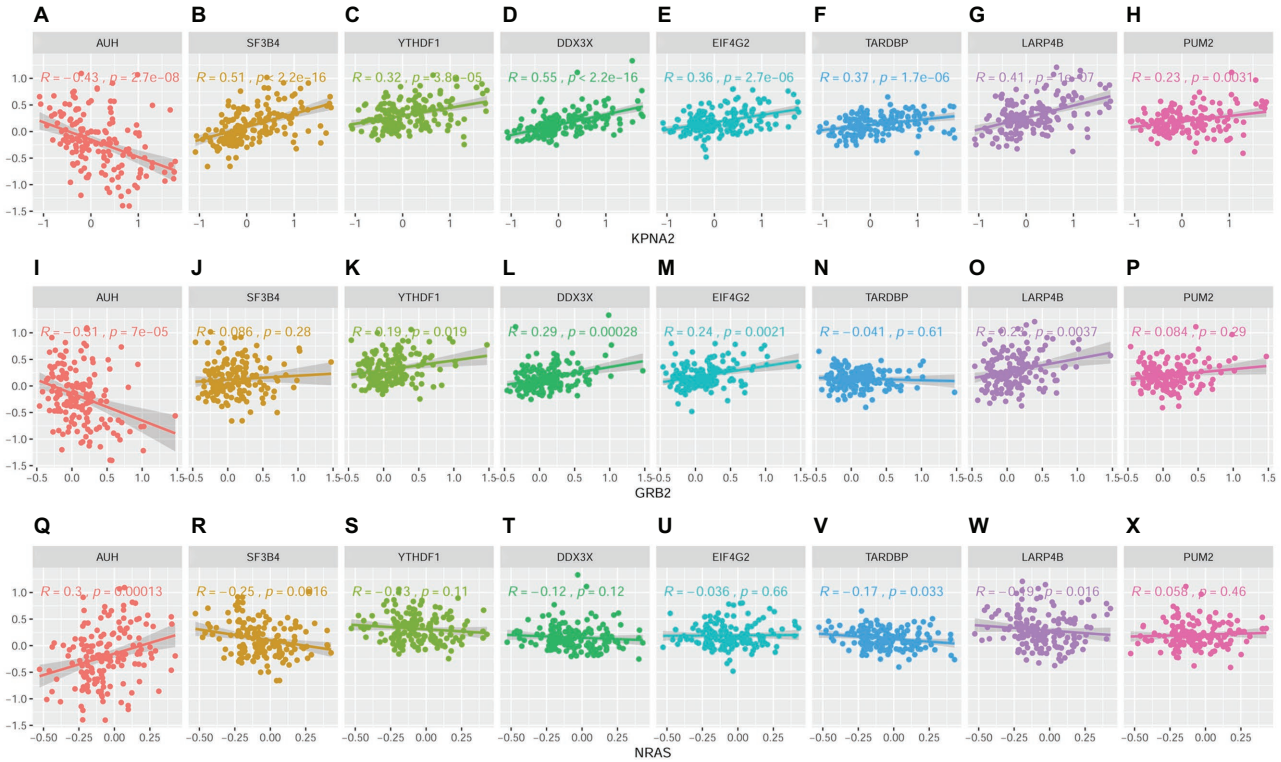

FIGURE 11 | Correlations between RBPs and KPNA2, GRB2, and NRAS at protein level in HCC. **(A–H)** Correlations between KPNA2 and AUH, SF3B4, YTHDF1, DDX3X, EIF4G2, TARDBP, LARP4B, and PUM2, respectively. **(I–P)** Correlations between GRB2 and AUH, SF3B4, YTHDF1, DDX3X, EIF4G2, TARDBP, LARP4B, and PUM2, respectively. **(Q–X)** Correlations between NRAS and AUH, SF3B4, YTHDF1, DDX3X, EIF4G2, TARDBP, LARP4B, and PUM2, respectively. The *x*-axis and *y*-axis represented the abundance of the RBPs in the HCC samples, with respect to the pooled reference sample, as $\log_2$ ratios. Spearman correlation analysis was used, and $p < 0.05$ was considered significant.

Fredericks and Ren, 2013). Recently, its tumor-promoting effects were shown in many malignancies including lung cancer (Yang et al., 2017a; Jiang et al., 2018; Mitra et al., 2018; Wang and Wang, 2020), gastric cancer (Ye et al., 2018), colorectal cancer (Ding et al., 2019), ovarian carcinoma (Huang et al., 2018), renal cell carcinoma (Gu et al., 2017), breast cancer (Lim et al., 2000; Haines et al., 2014; López-Cortés et al., 2020), and esophageal squamous cell carcinoma (Shi et al., 2018). In addition, in lung cancer (Chen et al., 2020), ovarian cancer (Xu et al., 2018), colorectal cancer (Agrawal et al., 2019), and breast cancer (Chen et al., 2018), its associations with the resistance of the tumors to chemotherapeutic drugs were presented, and its downregulation could reverse the resistant status or enhance the sensitivity to the drugs, indicating its potential as a chemotherapeutic target in the malignancies. Here, we presented the overexpression of GRB2, its positive correlation with cell proliferation marker MKI67, and its unfavorable prognostic roles in HCC, indicating its associations with HCC development and progression, consistent with the tumor-promoting activity of GRB2 in HCC reported in recent studies (Yang et al., 2018; Lv et al., 2020; Sun et al., 2020). Considering its positive correlation with KPNA2 and its important role in B-cell activation, we speculated they might play coordinate roles in HCC immunoregulation and progression. Besides its potential in HCC diagnosis and prognosis, GRB2 might be a target in HCC immunotherapy.

NRAS, a GTPase encoded by NRAS gene, originally identified in neuroblastoma cell lines, was considered as the third member of RAS family (Shimizu et al., 1983). Similar to the other two RAS members, KRAS and HRAS, NRAS was involved in cell growth, differentiation, and proliferation (Parker and Mattos, 2018; Murugan et al., 2019), and its deregulation was implicated in the metabolism of tumor cells, microenvironment remodeling, and the evasion of tumoral immune response (Mandalà et al., 2014). In fact, approximately 15–25% of all metastatic tumors were shown to have NRAS mutation (Boespflug et al., 2017), and the associations of NRAS with chemotherapy resistance were shown (Nguyen et al., 2020) in melanoma. In neuroblastoma, NRAS status was found to be associated with the sensitivity to SHP2 inhibitors (Valencia-Sama et al., 2020). In osteosarcoma, NRAS was reported to be associated with cell proliferation, migration, invasion, and methotrexate resistance (Li et al., 2020). Interestingly, in colorectal cancer, three NRAS mutants were shown to be associated with resistance to apoptosis, cytoskeletal reorganization, and loss of adhesion, whereas they have no effects on obvious effect on cell proliferation and motility (Yu and Garcia, 2020), indicating the complicated activities of the gene. In this study, we found the upregulation of NRAS and its unfavorable prognostic effects in HCC at mRNA level, consistent with a recent study that presented the overexpression of NRAS and its prognostic value in HCC (Dietrich et al., 2019). However, here, surprisingly,

we found the downregulation of NRAS and its favorable prognostic effects in HCC at protein level, opposite to its mRNA level. As we included tumor and paired normal liver samples from 159 HCC patients for investigation at protein level here, the sample size is much larger than the previous study that included 46 HCC patients (Dietrich et al., 2019), and the results are reliable. In addition, considering the variety of regulatory factors implicated in the process of translation (Grafanaki et al., 2019; Moutinho et al., 2019; Kim et al., 2020), it is understandable to see the inconsistence between the mRNA and protein levels of specific genes. As no significant correlation of NRAS with MKI67 was shown at protein level in HCC, further studies were needed to investigate the specific functions of NRAS in HCC.

RBPs are implicated in various RNA processes including RNA splicing, RNA translation, and RNA degradation (Mohibi et al., 2019), and to some extent, they can regulate the fate of specific mRNAs (Zhang et al., 2018). In this study, through investigation of the RBPs of KPNA2, GRB2, and NRAS, we identified an eight-RBP signature that could predicate HCC OS effectively. Among the eight RBPs, AUH was shown to be downregulated and have favorable prognostic effects on HCC OS. Within the 3' untranslated region of lymphokines transcripts and some protooncogenes, there are many AU-rich elements that contain various numbers of reiterated AUUUA pentamers, and as *cis* elements, they serve as signal for rapid mRNA degradation (Chen et al., 1994). As an AU-binding protein, the downregulation of AUH in this study might lead to the decrease in the degradation of its binding mRNAs and contribute to the upregulation of corresponding mRNAs including KPNA2, GRB2, and NRAS. Noticeably, here, we presented an inconsistency of NRAS between its mRNA and protein levels. Besides its RNA-binding activity, as an enoyl-CoA hydratase, AUH also plays key roles in leucine degradation (IJlst et al., 2002; Ly et al., 2003) and mitochondrial protein synthesis (Richman et al., 2014). It was demonstrated that decrease or overexpression of the AUH protein in cells could lead to defects in mitochondrial translation. Here, NRAS was shown to be downregulated issn HCC at protein level, and there was a significant positive correlation between NRAS and AUH at protein level. We speculated that the decrease in AUH protein might be associated with the dysregulation of NRAS in HCC. However, considering the dual functions of AUH in RNA binding and protein translation and its involvement in the immune response to lipopolysaccharide (Zhang et al., 2020), further investigation is needed for its specific roles in NRAS dysregulation and HCC immunoregulation.

YTHDF1, one of the readers of N-6-methyladenosine (m6A) RNA methylation (Fu et al., 2014), could promote protein synthesis by interacting with translation machinery (Wang et al., 2015). Here, we presented the upregulation of YTHDC1 and its positive correlations with KPNA2 and GRB2 at protein level in HCC, indicating its associations with the dysregulation of the two proteins. In addition, as one element of the eight-PBP prognostic signature in HCC in this study, its unfavorable effects were also shown, consistent with previous studies (Huang et al., 2020; Wu et al., 2020). Through the correlation analysis, the significant correlations between the RBPs were shown, indicating their complicated associations and their coordinating roles during HCC progression.

In summary, B-cell infiltration was positively correlated with KPNA2 expression and unfavorably prognostic for HCC survival, indicating its associations with HCC progression. The significant correlations of KPNA2 with BCR signaling pathway indicated its association with BCR signaling. The dysregulation of BCR signaling pathway genes in HCC tumors indicated the crucial roles of BCR signaling pathway in the immune response of HCC. GRB2 and NRAS were dysregulated and had prognostic roles in HCC at both mRNA and protein levels, and they should be considered during HCC immunotherapy. Among the RBPs of KPNA2, GRB2, and NRAS, an eight-RBP signature with independent prognostic effects was identified. As all the eight RBPs were dysregulated in HCC, they might be new markers for HCC diagnosis and prognosis predication. Considering the complicated correlations between the RBPs, they might coordinate with each other in various biological processes, and their specific roles in the regulation of KPNA2, GRB2, and NRAS expressions needed further investigation.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

XZ and LD conceived and designed the study. XZ, SF, JLZ, and FG collected and analyzed the data. XZ and JZZ interpreted the data. XZ drafted the manuscript. LD and SF reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Key Scientific Research Project of Colleges and Universities in Henan Province of China (grant number 20B320007), the Joint Construction Project of Henan Medical Science and Technology Research Plan (LHGJ20190717), Changsha Science and Technology Bureau Project (kq1701050), and Technology Plan of Youth Natural Science Foundation Hunan Province (S2017JJQNJJ710).

ACKNOWLEDGMENTS

We thank J. M. Zeng from University of Macau and his team for their help in the bioinformatics analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.593273/full#supplementary-material

REFERENCES

Agrawal, S., Woźniak, M., Łuc, M., Makuch, S., Pielka, E., Agrawal, A. K., et al. (2019). Insulin enhancement of the antitumor activity of chemotherapeutic agents in colorectal cancer is linked with downregulating PIK3CA and GRB2. *Sci. Rep.* 9:16647. doi: 10.1038/s41598-019-53145-x

Boespflug, A., Caramel, J., Dalle, S., and Thomas, L. (2017). Treatment of NRAS-mutated advanced or metastatic melanoma: rationale, current trials and evidence to date. *Ther. Adv. Med. Oncol.* 9, 481–492. doi: 10.1177/1758834017708160

Booth, D. G., Takagi, M., Sanchez-Pulido, L., Petfalski, E., Vargiu, G., Samejima, K., et al. (2014). Ki-67 is a PP1-interacting protein that organises the mitotic chromosome periphery. *eLife* 3:e01641. doi: 10.7554/eLife.01641

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 Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Chen, C. Y., Chen, T. M., and Shyu, A. B. (1994). Interplay of two functionally and structurally distinct domains of the c-fos AU-rich element specifies its mRNA-destabilizing function. *Mol. Cell. Biol.* 14, 416–426. doi: 10.1128/mcb.14.1.416

Chen, P. F., Li, Q. H., Zeng, L. R., Yang, X. Y., Peng, P. L., He, J. H., et al. (2019). A 4-gene prognostic signature predicting survival in hepatocellular carcinoma. *J. Cell. Biochem.* 120, 9117–9124. doi: 10.1002/jcb.28187

Chen, D., Si, W., Shen, J., Du, C., Lou, W., Bao, C., et al. (2018). miR-27b-3p inhibits proliferation and potentially reverses multi-chemoresistance by targeting CBLB/GRB2 in breast cancer cells. *Cell Death Dis.* 9:188. doi: 10.1038/s41419-017-0211-4

Chen, Y., Wu, J., Yan, H., Cheng, Y., Wang, Y., Yang, Y., et al. (2020). Lymecycline reverses acquired EGFR-TKI resistance in non-small-cell lung cancer by targeting GRB2. *Pharmacol. Res.* 159:105007. doi: 10.1016/j.phrs.2020.105007

Cheng, A. M., Saxton, T. M., Sakai, R., Kulkarni, S., Mbamalu, G., Vogel, W., et al. (1998). Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. *Cell* 95, 793–803. doi: 10.1016/s0092-8674(00)81702-x

Dietrich, P., Gaza, A., Wormser, L., Fritz, V., Hellerbrand, C., and Bosserhoff, A. K. (2019). Neuroblastoma RAS viral oncogene homolog (NRAS) is a novel prognostic marker and contributes to sorafenib resistance in hepatocellular carcinoma. *Neoplasia* 21, 257–268. doi: 10.1016/j.neo.2018.11.011

Ding, C., Tang, W., Wu, H., Fan, X., Luo, J., Feng, J., et al. (2019). The PEAK1-PPP1R12B axis inhibits tumor growth and metastasis by regulating Grb2/PI3K/Akt signalling in colorectal cancer. *Cancer Lett.* 442, 383–395. doi: 10.1016/j.canlet.2018.11.014

Downward, J. (1994). The GRB2/Sem-5 adaptor protein. *FEBS Lett.* 338, 113–117. doi: 10.1016/0014-5793(94)80346-3

Duan, M., Hu, F., Li, D., Wu, S., and Peng, N. (2020). Silencing KPNA2 inhibits IL-6-induced breast cancer exacerbation by blocking NF-κB signaling and c-Myc nuclear translocation in vitro. *Life Sci.* 253:117736. doi: 10.1016/j.lfs.2020.117736

El-Serag, H. B., Marrero, J. A., Rudolph, L., and Reddy, K. R. (2008). Diagnosis and treatment of hepatocellular carcinoma. *Gastroenterology* 134, 1752–1763. doi: 10.1053/j.gastro.2008.02.090

Feng, G. S., Hanley, K. L., Liang, Y., and Lin, X. (2020). Improving the efficacy of liver cancer immunotherapy: the power of combined preclinical and clinical studies. *Hepatology* doi: 10.1002/hep.31479 [Epub ahead of print]

Fredericks, J., and Ren, R. (2013). The role of RAS effectors in BCR/ABL induced chronic myelogenous leukemia. *Front. Med.* 7, 452–461. doi: 10.1007/s11684-013-0304-0

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m^6A RNA methylation. *Nat. Rev. Genet.* 15, 293–306. doi: 10.1038/nrg3724

Gao, C. L., Wang, G. W., Yang, G. Q., Yang, H., and Zhuang, L. (2018). Karyopherin subunit-α 2 expression accelerates cell cycle progression by upregulating CCNB2 and CDK1 in hepatocellular carcinoma. *Oncol. Lett.* 15, 2815–2820. doi: 10.3892/ol.2017.7691

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. doi: 10.1016/j.cell.2019.08.052

Gerdes, J., Schwab, U., Lemke, H., and Stein, H. (1983). Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. *Int. J. Cancer* 31, 13–20. doi: 10.1002/ijc.2910310104

Goldfarb, D. S., Corbett, A. H., Mason, D. A., Harreman, M. T., and Adam, S. A. (2004). Importin alpha: a multipurpose nuclear-transport receptor. *Trends Cell Biol.* 14, 505–514. doi: 10.1016/j.tcb.2004.07.016

Grafanaki, K., Anastasakis, D., Kyriakopoulos, G., Skeparnias, I., Georgiou, S., and Stathopoulos, C. (2019). Translation regulation in skin cancer from a tRNA point of view. *Epigenomics* 11, 215–245. doi: 10.2217/epi-2018-0176

Groheux, D., Biard, L., Lehmann-Che, J., Teixeira, L., Bouhidel, F. A., Poirot, B., et al. (2018). Tumor metabolism assessed by FDG-PET/CT and tumor proliferation assessed by genomic grade index to predict response to neoadjuvant chemotherapy in triple negative breast cancer. *Eur. J. Nucl. Med. Mol. Imaging* 45, 1279–1288. doi: 10.1007/s00259-018-3998-z

Gu, D. H., Mao, J. H., Pan, X. D., Zhu, H., Chen, X., Zheng, B., et al. (2017). microRNA-302c-3p inhibits renal cell carcinoma cell proliferation by targeting Grb2-associated binding 2 (Gab2). *Oncotarget* 8, 26334–26343. doi: 10.18632/oncotarget.15463

Guo, X., Wang, Z., Zhang, J., Xu, Q., Hou, G., Yang, Y., et al. (2019). Upregulated KPNA2 promotes hepatocellular carcinoma progression and indicates prognostic significance across human cancer types. *Acta Biochim. Biophys. Sin.* 51, 285–292. doi: 10.1093/abbs/gmz003

Haines, E., Saucier, C., and Claing, A. (2014). The adaptor proteins p66Shc and Grb2 regulate the activation of the GTPases ARF1 and ARF6 in invasive breast cancer cells. *J. Biol. Chem.* 289, 5687–5703. doi: 10.1074/jbc.M113.516047

Hua, S., Ji, Z., Quan, Y., Zhan, M., Wang, H., Li, W., et al. (2020). Identification of hub genes in hepatocellular carcinoma using integrated bioinformatic analysis. *Aging* 12, 5439–5468. doi: 10.18632/aging.102969

Huang, H., Bai, Y., Lu, X., Xu, Y., Zhao, H., and Sang, X. (2020). N6-methyladenosine associated prognostic model in hepatocellular carcinoma. *Ann. Transl. Med.* 8:633. doi: 10.21037/atm-20-2894

Huang, Y., Hu, Y., Jin, Z., and Shen, Z. (2018). LncRNA snaR upregulates GRB2-associated binding protein 2 and promotes proliferation of ovarian carcinoma cells. *Biochem. Biophys. Res. Commun.* 503, 2028–2032. doi: 10.1016/j.bbrc.2018.07.152

IJlst, L., Loupatty, F. J., Ruiter, J. P., Duran, M., Lehnert, W., and Wanders, R. J. (2002). 3-Methylglutaconic aciduria type I is caused by mutations in AUH. *Am. J. Hum. Genet.* 71, 1463–1466. doi: 10.1086/344712

Jeong, D., Kim, H., Ban, S., Oh, S., Ji, S., Kim, D., et al. (2017). Karyopherin α-2 is a reliable marker for identification of patients with high-risk stage II colorectal cancer. *J. Cancer Res. Clin. Oncol.* 143, 2493–2503. doi: 10.1007/s00432-017-2512-5

Jiang, P., Tang, Y., He, L., Tang, H., Liang, M., Mai, C., et al. (2014). Aberrant expression of nuclear KPNA2 is correlated with early recurrence and poor prognosis in patients with small hepatocellular carcinoma after hepatectomy. *Med. Oncol.* 31:131. doi: 10.1007/s12032-014-0131-4

Jiang, W., Wei, K., Pan, C., Li, H., Cao, J., Han, X., et al. (2018). MicroRNA-1258 suppresses tumour progression via GRB2/Ras/Erk pathway in non-small-cell lung cancer. *Cell Prolif.* 51:e12502. doi: 10.1111/cpr.12502

Keenan, B. P., Fong, L., and Kelley, R. K. (2019). Immunotherapy in hepatocellular carcinoma: the complex interface between inflammation, fibrosis, and the immune response. *J. Immunother. Cancer* 7:267. doi: 10.1186/s40425-019-0749-z

Kim, D. S., Challa, S., Jones, A., and Kraus, W. L. (2020). PARPs and ADP-ribosylation in RNA biology: from RNA expression and processing to protein translation and proteostasis. *Genes Dev.* 34, 302–320. doi: 10.1101/gad.334433.119

Kriegsmann, K., Baertsch, M. A., Awwad, M. H. S., Merz, M., Hose, D., Seckinger, A., et al. (2019). Cereblon-binding proteins expression levels correlate with hyperdiploidy in newly diagnosed multiple myeloma patients. *Blood Cancer J.* 9:13. doi: 10.1038/s41408-019-0174-z

Kwak, K., Akkaya, M., and Pierce, S. K. (2019). B cell signaling in context. *Nat. Immunol.* 20, 963–969. doi: 10.1038/s41590-019-0427-9

Li, X., Liu, Y., Zhang, X., Shen, J., Xu, R., Liu, Y., et al. (2020). Circular RNA hsa_circ_0000073 contributes to osteosarcoma cell proliferation, migration, invasion and methotrexate resistance by sponging miR-145-5p and miR-151-3p and upregulating NRAS. *Aging* 12, 14157–14173. doi: 10.18632/aging.103423

Lim, S. J., Lopez-Berestein, G., Hung, M. C., Lupu, R., and Tari, A. M. (2000). Grb2 downregulation leads to Akt inactivation in heregulin-stimulated and

ErbB2-overexpressing breast cancer cells. *Oncogene* 19, 6271–6276. doi: 10.1038/sj.onc.1204014

Lin, C. Z., Ou, R. W., and Hu, Y. H. (2018). Lentiviral-mediated microRNA-26b up-regulation inhibits proliferation and migration of hepatocellular carcinoma cells. *Kaohsiung J. Med. Sci.* 34, 547–555. doi: 10.1016/j.kjms.2018.05.003

Liu, J., Li, W., Zhang, J., Ma, Z., Wu, X., and Tang, L. (2019). Identification of key genes and long non-coding RNA associated ceRNA networks in hepatocellular carcinoma. *PeerJ* 7:e8021. doi: 10.7717/peerj.8021

Liu, S. K., and McGlade, C. J. (1998). Gads is a novel SH2 and SH3 domain-containing adaptor protein that binds to tyrosine-phosphorylated Shc. *Oncogene* 17, 3073–3082. doi: 10.1038/sj.onc.1202337

Liu, Z., Zhang, D., Sun, C., Tao, R., Xu, X., Xu, L., et al. (2015). KPNA2 contributes to the inflammatory processes in synovial tissue of patients with rheumatoid arthritis and SW982 cells. *Inflammation* 38, 2224–2234. doi: 10.1007/s10753-015-0205-2

López-Cortés, A., Paz, Y. M. C., Guerrero, S., Cabrera-Andrade, A., Barigye, S. J., Munteanu, C. R., et al. (2020). OncoOmics approaches to reveal essential genes in breast cancer: a panoramic view from pathogenesis to precision medicine. *Sci. Rep.* 10:5285. doi: 10.1038/s41598-020-62279-2

Lowenstein, E. J., Daly, R. J., Batzer, A. G., Li, W., Margolis, B., Lammers, R., et al. (1992). The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. *Cell* 70, 431–442. doi: 10.1016/0092-8674(92)90167-b

Lv, J., Zhang, S., Wu, H., Lu, J., Lu, Y., Wang, F., et al. (2020). Deubiquitinase PSMD14 enhances hepatocellular carcinoma growth and metastasis by stabilizing GRB2. *Cancer Lett.* 469, 22–34. doi: 10.1016/j.canlet.2019.10.025

Ly, T. B., Peters, V., Gibson, K. M., Liesert, M., Buckel, W., Wilcken, B., et al. (2003). Mutations in the AUH gene cause 3-methylglutaconic aciduria type I. *Hum. Mutat.* 21, 401–407. doi: 10.1002/humu.10202

Mandalà, M., Merelli, B., and Massi, D. (2014). Nras in melanoma: targeting the undruggable target. *Crit. Rev. Oncol. Hematol.* 92, 107–122. doi: 10.1016/j.critrevonc.2014.05.005

Miao, Y. -R., Zhang, Q., Lei, Q., Luo, M., Xie, G. -Y., Wang, H., et al. (2020). ImmuCellAI: a unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. *Adv. Sci.* 7:1902880. doi: 10.1002/advs.201902880

Mitra, P., Kalailingam, P., Tan, H. B., and Thanabalu, T. (2018). Overexpression of GRB2 enhances epithelial to mesenchymal transition of A549 cells by upregulating SNAIL expression. *Cells* 7:97. doi: 10.3390/cells7080097

Mohibi, S., Chen, X., and Zhang, J. (2019). Cancer the"RBP"eutics-RNA-binding proteins as therapeutic targets for cancer. *Pharmacol. Ther.* 203:107390. doi: 10.1016/j.pharmthera.2019.07.001

Moutinho, M., Codocedo, J. F., Puntambekar, S. S., and Landreth, G. E. (2019). Nuclear receptors as therapeutic targets for neurodegenerative diseases: lost in translation. *Annu. Rev. Pharmacol. Toxicol.* 59, 237–261. doi: 10.1146/annurev-pharmtox-010818-021807

Müller, T., Tolkach, Y., Stahl, D., Steiner, S., Hauser, S., Ellinger, J., et al. (2019). Karyopherin alpha 2 is an adverse prognostic factor in clear-cell and papillary renal-cell carcinoma. *Clin. Genitourin. Cancer* 17, e167–e175. doi: 10.1016/j.clgc.2018.10.008

Murugan, A. K., Grieco, M., and Tsuchida, N. (2019). RAS mutations in human cancers: roles in precision medicine. *Semin. Cancer Biol.* 59, 23–35. doi: 10.1016/j.semcancer.2019.06.007

Nguyen, M. Q., Teh, J. L. F., Purwin, T. J., Chervoneva, I., Davies, M. A., Nathanson, K. L., et al. (2020). Targeting PHGDH upregulation reduces glutathione levels and re-sensitizes resistant NRAS mutant melanoma to MEK inhibition. *J. Invest. Dermatol.* doi: 10.1016/j.jid.2020.02.047 [Epub ahead of print]

Ostasiewicz, B., Ostasiewicz, P., Duś-Szachniewicz, K., Ostasiewicz, K., and Ziółkowski, P. (2016). Quantitative analysis of gene expression in fixed colorectal carcinoma samples as a method for biomarker validation. *Mol. Med. Rep.* 13, 5084–5092. doi: 10.3892/mmr.2016.5200

Pallett, M. A., Ren, H., Zhang, R. Y., Scutts, S. R., Gonzalez, L., Zhu, Z., et al. (2019). Vaccinia virus BBK E3 ligase adaptor A55 targets importin-dependent NF-κB activation and inhibits CD8(+) T-cell memory. *J. Virol.* 93, e00051–e00019. doi: 10.1128/jvi.00051-19

Park, S. H., Ham, S., Lee, A., Möller, A., and Kim, T. S. (2019). NLRP3 negatively regulates Treg differentiation through Kpna2-mediated nuclear translocation. *J. Biol. Chem.* 294, 17951–17961. doi: 10.1074/jbc.RA119.010545

Parker, J. A., and Mattos, C. (2018). The K-Ras, N-Ras, and H-Ras isoforms: unique conformational preferences and implications for targeting oncogenic mutants. *Cold Spring Harb. Perspect. Med.* 8:a031427. doi: 10.1101/cshperspect.a031427

Peng, N., Yang, X., Zhu, C., Zhou, L., Yu, H., Li, M., et al. (2018). MicroRNA-302 cluster downregulates enterovirus 71-induced innate immune response by targeting KPNA2. *J. Immunol.* 201, 145–156. doi: 10.4049/jimmunol.1701692

Protzer, U., Maini, M. K., and Knolle, P. A. (2012). Living in the liver: hepatic infections. *Nat. Rev. Immunol.* 12, 201–213. doi: 10.1038/nri3169

Puri, K. D., Di Paolo, J. A., and Gold, M. R. (2013). B-cell receptor signaling inhibitors for treatment of autoimmune inflammatory diseases and B-cell malignancies. *Int. Rev. Immunol.* 32, 397–427. doi: 10.3109/08830185.2013.818140

Qin, H., Ni, H., Liu, Y., Yuan, Y., Xi, T., Li, X., et al. (2020). RNA-binding proteins in tumor progression. *J. Hematol. Oncol.* 13:90. doi: 10.1186/s13045-020-00927-w

Richman, T. R., Davies, S. M., Shearwood, A. M., Ermer, J. A., Scott, L. H., Hibbs, M. E., et al. (2014). A bifunctional protein regulates mitochondrial protein synthesis. *Nucleic Acids Res.* 42, 5483–5494. doi: 10.1093/nar/gku179

Ringelhan, M., Pfister, D., O'Connor, T., Pikarsky, E., and Heikenwalder, M. (2018). The immunology of hepatocellular carcinoma. *Nat. Immunol.* 19, 222–232. doi: 10.1038/s41590-018-0044-z

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. doi: 10.1186/1471-2105-12-77

Schlaepfer, D. D., Hanks, S. K., Hunter, T., and van der Geer, P. (1994). Integrin-mediated signal transduction linked to Ras pathway by GRB2 binding to focal adhesion kinase. *Nature* 372, 786–791. doi: 10.1038/372786a0

Scholzen, T., and Gerdes, J. (2000). The Ki-67 protein: from the known and the unknown. *J. Cell. Physiol.* 182, 311–322. doi: 10.1002/(sici)1097-4652(200003)182:3<311::aid-jcp1>3.0.co;2-9

Shi, Q., Wang, Y., Mu, Y., Wang, X., and Fan, Q. (2018). MiR-433-3p inhibits proliferation and invasion of esophageal squamous cell carcinoma by targeting GRB2. *Cell. Physiol. Biochem.* 46, 2187–2196. doi: 10.1159/000489548

Shimizu, K., Goldfarb, M., Suard, Y., Perucho, M., Li, Y., Kamata, T., et al. (1983). Three human transforming genes are related to the viral ras oncogenes. *Proc. Natl. Acad. Sci. U. S. A.* 80, 2112–2116. doi: 10.1073/pnas.80.8.2112

Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer statistics, 2018. *CA Cancer J. Clin.* 68, 7–30. doi: 10.3322/caac.21442

Skånland, S. S., Karlsen, L., and Taskén, K. (2020). B cell signaling pathways - new targets for precision medicine in CLL. *Scand. J. Immunol.* e12931. doi: 10.1111/sji.12931 [Epub ahead of print]

Song, K. H., Jung, S. Y., Kang, S. M., Kim, M. H., Ahn, J., Hwang, S. G., et al. (2016). Induction of immunogenic cell death by radiation-upregulated karyopherin alpha 2 in vitro. *Eur. J. Cell Biol.* 95, 219–227. doi: 10.1016/j.ejcb.2016.04.002

Song, H., Song, J., Lu, L., and Li, S. (2019). SNHG8 is upregulated in esophageal squamous cell carcinoma and directly sponges microRNA-411 to increase oncogenicity by upregulating KPNA2. *Onco. Targets. Ther.* 12, 6991–7004. doi: 10.2147/ott.s214881

Sternburg, E. L., and Karginov, F. V. (2020). Global approaches in studying RNA-binding protein interaction networks. *Trends Biochem. Sci.* 45, 593–603. doi: 10.1016/j.tibs.2020.03.005

Sun, C., Huang, S., Hou, Y., Li, Z., Xia, D., Zhang, L., et al. (2020). Long noncoding RNA AC092171.4 promotes hepatocellular carcinoma progression by sponging microRNA-1271 and upregulating GRB2. *Aging* 12, 14141–14156. doi: 10.18632/aging.103419

Tachita, T., Kinoshita, S., Ri, M., Aoki, S., Asano, A., Kanamori, T., et al. (2020). Expression, mutation, and methylation of cereblon-pathway genes at pre- and post-lenalidomide treatment in multiple myeloma. *Cancer Sci.* 111, 1333–1343. doi: 10.1111/cas.14352

Valencia-Sama, I., Ladumor, Y., Kee, L., Adderley, T., Christopher, G., Robinson, C. M., et al. (2020). NRAS status determines sensitivity to SHP2 inhibitor combination therapies targeting the RAS-MAPK pathway in neuroblastoma. *Cancer Res.* 80, 3413–3423. doi: 10.1158/0008-5472.can-19-3822

Verbeek, B. S., Adriaansen-Slot, S. S., Rijksen, G., and Vroom, T. M. (1997). Grb2 overexpression in nuclei and cytoplasm of human breast cells: a histochemical and biochemical study of normal and neoplastic mammary

tissue specimens. *J. Pathol.* 183, 195–203. doi: 10.1002/(sici)1096-9896(199710)183:2<195::aid-path901>3.0.co;2-y

Wang, J. J., Huang, Y. Q., Song, W., Li, Y. F., Wang, H., Wang, W. J., et al. (2019a). Comprehensive analysis of the lncRNA-associated competing endogenous RNA network in breast cancer. *Oncol. Rep.* 42, 2572–2582. doi: 10.3892/or.2019.7374

Wang, A., and Wang, J. (2020). E2F1-induced overexpression of long noncoding RNA SBF2-AS1 promotes non-small-cell lung Cancer metastasis through regulating miR-362-3p/GRB2 axis. *DNA Cell Biol.* 39, 1290–1298. doi: 10.1089/dna.2020.5426

Wang, C. I., Yu, C. J., Huang, Y., Yi, J. S., Cheng, H. W., Kao, H. K., et al. (2018). Association of overexpressed karyopherin alpha 2 with poor survival and its contribution to interleukin-1β-induced matrix metalloproteinase expression in oral cancer. *Head Neck* 40, 1719–1733. doi: 10.1002/hed.25145

Wang, S., Zhang, J., He, Z., Wu, K., and Liu, X. -S. (2019b). The predictive power of tumor mutational burden in lung cancer immunotherapy response is influenced by patients' sex. *Int. J. Cancer* 145, 2840–2849. doi: 10.1002/ijc.32327

Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., et al. (2015). N(6)-methyladenosine modulates messenger RNA translation efficiency. *Cell* 161, 1388–1399. doi: 10.1016/j.cell.2015.05.014

Wu, X., Zhang, X., Tao, L., Dai, X., and Chen, P. (2020). Prognostic value of an m6A RNA methylation regulator-based signature in patients with hepatocellular carcinoma. *Biomed. Res. Int.* 2020, 1–11. doi: 10.1155/2020/2053902

Xu, Z. H., Yao, T. Z., and Liu, W. (2018). miR-378a-3p sensitizes ovarian cancer cells to cisplatin through targeting MAPK1/GRB2. *Biomed. Pharmacother.* 107, 1410–1417. doi: 10.1016/j.biopha.2018.08.132

Yang, Y., Guo, J., Hao, Y., Wang, F., Li, F., Shuang, S., et al. (2017b). Silencing of karyopherin α2 inhibits cell growth and survival in human hepatocellular carcinoma. *Oncotarget* 8, 36289–36304. doi: 10.18632/oncotarget.16749

Yang, L., Wang, X., Xu, J., Wen, Y., Zhang, M., Lu, J., et al. (2018). Integrated transcriptomic and proteomic analyses reveal α-lipoic acid-regulated cell proliferation via Grb2-mediated signalling in hepatic cancer cells. *J. Cell. Mol. Med.* 22, 2981–2992. doi: 10.1111/jcmm.13447

Yang, L., Wen, Y., Lv, G., Lin, Y., Tang, J., Lu, J., et al. (2017a). α-Lipoic acid inhibits human lung cancer cell proliferation through Grb2-mediated EGFR downregulation. *Biochem. Biophys. Res. Commun.* 494, 325–331. doi: 10.1016/j.bbrc.2017.10.030

Ye, B., Duan, B., Deng, W., Wang, Y., Chen, Y., Cui, J., et al. (2018). EGF stimulates Rab35 activation and gastric cancer cell migration by regulating DENND1A-Grb2 complex formation. *Front. Pharmacol.* 9:1343. doi: 10.3389/fphar.2018.01343

Yu, R. T. D., and Garcia, R. L. (2020). NRAS mutant E132K identified in young-onset sporadic colorectal cancer and the canonical mutants G12D and Q61K affect distinct oncogenic phenotypes. *Sci. Rep.* 10:11028. doi: 10.1038/s41598-020-67796-8

Yu, B., Liang, H., Ye, Q., and Wang, Y. (2020). Establishment of a genomic-clinicopathologic nomogram for predicting early recurrence of hepatocellular carcinoma after R0 resection. *J. Gastrointest. Surg.* doi: 10.1007/s11605-020-04554-1 [Epub ahead of print]

Yue, C., Ren, Y., Ge, H., Liang, C., Xu, Y., Li, G., et al. (2019). Comprehensive analysis of potential prognostic genes for the construction of a competing endogenous RNA regulatory network in hepatocellular carcinoma. *Onco Targets Ther.* 12, 561–576. doi: 10.2147/ott.s188913

Zan, Y., Wang, B., Liang, L., Deng, Y., Tian, T., Dai, Z., et al. (2019). MicroRNA-139 inhibits hepatocellular carcinoma cell growth through down-regulating karyopherin alpha 2. *J. Exp. Clin. Cancer Res.* 38:182. doi: 10.1186/s13046-019-1175-2

Zhang, J., Xu, E., Ren, C., Yang, H. J., Zhang, Y., Sun, W., et al. (2018). Genetic ablation of Rbm38 promotes lymphomagenesis in the context of mutant p53 by downregulating PTEN. *Cancer Res.* 78, 1511–1521. doi: 10.1158/0008-5472.can-17-2457

Zhang, D., Zhao, P., Liu, J., Qi, T., Liu, Q., Jiang, S., et al. (2020). Transcriptome analysis reveals the tolerance mechanism of mantis shrimp (*Oratosquilla oratoria*) under a lipopolysaccharide challenge. *ACS Omega* 5, 2310–2317. doi: 10.1021/acsomega.9b03629

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 November 2020
doi: 10.3389/fgene.2020.597508

Serum MiR-4687-3p Has Potential for Diagnosis and Carcinogenesis in Non-small Cell Lung Cancer

Man Liu[1,6], Qiufang Si[3,6], Songyun Ouyang[4], Zhigang Zhou[5], Meng Wang[5], Chunling Zhao[4], Ting Yang[3,6], Yulin Wang[1,6], Xue Zhang[1,6], Wenbo Xie[7], Liping Dai[1,6] and Jitian Li[2]**

[1] Henan Institute of Medical and Pharmaceutical Sciences, Academy of Medical Sciences, Zhengzhou University, Zhengzhou, China, [2] Laboratory of Molecular Biology, Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), Zhengzhou, China, [3] BGI College, Zhengzhou University, Zhengzhou, China, [4] Department of Respiratory and Sleep Medicine, The First Affiliated Hospital in Zhengzhou University, Zhengzhou, China, [5] Department of Radiology, The First Affiliated Hospital in Zhengzhou University, Zhengzhou, China, [6] Henan Key Laboratory of Tumor Epidemiology & State Key Laboratory of Esophageal Cancer Prevention, Zhengzhou University, Zhengzhou, China, [7] Department of Computer Science, College of Engineering, University of Texas at El Paso, El Paso, TX, United States

OPEN ACCESS

Edited by:
Xiaofeng Dai,
Jiangnan University, China

Reviewed by:
Ximing Tang,
University of Texas MD Anderson Cancer Center, United States
Zhe Chen,
Shanghai Jiao Tong University, China

***Correspondence:**
Liping Dai
lpdai@zzu.edu.cn
Jitian Li
jitianlee@hotmail.com

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *21 August 2020*
Accepted: *08 October 2020*
Published: *23 November 2020*

Citation:
Liu M, Si Q, Ouyang S, Zhou Z, Wang M, Zhao C, Yang T, Wang Y, Zhang X, Xie W, Dai L and Li J (2020) Serum MiR-4687-3p Has Potential for Diagnosis and Carcinogenesis in Non-small Cell Lung Cancer. Front. Genet. 11:597508. doi: 10.3389/fgene.2020.597508

The lack of a useful biomarker partly contributes to the increased mortality of non-small cell lung cancer (NSCLC). MiRNAs have become increasingly appreciated in diagnosis of NSCLC. In the present study, we used microarray to screen 2,549 miRNAs in serum samples from the training cohort (NSCLC, n = 10; the healthy, n = 10) to discover differentially expressed miRNAs (DEMs). Quantitative reverse-transcription polymerase chain reaction (qRT-PCR) assay was applied to validate the expression level of selected overexpressed DEMs of NSCLC in a validation cohort (NSCLC, n = 30; the healthy, n = 30). Area under the receiver operating characteristic curve (AUC) was performed to evaluate diagnostic capability of the DEMs. The expression of the miRNAs in tissues was analyzed based on the TCGA database. Subsequently, the target genes of the miR-4687-3p were predicted by TargetScan. Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were tested by R software (ClusterProfiler package). NSCLC cells were transfected with inhibitor or mimic to down-regulate or up-regulate the miR-4687-3p level. The function of miR-4687-3p on proliferation, invasion, and migration of lung cancer cells were investigated through CCK-8 and Transwell assays, respectively. In the results, we identified serum miR-4687-3p that provided a high diagnostic accuracy of NSCLC (AUC = 0.679, 95%CI: 0.543–0.815) in the validation cohort. According to the TCGA database, we found that the miR-4687-3p level was significantly higher in NSCLC tissues than in normal lung tissues ($p < 0.05$). GO and KEGG pathway enrichment analysis showed that postsynaptic specialization and TGF-β signaling pathway were significantly enriched. Down-regulation of miR-4687-3p could suppress the proliferation, invasion, and migration of the NSCLC cells, compared with inhibitor negative control

(NC). Meanwhile, overexpression of miR-4687-3p could promote the proliferation, invasion, and migration of the NSCLC cells compared with mimic NC. As a conclusion, our study first discovered that serum miR-4687-3p might have clinical potential as a non-invasive diagnostic biomarker for NSCLC and play an important role in the development of NSCLC.

Keywords: NSCLC, miR-4687-3p, microarray, bioinformatics, biomarker

INTRODUCTION

Lung cancer is the most common malignancy and the leading cause of malignancy-related death in the world, with an estimated 2.3 million new cases and 1.4 million deaths (Siegel and Miller, 2019). In addition, lung cancer occupied the highest incidence and mortality in numerous malignant cancers in China (Zheng et al., 2019). In the absence of significant symptoms and practical diagnostic methods, most patients are at advanced disease (Zheng et al., 2019). In terms of the pathological type, non-small cell lung cancer (NSCLC) accounts for about 85% of lung cancer. Therefore, exploring early detection methods for NSCLC is crucial to reduce NSCLC-related deaths. Low-dose spiral computed tomography (LDCT) can detect early lung cancer patients and reduce lung cancer-related mortality effectively, and pathological diagnosis provides the gold standard for the diagnosis of lung cancer. However, high false positive rate limits the application of LDCT in clinic. Moreover, the pathological diagnosis belongs to the invasive detection method, which causes great pain to the patients. Hence, searching for a convenient and non-invasive method for the diagnosis of NSCLC can play an essential role in improving the prognosis of NSCLC (Grilley-Olson et al., 2013). Serum contains types of molecular markers, and serum test is non-invasive and cheap, so it is one of the best sources of biomarkers. Moreover, many serum markers have been applied to detect lung cancer in clinic, such as CEA and SCC. However, the serum biomarkers are always limited by low sensitivity and specificity.

MiRNA, a small non-protein coding RNA, targets to 3' UTR of mRNA, degrades the mRNA and alters the expression of the gene and protein encoded (Bartel, 2004). MiRNA could be released by cancer cells into serum, plasma, or other body fluids to a participant in carcinogenesis, which have been reported in multiple cancers, such as hepatoma carcinoma, NSCLC (Zhang et al., 2015; Zhao et al., 2016; Usuba et al., 2019; Valihrach et al., 2019). As a biomarker, serum miRNAs have particular merits: resistant to RNase digestion and repeated freezing and thawing (Chen et al., 2008; Mitchell et al., 2008).

Several serum miRNAs with differential expression in patients with NSCLC and the healthy were reported recently, such as miR-16 (Fan et al., 2016), miR-504 (Szpechcinski et al., 2019), and miR-21 (Zhao et al., 2015). As a favorable biomarker for the disease, the miRNA should possess certain biological functions. However, most previous studies merely expound the differential expression of the miRNA in NSCLC and the healthy, but rarely explore the biological function of the miRNA in NSCLC.

The present study screened serum miRNA expression profiles (2,549 miRNAs) in 20 serum samples, and the selected miRNAs were validated in a cohort (NSCLC, $n = 30$; the healthy, $n = 30$). Moreover, we identified the expression level of tissue miR-4687-3p in lung adenocarcinoma (LUAD) and squamous cell lung carcinoma (LUSC) based on the TCGA database. Furthermore, we explored its function on the proliferation, invasion, and migration of NSCLC cells. The study design is illustrated in **Figure 1**.

MATERIALS AND METHODS

Study Participants

In this study, 50 patients with NSCLC (**Table 1**) were recruited at the time of diagnosis, before any medical or surgical treatment, from the First Affiliated Hospital of Zhengzhou University (Zhengzhou, China) between January 2018 and December 2018. Moreover, 50 healthy participants (**Table 1**) were collected from health examination populations without pulmonary-related diseases or other cancers from Zhengzhou Hospital of Traditional Chinese Medicine (Zhengzhou, China). Written informed consent was obtained from all subjects (NSCLC patients and healthy controls) before the study began, and the study protocol was approved by Medical Ethics Committee of Zhengzhou University (Zhengzhou, China). The detailed information of the participants from training (NSCLC, $n = 20$; the healthy, $n = 20$) and validation cohort (NSCLC, $n = 30$; the healthy, $n = 30$) were illustrated in **Supplementary Tables 1**, **2** respectively.

After centrifugation at 3,000 rpm for 5 min, the serum sample was divided into 500 μL/tube and stored at −80°C immediately to avoid repeated freezing and thawing.

Microarray Assay

Human miRNA microarray 2.0 from Agilent Technologies (Santa Clara, CA) including 2,549 miRNAs were applied to screen candidate miRNAs for diagnosing NSCLC in 20 serum samples. The aim was to reduce bias caused by sample heterogeneity, and sera from two participants were merged as a serum sample. Total RNA was harvested using TRIzol (Invitrogen) and the RNeasy Mini Kit (Qiagen), and then labeled and hybridized on the human miRNA microarray. After microarray washing

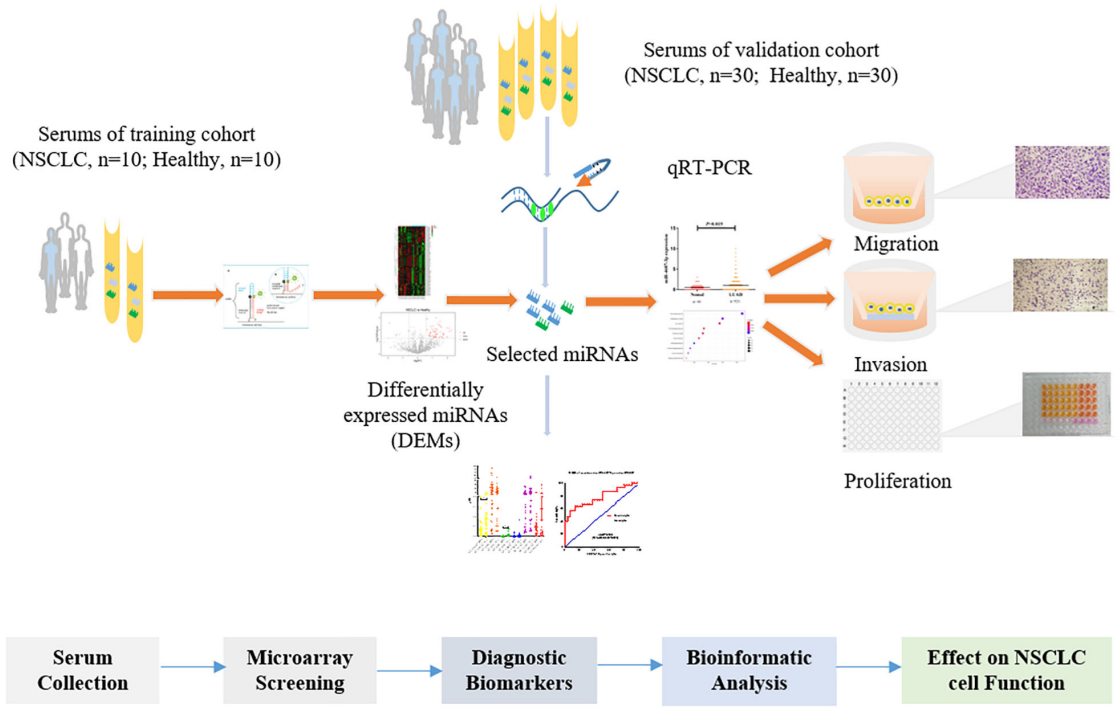

FIGURE 1 | An overview of the workflow of the study design.

TABLE 1 | Characteristics of study participants in the training and validation cohort.

Factors	Subgroup	Training cohort No. of participants (%)	Validation cohort No. of participants (%)
Healthy		20	30
Age, year			
	Mean	53.9	54.1
		4.3	12.0
Sex			
	Male	10 (50%)	14 (46.7%)
	Female	10 (50%)	16 (53.3%)
NSCLC		20	30
Age, year			
	Mean	56.1	60.8
		5.9	9.5
Sex			
	Male	10 (50%)	19 (63.3%)
	Female	10 (50%)	11 (36.7%)
Histopathological type			
	LUSC	6 (30%)	13 (43.3%)
	LUAD	14 (70%)	17 (56.7%)
Stage			
	I+II	5 (25%)	1 (3%)
	III+IV	10 (50%)	24 (80%)
	Undetermined	5 (25%)	5 (17%)

LUSC, Lung Squamous Carcinoma; LUAD, Lung Adenocarcinoma.

and scanning, the data were extracted with Agilent Feature Extraction Software. The assay was conducted by KangChen Bio-tech (Shanghai, China).

RNA Extraction and qRT-PCR

Total RNA was harvested from the serum samples using TRIzol (Takara, Japan) and then used to synthesize cDNA by the stem-loop method using the PrimeScript RT Master Mix Kit (Promega, United States). qRT-PCR was carried out by using an ABI Q3 system (Applied Biosystems, Foster City, CA). The primers used in the study were purchased from Sangya Corporation (Sangya, China) or RiboBio Corporation (Ribo, China). Thereinto, U6 primer was using the Bulge-Loop miRNA qRT-PCR Primer Set (RiboBio, China) according to the manufacturer's instructions.

U6 and serum miR-484 (Zhou et al., 2015) expression were used as a stable endogenous control for normalization in cells and serum samples, respectively.

TCGA Data Analysis

The Cancer Genome Atlas (TCGA) contains multiple primary cancer and corresponding normal samples, including lung cancer. There are two datasets for lung cancer (TCGA LUAD and TCGA LUSC), which include LUAD (n = 521) with corresponding normal tissues (n = 46), and LUSC (n = 478) with corresponding normal tissues (n = 45), respectively. After we downloaded and normalized the non-transcriptome expression data of TCGA lung cancer, we compared tissue miR-4687-3p expression in LUAD, LUSC, and corresponding normal tissue[1].

[1]http://cancergenome.nih.gov

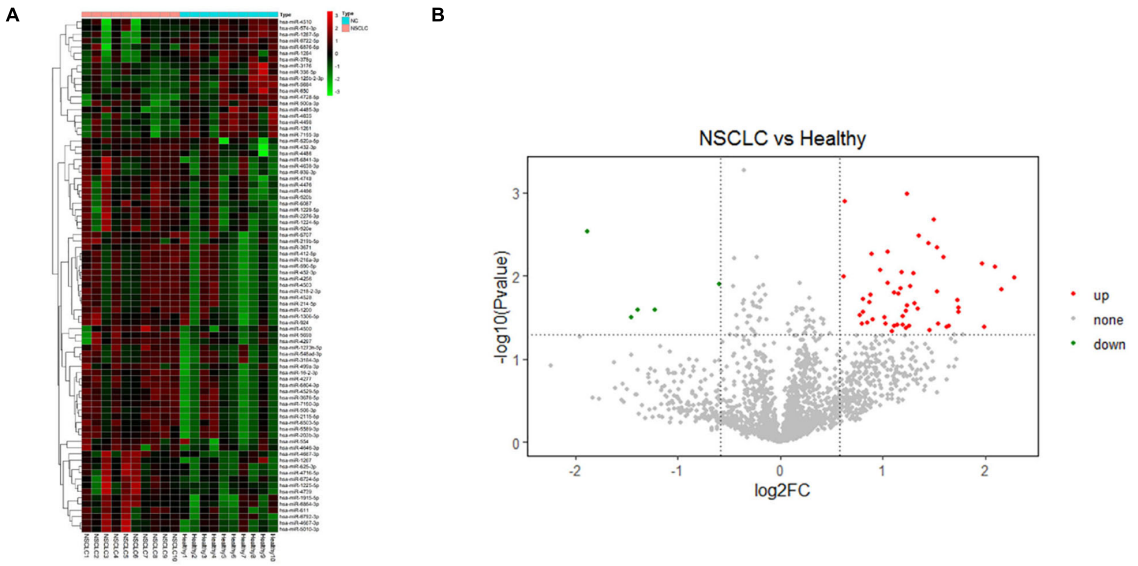

FIGURE 2 | (A) Hierarchical clustering and **(B)** Volcano Plot of the differentially expressed miRNAs (DEMs) in the comparison of NSCLC and the healthy (fold change > 1.5, $p < 0.05$).

TABLE 2 | Six differentially overexpressed serum miRNAs in NSCLC.

Name	*P*-value	Fold change	NSCLC (Raw)	Healthy (Raw)
MiR-1915-5p	0.027	3.3	10.2	6.6
MiR-432-3p	0.042	2.3	25.5	18.6
MiR-4488	0.022	2.5	23.7	16.4
MiR-4687-3p	0.010	1.5	470.8	301.1
MiR-520a-5p	0.040	3.1	28.3	23.2
MiR-6087	0.001	1.5	2764.2	1772.4

Target Prediction and Enrichment Analysis

TargetScan was used to perform the target prediction of miR-4687-3p. For the target genes, GO and KEGG enrichment analysis were conducted by R software (ClusterProfiler package) to clarify the potential function of miRNA.

Cell Culture

Human NSCLC cell lines A549, PC-9, H1299, H1975 were cultured in RPMI 1640 (BI, Israel) medium with 10% FBS (BI, Israel) and CALU-3 in DMEM (BI, Israel) medium with 10% FBS. These cell lines were grown in a cell incubator, and the medium was replaced every 2 days.

MiRNA Transfection

MiR-4687-3p mimics, inhibitors, and corresponding controls were obtained from RiboBio (RiboBio, China). Cells were added into six-well plates with 2.5×10^5 cells per well and cultured overnight. When the cell confluence reached 30–40%, cells were transfected with inhibitor negative control (concentration: 100 nM), miR-4687-3p inhibitor (concentration: 100 nM), or mimic negative control (concentration: 50 nM), miR-4687-3p mimic (concentration: 50 nM), respectively. Lipofectamine^{TM}3000 (Invitrogen, United States) was used as the transfection reagent.

CCK-8 Cell Proliferation Assay

Cells were seeded into 96-well plates ($3–4 \times 10^3$ cells/well) and supplemented with corresponding transfection reagents after 6–8 h. Each group included six parallel wells. After cells

TABLE 3 | Primer sequence of the selected six miRNAs and reference miRNA.

Name	Specific reverse primer	Forward primer	Reverse primer
MiR-1915-5p	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGGCCCG	ATATCGACCTTGCCTTGCTGCC	AGTGCAGGGTCCGAGGTATT
MiR-432-3p	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGACAT	AATCCGCTGGATGGCTCCTCC	AGTGCAGGGTCCGAGGTATT
MiR-4488	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCGCCGG	ATATATCGAGGGGGCGGGCT	AGTGCAGGGTCCGAGGTATT
MiR-4687-3p	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGCCTGC	ATATCCGTGGCTGTTGGAGGGG	AGTGCAGGGTCCGAGGTATT
MiR-520a-5p	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGAAAG	CCGCGCTCCAGAGGGAAGTA	AGTGCAGGGTCCGAGGTATT
MiR-6087	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGCTCGC	ATATATCGTGAGGCGGGGGG	AGTGCAGGGTCCGAGGTATT
MiR-484	CTCAACTGGTGTCGTGGAGTCGGCAATTCAGTTGAGATCGGGAG	ACACTCCAGCTGGGTCAGGCTC AGTCCCCT	TGGTGTCGTGGAGTCG

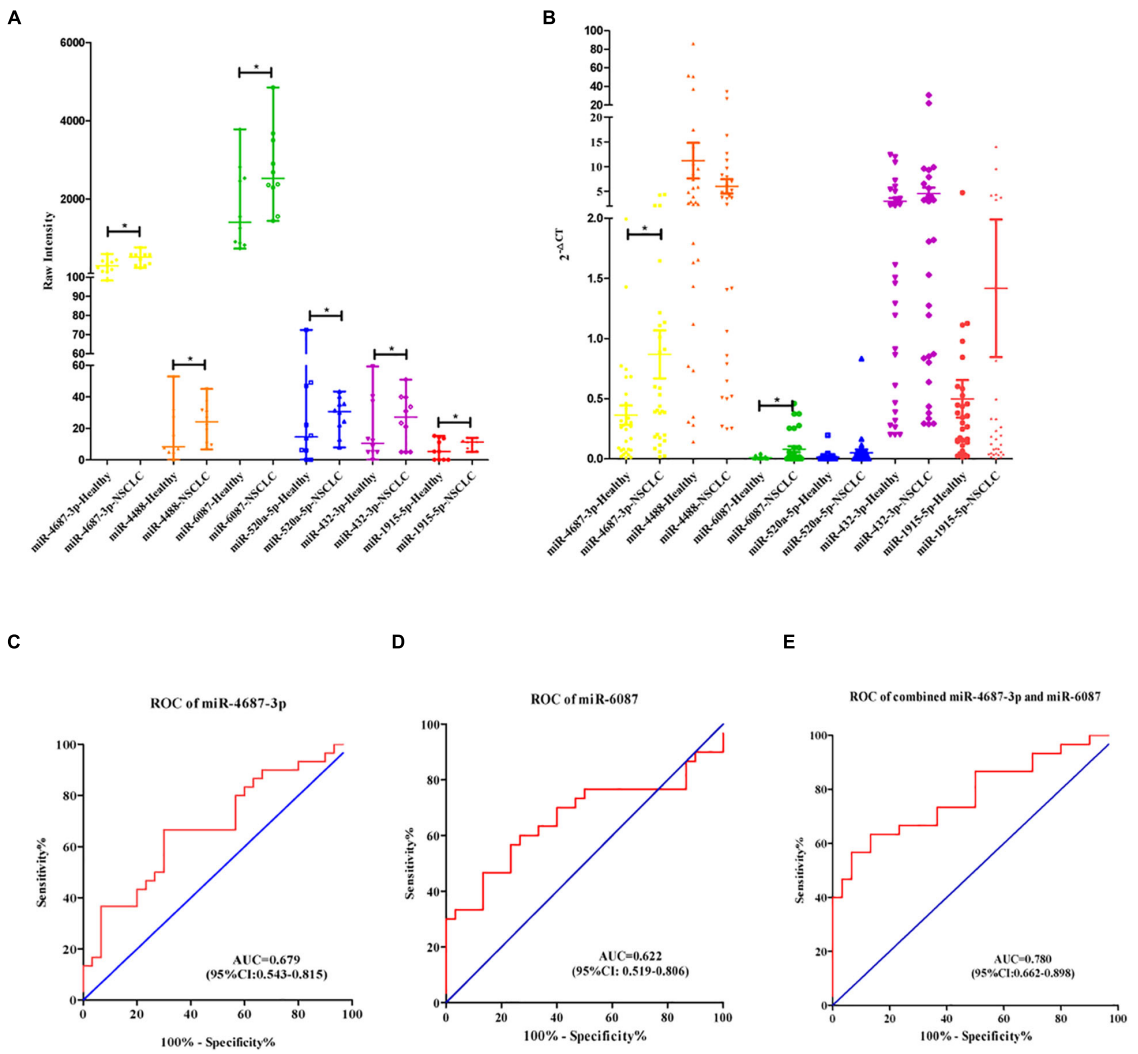

FIGURE 3 | (A) Scatter diagram exhibited the row intensity of six miRNAs in NSCLC serum samples (n = 10) and healthy serum samples (n = 10) in training cohort. **(B)** Scatter diagram exhibited the relative expression of six miRNAs of NSCLC serum samples (n = 30) and the healthy serum samples (n = 30) in validation cohort. **(C–E)** Receiver operating characteristic curve analysis of miR-4687-3p, miR-6087, and combined with the two miRNAs for NSCLC diagnosis in the validation cohort.

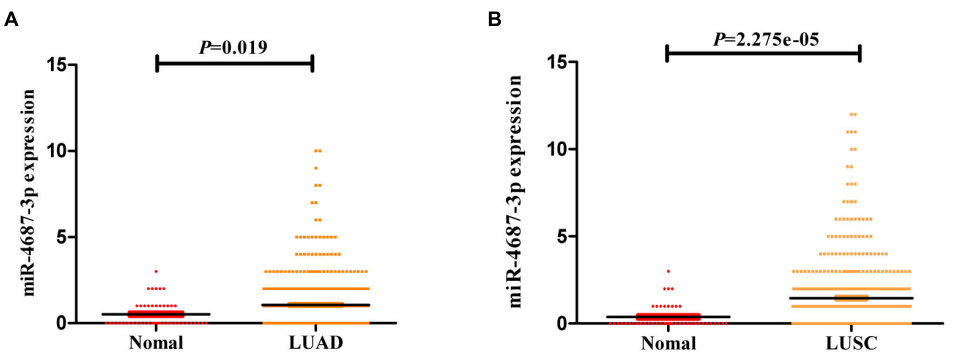

FIGURE 4 | (A) The expression of miR-4687-3p in LUAD tissues (n = 521) and normal lung tissues (n = 46). **(B)** The expression of miR-4687-3p in LUSC tissues (n = 478) and normal lung tissues (n = 45). Data were from TCGA database. LUAD, Lung Adenocarcinoma; LUSC, Lung Squamous Cell Carcinomas.

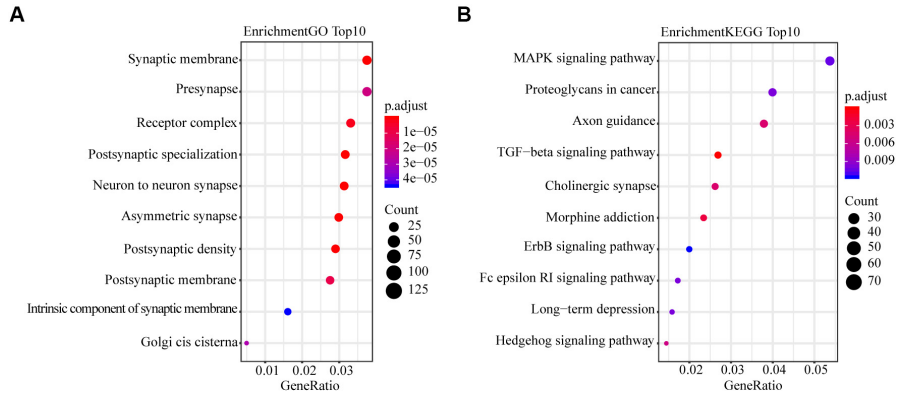

FIGURE 5 | (A) Plot of the enriched GO terms. Go enrichment analysis for miR-4687-3p-related mRNAs. **(B)** Plot of the KEGG pathways. KEGG pathway enrichment analysis for miR-4687-3p-related mRNAs. The bubble color and size represent enrichment significance and the number of target mRNAs enriched in a GO term or pathway, respectively. $p < 0.05$ was used as the threshold to select GO and KEGG terms.

were incubated at 37°C for 24, 48, and 72 h, 100 μL CCK-8 mixture (Dalian Mellon, China) (CCK-8: DMEM/RPMI1640 medium = 1:9) was added into the wells. The absorbance at 450 nm was measured by microplate reader.

Transwell Invasion and Migration Assay

After incubation for 24 h, the transfected cells were trypsinized and resuspended in DMEM/RPMI 1640 without FBS. Then 70 μL of Matrigel (BD, United States) (Matrigel: DMEM/RPMI1640 medium = 1:9) was added into the upper chamber and the transwell plate (Corning, United States) was placed into the incubator for 30 min. After that, excess unset Matrigel was removed followed by adding 200 μL cell suspension (5×10^4 CALU-3 cells or 10^5 PC-9 cells) into the upper chamber when the lower chamber was filled with DMEM/RPMI1640 medium containing 20% FBS. The transwell chamber was then placed into a 24-well transwell plate. After 24 h of incubation, the chamber was washed by PBS, fixed with 4% paraformaldehyde for 15 min, and stained with 0.1% crystal violet (Solarbio, China) for 30 min. Finally, we photographed the cells under the microscope and counted the number of cells that passed through the filter membrane in five random fields.

The process of migration assay was similar to the invasion assay, except for laying Matrigel into the chamber. In addition, in the migration assay, we added 1.5×10^4 CALU-3 or 5×10^4 PC-9 cells into the upper chamber, respectively.

Statistical Analysis

The differential expression of miRNAs between NSCLC and healthy group in the microarray analysis and qRT-PCR analysis was analyzed by Mann-Whitney unpaired test. Independent sample T test and analysis of variance (ANOVA) were used to analyze the data from cell function experiments. GraphPad prism 5 software (La Jolla, CA, United States) was used to present the data. The predicted probability of being diagnosed with NSCLC was used as a surrogate marker to construct receiver operating characteristic (ROC) curve. Area under the ROC curve (AUC) was used as an accuracy index for evaluating the diagnostic performance of the selected miRNAs. The ROC and regression analysis were performed by SPSS 19.0 software (IBM, United States). Two sides $P < 0.05$ was considered statistically significant.

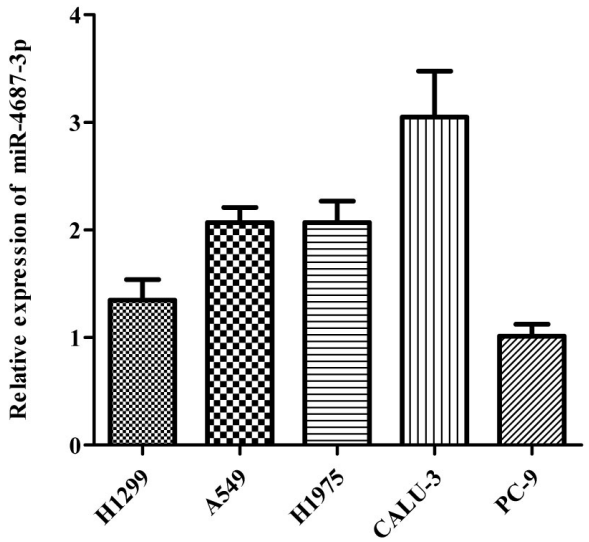

FIGURE 6 | Comparison of miR-4687-3p in five NSCLC cells. The PC-9 cells were used as the control.

RESULTS

MiRNA Screening

A microarray, including probes for 2,549 human miRNAs, was used to screen the significant differential miRNAs (DEMs) (fold change > 1.5, $p < 0.05$) between the NSCLC and the healthy, which were displayed through the hierarchical clustering and Volcano Plot in **Figure 2**. We obtained 53 up-regulated DEMs and five down-regulated DEMs in NSCLC, compared with the

healthy (**Supplementary Table 3**). Taken fold change > 1.5, $p < 0.05$, raw value > 5 as a criterion, we screened six miRNAs (**Tables 2**, **3**) significantly up-regulated in NSCLC, compared with the healthy. Then, we detected the expression of the selected miRNAs in the validation cohort via qRT-PCR.

Expression of the Six Selected MiRNAs

Expression of the selected six individual miRNAs according to the training cohort (**Figure 3A**) was evaluated with qRT-PCR in 60 serum samples (NSCLC, $n = 30$; the healthy, $n = 30$) (**Figure 3B**). Serum miR-4687-3p and miR-6087 owned a high expression

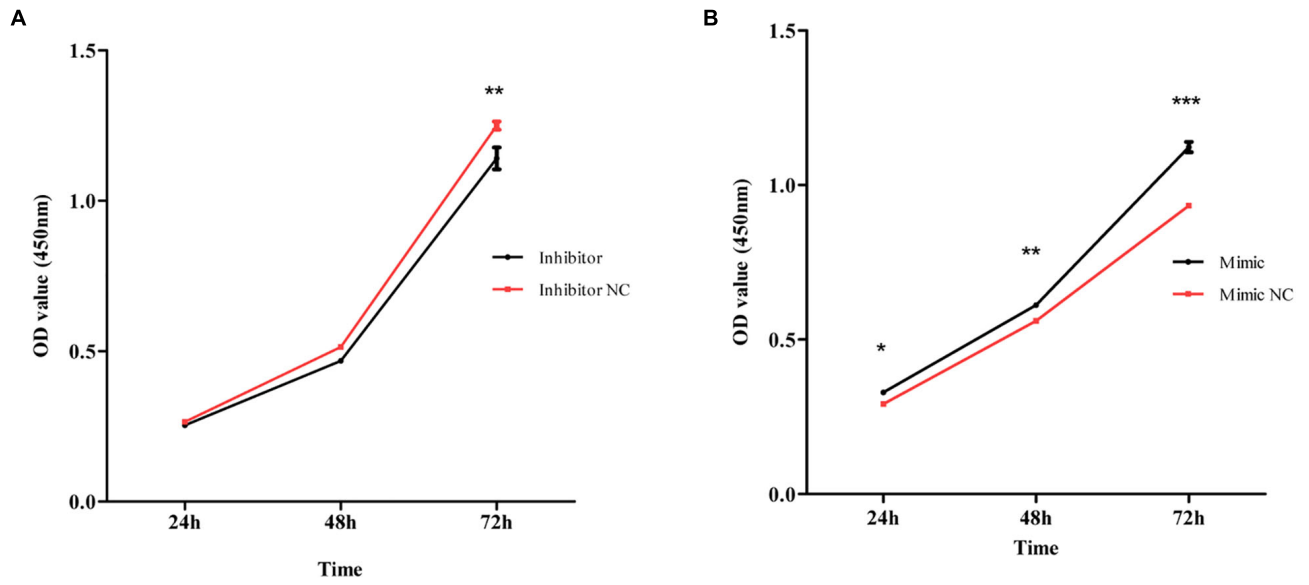

FIGURE 7 | (A) CALU-3 cells were transfected with inhibitor control and miR-4687-3p inhibitor, respectively. Cell proliferation was evaluated by the CCK-8 assay at 24, 48, and 72 h. **(B)** PC-9 Cells were transfected with mimic control and miR-4687-3p mimic, respectively. Cell proliferation was evaluated by the CCK-8 assay at 24, 48, and 72 h. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

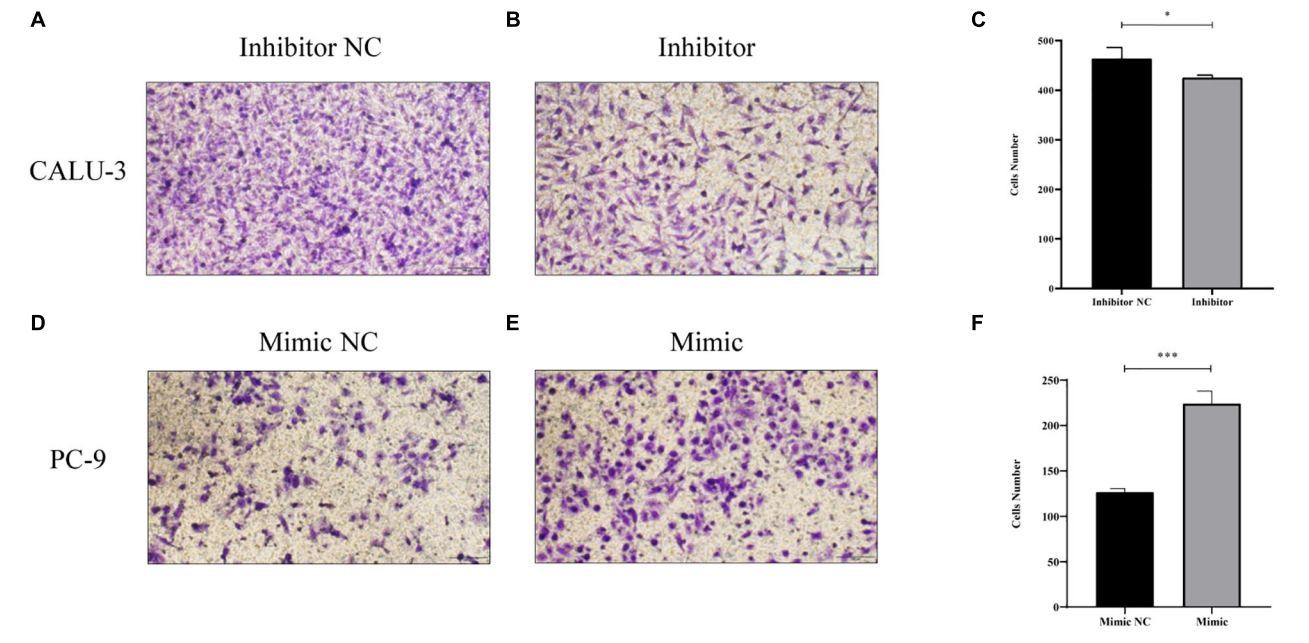

FIGURE 8 | (A–C) Transwell invasion assay with Matrigel was performed in miR-4687-3p inhibitor or inhibitor control transfected CALU-3 cells (Magnification ×400). **(D–F)** Transwell invasion assay with Matrigel was performed in miR-4687-3p mimic or mimic control transfected PC-9 cells (Magnification ×400). *$p < 0.05$, ***$p < 0.001$.

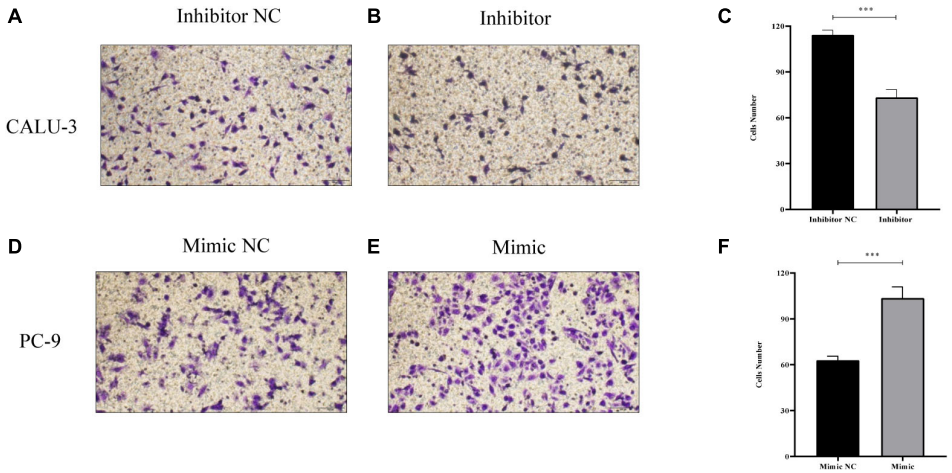

FIGURE 9 | (A–C) Transwell migration assay without Matrigel was performed in miR-4687-3p inhibitor or inhibitor control transfected CALU-3 cells (Magnification ×400). **(D–F)** Transwell migration assay without Matrigel was performed in miR-4687-3p mimic or mimic control transfected PC-9 cells (Magnification ×400). ***$p < 0.001$.

level in NSCLC compared with the healthy. In addition, miR-4687-3p provided a higher diagnostic accuracy of NSCLC than the other five miRNAs (AUC = 0.679, 95% CI: 0.543–0.815) (**Figures 3C,D**). The AUC of the combination of miR-4687-3p and miR-6087 using logistic regression model reached 0.780 (95%CI: 0.662–0.898) (**Figure 3E**).

The Expression of Tissue MiR-4687-3p Was Validated in TCGA Database

According to the TCGA database of NSCLC, the expression of the tissue miR-4687-3p was higher both in LUAD (**Figure 4A**) and LUSC (**Figure 4B**) than corresponding normal tissues ($p < 0.05$).

GO and KEGG Pathway Enrichment

To explore the mechanism in the process of miR-4687-3p regulating NSCLC cell progression, we predicted the target genes of miR-4687-3p by using TargetScan, which provided 3,851 target genes. Meanwhile, we proceeded with GO and KEGG pathway enrichment analysis for the 3,851 target genes, which showed that postsynaptic specialization and TGF-β signaling pathway were significantly enriched (**Figures 5A,B**).

MiR-4687-3p Promoted the Proliferation, Invasion, and Migration of the NSCLC Cells

MiR-4687-3p expression was detected in five NSCLC cell lines (A549, PC-9, CALU-3, H1299, H1975), which showed that the CALU-3 and PC-9 cells were the highest and lowest observed expression of miR-4687-3p, respectively (**Figure 6**). Therefore, CALU-3 and PC-9 cells were selected for the next experiments.

MiR-4687-3p inhibitor and miR-4687-3p mimic was transfected into CALU-3 and PC-9 cells respectively, to down-regulate or up-regulate the miR-4687-3p level. The results showed that down-regulation of miR-4687-3p with inhibitor could markedly suppress the proliferation (**Figure 7A**), invasion (**Figures 8A–C**), and migration (**Figures 9A–C**) in CALU-3 cells compared to inhibitor NC, respectively. Overexpressed miR-4687-3p by using mimic could promote the proliferation (**Figure 7B**), invasion (**Figures 8D–F**), and migration (**Figures 9D–F**) of PC-9 cells, compared to the mimic NC, respectively.

DISCUSSION

In the study, serum miR-4687-3p was discovered up-regulated in NSCLC, compared to the healthy, and showed a remarkable differential diagnosis value for the first time. Zsófia Brigitta Nagy revealed that the plasma miR-4687-3p overexpressed in colorectal cancer when compared with tubulovillous adenoma, via microarray screening (Nagy et al., 2019). However, they did not conduct the further validation.

For diagnosis of NSCLC, compared with biopsy, the serum-based approach was found to have more advantages, such as easy to access and acceptable to patients. MiRNAs were demonstrated to play an important role in carcinogenesis and have the potential for diagnosis of multiple cancers (Cheng, 2015; Maia et al., 2015; Gao et al., 2016; Li et al., 2017; Romano and Kwong, 2018). Moreover, endogenous circulating miRNAs have attracted widespread attention (de Souza et al., 2017).

In the training cohort, microarray assay was used to screen the serum differentially expressed miRNAs (DEMs) between NSCLC patients and the healthy, directly. Furthermore, we sorted 60 serum samples to verify the differential expression and found that serum miR-4687-3p and serum miR-6087 owned a higher expression in the NSCLC group than that of the healthy. As to differential diagnostic ability, serum miR-4687-3p (AUC = 0.679) might be a potential diagnostic biomarker for NSCLC.

Abnormally expressed serum cancer-related miRNAs were usually closely related to its expression in cancer tissues. According to the TCGA NSCLC database, compared to the

corresponding normal lung tissues, the miR-4687-3p was up-regulated in LUAD and LUSC tissues, which indicated that miR-4687-3p was NSCLC-related miRNA.

Otherwise, as a favorable NSCLC diagnostic biomarker, it is vital to certain biological function to participate in the occurrence and development of the NSCLC. In addition, serum miR-4687-3p has not been reported as a diagnostic biomarker in several malignancies. Research of miR-4687-3p participating in the occurrence and development of tumors has not been reported yet. Although the specific mechanism of the miRNA on mRNA was unascertained, it was explicated that miRNAs triggered target mRNA degradation by complementary combining to 3' UTR of mRNA.

By bioinformatics analysis, we revealed that miR-4687-3p owned 3,851 target genes, which significantly enriched the TGF-β signaling pathway. Previous researches had reported that the TGF-β pathway regulated the invasion and migration of cancer cells via epithelial-mesenchymal transition (EMT) (Xu et al., 2009; Schneider et al., 2011; Ma et al., 2015; Zhao et al., 2018; Camerlingo et al., 2019) and TGF-β could form a complex with Smad4 and translocate into the nucleus to regulate gene transcription (Wang et al., 2015; Yang et al., 2019), which suggested that the TGF-β pathway was vital in carcinogenesis. Some specific molecules served as tumor promoter or suppressor and can regulate the cancer cells proliferation, invasion, and migration via the pathway (Colak and Ten Dijke, 2017; Seoane and Gomis, 2017). In the present study, the target genes enriched in the TGF-β pathway implied that the miR-4687-3p might serve as a promoter affected the carcinogenesis of NSCLC. Moreover, the results proved the above hypothesis. Down-regulation miR-4687-3p could suppress the proliferation, invasion, and migration of NSCLC cells, while overexpression miR-4687-3p had the opposite effects. Importantly, data from the present study revealed that miR-4687-3p as a tumor promoter could promote tumor growth.

In conclusion, our results strongly suggested that serum miR-4687-3p overexpressed in NSCLC and miR-4687-3p could promote the growth and migration of NSCLC cells, which demonstrated that serum miR-4687-3p could be a novel and favorable biomarker for NSCLC. Compared with those studies of circulating miRNAs in diagnosing NSCLC, our study is unique for the following reasons. First, we screened a large number of serum miRNAs via microarrays, which enabled us to identify potential diagnostic serum miRNAs. Furthermore, we compared the expression of tissue miR-4687-3p in NSCLC and the healthy based on the TCGA database. We found that miR-4687-3p possesses the potency to promote NSCLC cells growth, migration, and invasion. Our findings suggest that miR-4687-3p functions as a tumor promoter in NSCLC and holds promise as a prognostic biomarker and potential therapeutic target for NSCLC.

However, our study has some limitations. Currently, there is no standard endogenous control for circulating miRNA studies. The stable control (miR-484) needs to be validated in more studies. In this research, we evaluated the role of miR-4687-3p in NSCLC, but we did not identify the underlying molecular mechanisms. In future research, we will focus on the specific mechanism of miR-4687-3p.

DATA AVAILABILITY STATEMENT

We have submitted the microarray data to GEO repository (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157074).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Zhengzhou University. 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

ML conducted the research and writing the manuscript. LD and JL designed the study and contributed to the writing. QS assisting with the research. SO, ZZ, MW, and CZ provided the serum samples. TY, YW, XZ, and WX collected the serum samples and analysis the data. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 8167291), the Leading Talents of Science and Technology Innovation in Central China (Grant No. 20420051008), the Key project of discipline construction of Zhengzhou University (Grant No. XKZDQY202009), and Project of Basic Research Fund of the Henan Institute of Medical and Pharmacological Sciences (Grant No. 2020BP0202).

ACKNOWLEDGMENTS

We gratefully acknowledge Henan Key Laboratory of Pharmacology for Liver Diseases for providing experimental platform support and Academy of Medical Science of Zhengzhou University for providing the NSCLC cells in the experiment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.597508/full#supplementary-material

Supplementary Table 1 | The information for samples from training cohort.

Supplementary Table 2 | The information for samples from validation cohort.

Supplementary Table 3 | The information for differently expressed miRNAs between NSCLC and healthy based on the microarray (fold change > 1.5, $P < 0.05$).

REFERENCES

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

Camerlingo, R., Miceli, R., Marra, L., Rea, G., D'Agnano, I., Nardella, M., et al. (2019). Conditioned medium of primary lung cancer cells induces EMT in A549 lung cancer cell line by TGF-ss1 and miRNA21 cooperation. *PLoS One* 14:e0219597. doi: 10.1371/journal.pone.0219597

Chen, X., Ba, Y., Ma, L. J., Cai, X., Yin, Y., Wang, K. H., et al. (2008). Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* 18, 997–1006. doi: 10.1038/cr.2008.282

Cheng, G. F. (2015). Circulating miRNAs: roles in cancer diagnosis, prognosis and therapy. *Adv. Drug Deliver Rev.* 81, 75–93. doi: 10.1016/j.addr.2014.09.001

Colak, S., and Ten Dijke, P. (2017). Targeting TGF-beta signaling in cancer. *Trends Cancer* 3, 56–71. doi: 10.1016/j.trecan.2016.11.008

de Souza, M. F., Kuasne, H., Barros, M. D., Ciliao, H. L., Marchi, F. A., Fuganti, P. E., et al. (2017). Circulating mRNAs and miRNAs as candidate markers for the diagnosis and prognosis of prostate cancer. *PLoS One* 12:e0184094. doi: 10.1371/journal.pone.0184094

Fan, L. H., Qi, H. W., Teng, J. L., Su, B., Chen, H., Wang, C. H., et al. (2016). Identification of serum miRNAs by nano-quantum dots microarray as diagnostic biomarkers for early detection of non-small cell lung cancer. *Tumor Biol.* 37, 7777–7784. doi: 10.1007/s13277-015-4608-3

Gao, Y., Guo, Y., Wang, Z., Dai, Z., Xu, Y., Zhang, W., et al. (2016). Analysis of circulating miRNAs 21 and 375 as potential biomarkers for early diagnosis of prostate cancer. *Neoplasma* 63, 623–628. doi: 10.4149/neo_2016_417

Grilley-Olson, J. E., Hayes, D. N., Moore, D. T., Leslie, K. O., Wilkerson, M. D., Qaqish, B. F., et al. (2013). Validation of interobserver agreement in lung cancer assessment: hematoxylin-eosin diagnostic reproducibility for non-small cell lung cancer: the 2004 World Health Organization classification and therapeutically relevant subsets. *Arch. Pathol. Lab. Med.* 137, 32–40. doi: 10.5858/arpa.2012-0033-OA

Li, G., Fang, J., Wang, Y., Wang, H., and Sun, C. C. (2017). MiRNA-based therapeutic strategy in lung cancer. *Curr. Pharm. Design.* 23, 6011–6018. doi: 10.2174/1381612823666170725141954

Ma, Y., Liu, H., Zhang, H., and Shao, R. G. (2015). [The TGF-beta signaling pathway induced EMT in breast cancer]. *Acta Pharmaceut. Sin.* 50, 385–392.

Maia, B. D., Ling, H., Monroig, P., Ciccone, M., Soares, F. A., Calin, G. A., et al. (2015). Design of a miRNA sponge for the miR-17 miRNA family as a therapeutic strategy against vulvar carcinoma. *Mol. Cell Probe* 29, 420–426. doi: 10.1016/j.mcp.2015.08.002

Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L., et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. U.S. A.* 105, 10513–10518. doi: 10.1073/pnas.0804549105

Nagy, Z. B., Bartak, B. K., Kalmar, A., Galamb, O., Wichmann, B., Dank, M., et al. (2019). Comparison of circulating miRNAs expression alterations in matched tissue and plasma samples during colorectal cancer progression. *Pathol. Oncol. Res.* 25, 97–105. doi: 10.1007/s12253-017-0308-1

Romano, G., and Kwong, L. N. (2018). Diagnostic and therapeutic applications of miRNA-based strategies to cancer immunotherapy. *Cancer Metast. Rev.* 37, 45–53. doi: 10.1007/s10555-017-9716-7

Schneider, D., Tarantola, M., and Janshoff, A. (2011). Morphological and dynamical changes during TGF-beta induced epithelial-to-mesenchymal transition. *Eur. Biophys. J. Biophy.* 40, 163–163.

Seoane, J., and Gomis, R. R. (2017). TGF-beta family signaling in tumor suppression and cancer progression. *Cold Spring Harb. Perspect. Biol.* 9:a022277. doi: 10.1101/cshperspect.a022277

Siegel, R. L., and Miller, K. D. (2019). Cancer statistics, 2019. *CA Cancer J. Clin.* 69, 7–34. doi: 10.3322/caac.21551

Szpechcinski, A., Florczuk, M., Duk, K., Zdral, A., Rudzinski, S., Bryl, M., et al. (2019). The expression of circulating miR-504 in plasma is associated with EGFR mutation status in non-small-cell lung carcinoma patients. *Cell Mol. Life Sci.* 76, 3641–3656. doi: 10.1007/s00018-019-03089-2

Usuba, W., Urabe, F., Yamamoto, Y., Matsuzaki, J., Sasaki, H., Ichikawa, M., et al. (2019). Circulating miRNA panels for specific and early detection in bladder cancer. *Cancer Sci.* 110, 408–419. doi: 10.1111/cas.13856

Valihrach, L., Androvic, P., and Kubista, M. (2019). Circulating miRNA analysis for cancer diagnostics and therapy. *Mol. Aspects Med.* 72:100825. doi: 10.1016/j.mam.2019.10.002

Wang, X. Y., Chen, X. S., Meng, Q. W., Jing, H., Lu, H. L., Yang, Y. M., et al. (2015). MiR-181b regulates cisplatin chemosensitivity and metastasis by targeting TGF beta R1/Smad signaling pathway in NSCLC. *Sci. Rep.* 5:17618. doi: 10.1038/Srep17618

Xu, J., Lamouille, S., and Derynck, R. (2009). TGF-beta-induced epithelial to mesenchymal transition. *Cell Res.* 19, 156–172. doi: 10.1038/cr.2009.5

Yang, X. Y., Liao, J. J., and Xue, W. R. (2019). FMNL1 down-regulation suppresses bone metastasis through reducing TGF beta-1 expression in non-small cell lung cancer (NSCLC). *Biomed. Pharmacother.* 117:109126. doi: 10.1016/J.Biopha.2019.109126

Zhang, L., Xu, Y., Jin, X., Wang, Z., Wu, Y., Zhao, D., et al. (2015). A circulating miRNA signature as a diagnostic biomarker for non-invasive early detection of breast cancer. *Breast Cancer Res. Treatment* 154, 423–434. doi: 10.1007/s10549-015-3591-0

Zhao, C., Lu, F., Chen, H., Zhao, F., Zhu, Z., Zhao, X., et al. (2016). Clinical significance of circulating miRNA detection in lung cancer. *Med. Oncol.* 33:41. doi: 10.1007/s12032-016-0757-5

Zhao, L. M., Lie, J., Liu, Y. P., Zhou, W., Shan, Y. N., Fan, X. Y., et al. (2018). Flotillin1 promotes EMT of human small cell lung cancer via TGF-beta signaling pathway. *Cancer Biol. Med.* 15, 400–414. doi: 10.20892/j.issn.2095-3941.2018.0053

Zhao, W., Zhao, J. J., Zhang, L., Xu, Q. F., Zhao, Y. M., Shi, X. Y., et al. (2015). Serum miR-21 level: a potential diagnostic and prognostic biomarker for non-small cell lung cancer. *Int. J. Clin. Exp. Med.* 8, 14759–14763.

Zheng, R. S., Sun, K. X., Zhang, S. W., Zeng, H. M., Zou, X. N., Chen, R., et al. (2019). [Report of cancer epidemiology in China, 2015]. *Chinese J. Oncol.* 41, 19–28. doi: 10.3760/cma.j.issn.0253-3766.2019.01.005

Zhou, C., Chen, Z., Dong, J., Li, J., Shi, X., Sun, N., et al. (2015). Combination of serum miRNAs with Cyfra21-1 for the diagnosis of non-small cell lung cancer. *Cancer Lett.* 367, 138–146. doi: 10.1016/j.canlet.2015.07.015

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 December 2020
doi: 10.3389/fgene.2020.562114

MicroRNA-1291 Is Associated With Locoregional Metastases in Patients With Early-Stage Breast Cancer

Daniel Escuin[1]†, Laura López-Vilaró[1,2]†, Olga Bell[1], Josefina Mora[2], Antonio Moral[2,3], José Ignacio Pérez[2], Cristina Arqueros[2], Teresa Ramón y Cajal[2], Enrique Lerma[1,2,3] and Agustí Barnadas[1,2,3,4]**

[1] Institut d'Investigació Biomèdica Sant Pau, Barcelona, Spain, [2] Hospital de la Santa Creu i Sant Pau, Barcelona, Spain, [3] Department of Medicine, Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Spain, [4] Centro de Investigación Biomédica en Red Cáncer, Madrid, Spain

Evidence that microRNAs (miRNAs) regulate the various steps of metastasis is increasing. Several studies have looked at the miRNA expression profile in primary breast tumors but few have compared primary tumor and sentinel lymph node (SLN) metastasis. We correlated the expression of miRNAs with the SLN status and the outcome of axillary lymph node dissection (ALND) in 60 patients with early breast cancer. We profiled the expression of miRNAs in paired breast tumor samples and SLNs using the NextSeq500 Illumina platform and key findings were validated by qPCR. MultiMiR Bioconductor and Reactome pathways analysis were performed to identify target genes and signaling pathways affected by altered expressed miRNAs. Our results show that nine miRNAs were differentially expressed in tumor tissues ($q \leq 0.05$). In tumor samples, a 13.5-fold up-regulation of miR-7641-2 ($q < 0.001$) and a 2.9-fold down-regulation of miR-1291 ($q < 0.001$) were associated with tumors with positive SLNs. However, only down-regulation of miR-1291 ($q = 0.048$) remained significant in paired SLNs samples. Interestingly, a 10.5 up-regulation of miR-1291 in SLNs samples was associated with additional axillary lymph node involvement ($q < 0.001$). The enrichment analyses showed that canonical and non-canonical WNT pathways and negative regulation of various receptor tyrosine kinases signaling pathways were targets of miR-1291 and supports the role of miR-1291 as a tumor suppressor gene (TSG). Further studies are warranted to investigate the use of miR-1291 as a surrogate biomarker of SLN node metastasis in patients with early-stage breast cancer.

Keywords: axillary lymph node dissection, breast cancer, sentinel lymph node, metastasis, microRNAs

OPEN ACCESS

Edited by:
Dongqing Wei,
Shanghai Jiao Tong University, China

Reviewed by:
Leda Torres,
National Institute of Pediatrics, Mexico
Jesús Espinal-Enríquez,
Instituto Nacional de Medicina Genómica (INMEGEN), Mexico

***Correspondence:**
Daniel Escuin
descuin@santpau.cat
Agustí Barnadas
abarnadasm@santpau.cat

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *20 May 2020*
Accepted: *10 November 2020*
Published: *02 December 2020*

Citation:
Escuin D, López-Vilaró L, Bell O, Mora J, Moral A, Pérez JI, Arqueros C, Ramón y Cajal T, Lerma E and Barnadas A (2020) MicroRNA-1291 Is Associated With Locoregional Metastases in Patients With Early-Stage Breast Cancer. Front. Genet. 11:562114. doi: 10.3389/fgene.2020.562114

INTRODUCTION

MicroRNAs (miRNAs) are a small (19–25 nt) non-coding RNA (ncRNA), expressed in a wide variety of organisms and highly conserved across species. MiRNAs regulate the expression of target genes by binding to complementary regions of messenger transcripts to repress their translation or regulate their degradation. The number of human miRNAs is currently over 1900 and the number of predicted target genes are in the range of thousands, with some estimates indicating that miRNAs target over 30% of the human genome. The overall emerging picture is that of a complex regulation level of gene expression, in which a single miRNA may control hundreds of targets (Condrat et al., 2020).

Several studies have shown that the expression signatures of miRNAs in human cancers are distinct from those in normal tissues (Janssen et al., 2010). In breast cancer, various studies have revealed a deregulation of miRNAs with clusters of miRNAs frequently being either over-expressed or down-regulated (Nana-Sinkam and Croce, 2011; Condrat et al., 2020). In addition, many cellular pathways are affected by the regulatory function of miRNAs and several human pathologies including cancers, have been associated with deregulation of the miRNAs (Dai and Ahmed, 2011) and play a pivotal role in various steps of the metastatic process (Nicoloso et al., 2009; Valastyan, 2012). Furthermore, analysis of miRNA expression correlated these with various clinicopathological factors (Blenkiron et al., 2007). However, few studies have compared primary tumor and lymph node (LN) metastasis (Smeets et al., 2011; Gravgaard et al., 2012; Cascione et al., 2013). Gravgaard et al. (2012) found a differential expression of miRNA-200 family and miRNA-9 in LN associated with the metastatic process. Cascione et al. (2013) reported analysis of 173 formalin-fixed paraffin embedded (FFPE) tumors and 53 matched LN in TNBC. They found two miRNA signatures that were independent predictors for overall survival (OS) and distant-disease free survival, respectively. In another study, Smeets et al. (2011) identified eight with measurable differences in gene and miRNA expression between N0 and N+ patients, suggesting that LN involvement is not a genetically random process.

Herein, we studied the miRNA expression profile in paired tumor and sentinel lymph node (SLN) from patients with early breast cancer. We used the one-step nucleic amplification (OSNA) assay (Tsujimoto et al., 2007) to accurately measure total metastatic volume in the SLN (Cserni, 2012), as an alternative to intraoperative microscopy-based pathological assessment of the SLN. The OSNA assay is a rapid molecular detection of SLN metastasis based on the semi-quantification of cytokeratin 19 (CK19) mRNA copy numbers (Tsujimoto et al., 2007). The CK19 mRNA copy number, also defined as total tumor load (TTL), has been shown to be a statistically significant parameter in predicting the risks of further positive axillary LN (aLN). Thus, only patients diagnosed with more than two macrometastatic SLN are further treated with axillary lymph node dissection (ALND) (Peg et al., 2013), the golden standard procedure for invasive breast cancer. The aLN status is the most powerful prognostic factor in breast cancer and knowledge of this is essential for making decisions about adjuvant therapy (Shek and Godolphin, 1988). However, ALND has been questioned in recent years because of inherent morbidity following the procedure without directly contributing to survival (Giuliano et al., 2010; Jagsi et al., 2014).

In this study, we investigated the potential use of miRNAs as surrogate biomarkers for the presence of metastases in the SLN and the outcome of ALND, in patients with early-stage breast cancer. We found a differential expression of various miRNAs associated with both the metastatic status of the SLN and the occurrence of further aLN metastases. This study provides a new framework to study the role of miRNAs in the regulation of tumor metastases and their impact on patient outcome.

MATERIALS AND METHODS

Ethics Approval, Consent to Participate and Institutional Safety Procedures

This study was conducted according to the Declaration of Helsinki principles, with approval from the Clinical Research Ethics Committee at "Institut d'Investigacions Biomèdiques Sant Pau" (IIB Sant Pau). Written informed consent was obtained from all patients under institutional review board-approved protocols. All methods were performed in accordance with the relevant guidelines and regulations.

Patient and Study Samples

This study included 60 patients with early-stage breast cancer treated with surgery. None of the patients had prior treatment with chemotherapy or radiation. Patients were chosen on the basis of paired biological samples availability and confirmed diagnose based on the histopathology of primary tumors and OSNA-diagnosed SLNs. Information about completion of ALND was available for all patients. Samples included hormone receptor (HR) positive [estrogen receptor (ER) or progesterone receptor (PR) positive], HER2 positive and triple negative (TN) tumors. We collected clinical and pathological parameters and clinical follow-up. Tumor stage was determined according to the AJCC/UICC system (Webber et al., 2014) and histological grade was determined using the Elston–Ellis grading system (Elston and Ellis, 1991). We collected paired tumor and SLNs for all 60 patients (n = 120 samples). All samples were classified according to the SLN status as negative (n = 20) or positive (n = 40). Positive samples were sub-classified as macrometastatic (n = 20) or micrometastatic (n = 20) (Tsujimoto et al., 2007).

Sample Processing

Primary tumor samples were collected and processed within 30 min after surgery. A portion of each tumor was rapidly embedded in Tissue-Tek OCT (Sakura Finetek, Netherlands) and frozen using a histobath (Thermo Fisher Scientific, Waltham, MA, United States). Frozen tumor blocks were thin-sectioned and stained with hematoxylin/eosin and only those that were judged to contain at least 70% viable tumor by area were carried on for RNA extraction. Intraoperative SLN were evaluated by OSNA assay (Tsujimoto et al., 2007). The entire LN was homogenized in 4 mL of a lysis solution (Sysmex, Spain) and centrifuged at 10,000 × g at room temperature. A 2 μL sample of the supernatant was analyzed with the RD-100i system (Sysmex) and the remaining SLN homogenate was stored immediately at −80C.

RNA Isolation

Total RNA was purified using TRIzol (Invitrogen) and mirVana (Ambion) reagents according to manufacturer's instructions with some modifications. Briefly, Trizol homogenates were mixed with 1/10 volume of mirVana homogenizing solution to retain low molecular weight RNA species. The clear fraction containing total RNA was then purified using the mirVana miRNA isolation kit,

following manufacturer's instructions. Total RNA was eluted in 100 μl nuclease-free water.

RNA-Sequencing

Total RNA concentration and purity was measured using the ribogreen RNA assay kit (Life Technologies) and the integrity was visualized with a TapeStation Bioanalyzer (Agilent Technologies, Inc.). Next, RNA pools were precipitated overnight with 2× volumes of absolute ethanol and 0.1× volume of 0.3 M sodium acetate at −80°C for cDNA library construction. Double-stranded cDNA libraries were constructed using the NEBNext Small RNA library Prep Set for Illumina (New England Biolabs, Ipswich, MA, United States) following manufacturer's instructions. A quality check (QC) and size selection of the PCR amplified cDNA construct was performed using 6% polyacrylamide gel. Two biological replicates for each developmental stage were separately sequenced by the MiSeq (Illumina, San Diego, CA, United States) platform using sequenced runs of 2 × 75 paired-end reads and 1000× coverage to ensure proper quantification of the miRNA expression. A total of 59 tumors and 58 SLNs (117 paired-end, 2 × 75) sequences were successfully sequenced.

Genome Annotation and Quantification of MiRNAs

Paired-end (forward-reverse) sample merging and initial bioinformatics analysis were performed with the CLCBio Genomics Workbench® version 8.0.2 (Qiagen, Germany). A total of 234 fastq input files were generated and used in the analyses. The CLCBio software was used to align and map the trimmed reads to the human and mouse miRBases (version 19) and the Homo_sapiens.GRCh37.57 tracks from Ensembl. Up to two mismatches were allowed on the sequences. Mapping options were set as the program's default. Count tables were generated with R programming language (R Development Core Team, 2011) and the EdgeR package (Bioconductor repository) (Robinson et al., 2010), using non-specific filtering for sequences having a reads-per-million value higher than 0, in at least half of the samples included on each experimental comparison. Transcript per million (TPM) was used as a normalization procedure to correct for differences in sequencing depth and to quantified RNA species.

Differential Expression Analysis

Differential expression analyses were performed using the trimmed mean of M-values normalization method (TMM) (Robinson and Oshlack, 2010), based on the log-fold and absolute gene-wise changes in expression between samples. Differential expression analysis was performed using the EdgeR statistical software package (Bioconductor[1]). Principal component analysis (PCA) was performed using R programming and TMM-normalized quantifications from defined collections of samples as input. Volcano plots were constructed plotting the *p*-value (−log10) on the *y*-axis and the expression fold change (log2) between the two experimental groups on the *x*-axis. Wherever indicated, we have used fold regulation throughout the text to represent positive FC values as up-regulation (fold regulation is equal to 2^{FC}) and negative FC values to indicate a down-regulation (fold regulation is equal to 2^{-FC}).

Quantitative Real-Time RT–PCR Validation Analysis

Selected miRNAs were validated by quantitative real-time RT–PCR (qPCR) in an ABI Prism 7500 Sequence Detection System using specific LNA PCR primers (Exiqon). The cDNA was constructed using the miRCURY LNA™ Universal RT cDNA Synthesis Kit (Exiqon), diluted 40× and assayed in 10 μl PCR reactions according to manufacturer's instructions. Each qPCR was assayed in triplicates and a no-template control (NTC) of water was purified and profiled like the rest of the samples. Analysis of the data was performed using the relative miRNA expression according to the comparative Ct (ΔΔCt) method using negative metastatic samples as reference. We used the geNorm (Andersen et al., 2004) or the Normfinder algorithm (Vandesompele et al., 2002) to select the best combination of two reference genes based on our qPCR data. Data from multiples plates were normalized using UniSp3 spike-in as interplate calibrators.

Gene Targets Prediction and Enrichment Analysis of Gene Targets

The multiMiR Bioconductor's package (Ru et al., 2014) was used to retrieve miRNA-target interactions from 14 external databases[2] and the Reactome pathway database (Jassal et al., 2020) was use to performed enrichment analysis of target pathways and genes.[3]

Statistics

Differentially expressed miRNAs obtained by next generation sequencing (NGS) were detected by an exact test based on conditional maximum likelihood (CML) included in the R Bioconductor package edgeR (Robinson et al., 2010). *p*-Values from NGS were corrected (*q*-values) for multiple testing using the Benjamini–Hochberg method (Benjamini and Hochberg, 1995). A q-value $\leq$ 0.05 was considered significant. In all group comparisons missing expression values were treated as zero. Differences in total numbers of miRNAs between groups were analyzed by two-sided parametric *t*-tests. The analysis of clinicopathological was performed using the Student's *t*-test to compare quantitative variables, and the X2 or Fisher exact tests to compared qualitative variables. Disease free survival (DFS) was defined as the time from diagnosis to date of first relapse (local, regional, contralateral, or metastatic) or second primary cancer. OS was defined as the time from sample collection to death resulting from any cause. Patients lost to follow-up were censored at the last contact. Kaplan–Meier and log-rank analyses were used to compare DFS and OS. Differential

[1]http://www.bioconductor.org/

[2]http://multimir.ucdenver.edu

[3]https://reactome.org/

expression by qPCR of selected miRNAs was analyzed using an independent sample *t*-test with a Levene's test for equality of variances. The *p*-values were calculated using a Student's *t*-test of the replicate $2^{-\Delta CT}$ values for each miRNA in the different groups compared. A two-sided *p*-value ≤ 0.05 was considered significant.

RESULTS

Patients Characteristics

Patient and tumor characteristics are summarized in **Table 1**. We have analyzed 60 patients for whom paired tumor and SLN samples were available. A total of 20 patients had negative

TABLE 1 | Basic patient and tumor characteristics.

Variable		Total (%)	LN negative (%)	LN positive (%)
N	*N* (%)	60 (100)	20 (100)	40 (100)
Age (years)	≤50	9 (15)	2 (10)	7 (18)
	>50	51 (85)	18 (90)	33 (82)
	Mean + SD	63.6 ± 14.2	68.8 ± 11.9	60.9 ± 14.7
	Median (range)	64 (26-88)	71 (47-87)	60 (26-88)
Tumor stage	I	18 (30)	12 (60)	6 (15)
	II	40 (67)	8 (40)	32 (80)
	III	2 (3)	0 (0)	2 (5)
Tumor status	T1	32 (53)	12 (60)	20 (50)
	T2	27 (45)	8 (40)	19 (48)
	T3	1 (2)	0 (0)	1 (2)
Tumor grade	1	3 (5)	1 (5)	2 (5)
	2	35 (58)	14 (70)	21 (52)
	3	22 (37)	5 (25)	17 (43)
Node status	N0	20 (33)	20 (100)	0 (0)
	N+	40 (67)	0 (0)	40 (100)
SLN OSNA diagnosis	Negative	20 (33)	20 (100)	0 (0)
	Micrometastasis	20 (33)	0 (0)	20 (50)
	Macrometastasis	20 (33)	0 (0)	20 (50)
aLN status*	Negative	15 (75)	0 (0)	15 (75)
	Positive	5 (25)	0 (0)	5 (25)
CK19 status**	Low	12 (60)	0 (0)	12 (60)
	High	8 (40)	0 (0)	8 (40)
Tumor type	Unifocal	36 (60)	13 (65)	23 (58)
	Multifocal	20 (33)	7 (35)	13 (32)
	Multicentric	4 (7)	0 (0)	4 (10)
ER status	Negative	6 (10)	2 (10)	4 (10)
	Positive	54 (90)	18 (90)	36 (90)
PR status	Negative	18 (30)	7 (35)	11 (28)
	Positive	42 (70)	13 (65)	29 (72)
Receptor status	HR-positive	46 (77)	16 (80)	30 (75)
	HER-2 positive	9 (15)	3 (15)	6 (15)
	TN	5 (8)	1 (5)	4 (10)
Ki67 status	>20%	14 (23)	4 (20)	10 (25)
	≤20%	46 (77)	16 (80)	30 (75)
Lymphovascular invasion	Negative	48 (80)	18 (90)	30 (75)
	Positive	12 (20)	2 (10)	10 (25)
Menopausal status	Premenopausal	10 (17)	2 (10)	8 (20)
	Postmenopausal	50 (83)	18 (90)	32 (80)
Breast affected	Left	41 (68)	12 (60)	29 (73)
	Right	19 (32)	8 (40)	11 (27)
Breast surgery	Mastectomy	19 (32)	5 (25)	14 (35)
	Lumpectomy	41 (68)	15 (75)	26 (65)

aLN, axillary lymph node; ER, estrogen receptor; LN, lymph node; PR, progesterone receptor; TN, triple negative.
**CK19 mRNA copy number >15,000 in the subgroup of patients with macrometastatic SLNs.*
***Presence of additional aLN in the subgroup of patients with macrometastases.*

SLNs (33%) and 40 patients had positive SLNs (67%). Of those patients with positive SLNs, 20 (33%) SLNs were diagnosed as macrometastasis and 20 (33%) SLNs as micrometastasis. All patients diagnosed with macrometastasis received ALND of whom 5 (25%) had further aLN involvement. Our study included 46 patients with HR-positive tumors (77%), 9 patients with HER-2 positive carcinomas (15%), and 5 patients with TN (8%) tumors.

RNA-Sequencing

Prior to sequencing the samples were subjected to a QC. Three samples (one tumor and two SLNs) were excluded from further analyses because they did not pass all of the QC metrics, including the average read quality, the average base quality and the read length distribution with a *Q*-score > 30 (99.9% correct) (Cock et al., 2010). The remaining 59 tumors and 58 SLNs were successfully sequenced resulting in a total of 117 paired samples. All samples were sequenced in nine runs with a minimum and maximum read number of 0.33 and 34.5 million reads number, respectively. This resulted in a median 2.5 million read number per sample (**Supplementary Table 1**). Following sequence trimming, all reads containing identical insert sequences were collapsed into a single read, which were passed into the analysis pipeline. On average we obtained 0.92 million reads for each sample resulting in an average genome mapping rate of 26.9%. The remaining unmapped reads were usually from degraded RNAs that could not be uniquely mapped. After mapping and counting to relevant entries in mirbase_20 database, the number of known miRNAs was calculated using TPM to measure expression.

Differential Expression

To identify differentially expressed miRNAs between paired samples. First, we performed a data reduction analysis to compare the miRNA expression profile in paired tumor and SLN samples. Our results show that the two types of samples separated into two different groups, suggesting that the miRNA expression profile between paired tumor and SLNs samples from a same patient is different. Other prognostic factor such as tumor stage or tumor grade were in part responsible for grouping the tumor samples (**Supplementary Figure 1**). We observed significant differences in the expression of 9 miRNAs ($q < 0.05$) in tumors samples compared to SLNs. Six miRNAs were up-regulated (miR-182, miR-1291, miR-3651, miR-6240, miR-7641-1, and miR-6516) and three miRNAs were down-regulated (miR-3653, miR-3535, and miR-3607) (**Figure 1A**). We further validated the expression for all nine miRNAs using specific qPCR assays.

Next, we analyzed tumor and SLN samples independently in relation to the metastatic SLN status. We first compared negative ($n = 20$) vs. positive ($n = 40$) samples (**Figure 1B** and **Supplementary Table 2**). Our results show a 13.9-fold up-regulation of miR-7641-2 ($q < 0.001$) and a fourfold down-regulation of miR-1291 ($q < 0.001$) in positive tumor samples. A non-significant 1.9-fold down-regulation was also observed for miR-6240 ($q = 0.076$). In contrast, in positive SLNs samples only a 2.8-fold down-regulation of miR-1291 remained significant ($q = 0.048$). Next, we determined whether the differential expression of miRNAs observed in positive samples compared to negative samples was retained when samples were sub-classified as either macro- or micrometastasis. We observed a similar deregulation pattern in tumor samples. Macrometastatic tumors showed an 18.4-fold up-regulation ($q < 0.001$) of miR-7641-2, a 3.2-fold down-regulation of miR-1291 ($q = 0.014$) and a 2.3-fold down-regulation of miR-6240 ($q = 0.038$). In contrast, no differential expression was observed for SLNs diagnosed as macrometastasis (**Figure 1C** and **Supplementary Table 2**). Similar results were obtained when we compared micrometastatic and negative tumors or SLNs (**Figure 1D** and **Supplementary Table 2**). Interestingly, no significant differences were observed when comparing macrometastatic to micrometastatic tumor or SLNs (**Supplementary Table 2**), suggesting that the volume of the metastatic lesion do not translate in differences in the expression profile of miRNAs.

It has been reported that patients with $T_{1-2}N_{0-1}$ invasive non-metastatic breast cancer treated with breast conserving surgery and randomized to undergo ALND after SLN dissection vs. SLN dissection alone, showed no significant differences in local recurrence or regional recurrence (Giuliano et al., 2010; Jagsi et al., 2014). Therefore, we investigated the expression of miRNAs in tumors ($n = 19$) or SLNs ($n = 20$) from patients diagnosed as macrometastasis and with additional positive aLNs ($n = 5$). Our results show that in SLNs samples, a 10.6-fold up-regulation of miR-1291 ($q < 0.001$) was associated with additional aLN metastases (**Figure 1E**). In tumor samples, we found a differential expression of miR-182 ($p = 0.013$) and miR-7641-2 ($p = 0.029$) but none were significant after FDR correction ($q > 0.05$) (**Figure 1E**). We have further analyzed this subset of samples in relation to the TTL since it has been reported that a TTL cut-off value >15,000 copies of CK19 is associated with additional aLN metastasis (Peg et al., 2013). Our results show a negative correlation of miR-3535 and miR-3653 in tumors and a positive association of miR-1291in SLNs in relation to the TTL. However, none of these miRNAs passed the FDR correction ($q > 0.05$) (data not shown).

Our tumor samples included 46 HR-positive (77%), 9 HER2-positive (15%), and 5 TN (8%) tumors (**Table 1**). No differential expression of miRNAs was found between molecular subtypes in tumor samples ($n = 59$). However, in SLN samples, miR-1291 showed a 10-fold up-regulation in HER2-positive samples ($q = 0.0046$) compared to HR-positive SLNs and a 3.5-fold up-regulation of miR-3607 in HER2-positive SLNs compared to TN SLNs ($q = 0.033$) (**Table 2**). Next, we investigated the expression profile of miRNAs according to the molecular subtype and the SLN metastatic status. In TN tumor samples with positive SLNs, an 18.1 down-regulation of miR-3535 ($q = 0.048$) was observed compared with HR-positive tumors. A similar non-significant 20.2-fold down-regulation was observed when comparing TN with HER-2 positive tumors with positive LNs ($q = 0.06$). No differences were observed in SLN samples (data not shown).

Gene Target Prediction and Enrichment Analyses

To investigate target genes of miR-1291 we used the multiMiR Bioconductor's package to retrieve validated targets from

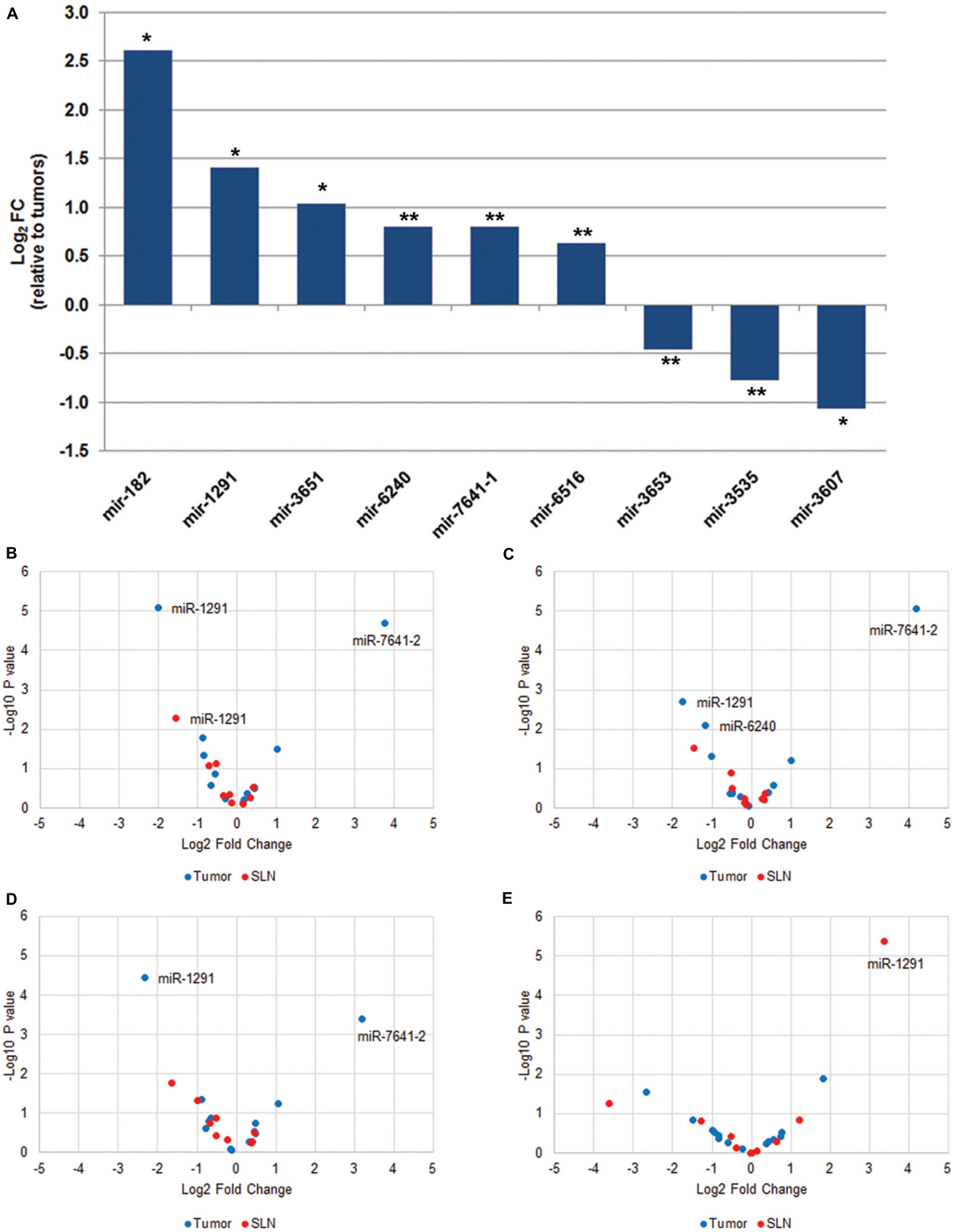

FIGURE 1 | MiRNA expression profile in paired tumor and SLN samples from breast cancer patients. **(A)** Total number of miRNAs with significant differential expression ($q < 0.05$) in tumor samples compared to paired SLN samples. Data is expressed as log2 fold change relative to tumor samples. *$q < 0.001$, **$q < 0.05$. The volcano plots show differentially expressed miRNAs in tumor (blue) and SLN (red) samples according to the locoregional metastatic status: positive vs. negative **(B)**, macrometastasis vs. negative **(C)**, micrometastasis vs. negative **(D)**, and subgroup of patients with macrometastasis that were treated with ALND and grouped according to the aLN status (positive vs. negative) **(E)**. The data show the relationship between the *p*-values (*y*-axis) and the fold change (*x*-axis) between the experimental groups. Only miRNAs with $q < 0.05$ are shown in the plots.

TABLE 2 | Differential expression of miRNAs in tumor samples (n = 59) and SLN (n = 58) according to the breast cancer molecular subtypes.

Sample	Comparison	miRNA	log2 FC	p-Value	q-Value		Fold regulation
Tumor	HR+ vs. HER2+	mir-21	−1.28	0.012	0.175	2.4	Down-regulation
	HR+ vs. TNBC	mir-3651	1.75	0.053	0.596	3.4	Up-regulation
	HER2+ vs. TNBC	mir-3651	2.13	0.032	0.352	4.4	Up-regulation
		mir-21	1.88	0.050	0.352	3.7	Up-regulation
SLN	HR+ vs. HER2+	mir-1291	3.323	0.0005	0.005	10.0	Up-regulation
		mir-3607	−0.756	0.032	0.143	1.7	Down-regulation
	HR+ vs. TNBC	mir-3607	1.03	0.051	0.319	2.1	Up-regulation
		mir-6516	1.56	0.074	0.319	3.0	Up-regulation
	HER2+ vs. TNBC	mir-3607	1.79	0.004	0.033	3.5	Up-regulation
		mir-6516	2.27	0.017	0.075	4.9	Up-regulation
		mir-3653	1.09	0.044	0.132	2.1	Up-regulation
		mir-1291	−1.98	0.073	0.164	3.9	Down-regulation

FC, fold change.

mirecords, mirtarbase, and tarbase databases and predicted targets from diana-microt, elmmo, microcosm, miranda, mirdb, pictar, pita, and targetscan databases. A total of 171 validated targets were found (**Supplementary Table 3**). To investigate the pathways and genes associated with miR-1291 we used the Reactome pathway database. The analyses show the most relevant pathways and the number of significant target genes within each category (**Figure 2A** and **Supplementary Table 4**), the interactions between these pathways (**Figure 2B**) and the targets genes associated with these pathways and how these genes and signaling pathways are linked (**Figure 2C**). Overall, our results show that the top five significant pathways regulated by miR-1291 are signaling by WNT, planar cell polarity/convergent extension (PCP/CE) pathway, β-catenin independent WNT signaling, diseases of signal transduction and signaling by receptor tyrosine kinases (RTKs). Interestingly, all target genes within these categories (**Figure 2C**) were included in the validated mirtabase and tabase databases (**Supplementary Table 3**).

Clinical Status

Our series include 60 patients with early breast cancer and we reported recurrence in 10/60 (17%) patients. Four patients had locoregional recurrences (breast) and six patients had distant metastasis (two cases in the liver and one case in the lung, pancreas, bone, or brain). At last follow-up, three patients with recurrences in their breast were reported alive without any evidence of disease, whereas 5/10 other recurrences (breast, lung, bone, and pancreas) were alive with disease. We reported seven deaths in our series, three deaths were due to complications related to the disease (two deaths were related to liver metastases and one death related to lung metastases), and four deaths non-related to disease (general deterioration related to age and/or associated comorbidity). We performed a survival analysis using the expression of miR-1291. The Kaplan–Meier and log-rank analyses showed that the expression levels of miR-1291 (high vs. low expression) were not significantly associated with DFS nor OS (**Figure 3**).

DISCUSSION

In this study we have correlated the expression of miRNAs with SLN metastatic status and the outcome of ALND. We profiled the miRNA expression using NGS and key findings were validated by qPCR. We analyzed samples from 60 patients with early breast cancer for whom paired primary tumor and SLNs were available. Overall, we found a poor correlation in the miRNA expression profile from SLNs compared to match primary tumors samples from the same patient. These differences are likely due to histological differences between the LN and the tumor. Whereas the SLN is constituted mainly by lymphoid and monocytes cells the tumor tissue is formed mostly by epithelial and mesenchymal cells. Nonetheless, we found nine differentially expressed miRNAs in tumor tissues compared to SLN ($q \leq 0.05$). Of those, only four miRNAs showed a fold de-regulation $\geq$2.0 (either up-regulation or down-regulation). The two most de-regulated miRNAs were miR-182 (6.2-fold up-regulation, $p < 0.001$) and miR-3607 (2.1-fold down-regulation, $p < 0.001$). Our data agrees with a previous report showing that miR-182 directly targets MIM (Missing in Metastasis), which suppresses metastasis by inhibiting Ras homolog family member A (RhoA) activity and stress fiber formation in breast cancer cells (Lei et al., 2014). In addition, miR-182 expression was reported to be higher in primary tumors with positive SLNs metastasis and correlated with earlier relapse-free and metastasis-free survivals (Lei et al., 2014). On the other hand, miR-3607 has been shown to be widely attenuated in prostate cancer (PCa) patients and low expression levels were significantly associated with higher PCa stage, Gleason score, serum PSA levels, tumor progression, and poor survival outcome (Saini et al., 2014).

Given that paired tumor and SLN samples in our study showed different miRNA expression profiles, we analyzed both types of samples independently in relation to the locoregional metastatic status. We show that down-regulation of miR-1291 in both tumor and SLN samples was associated with positive SLN metastases. Interestingly, we found that in the subgroup of positive patients that underwent ALND and had further aLN metastases the

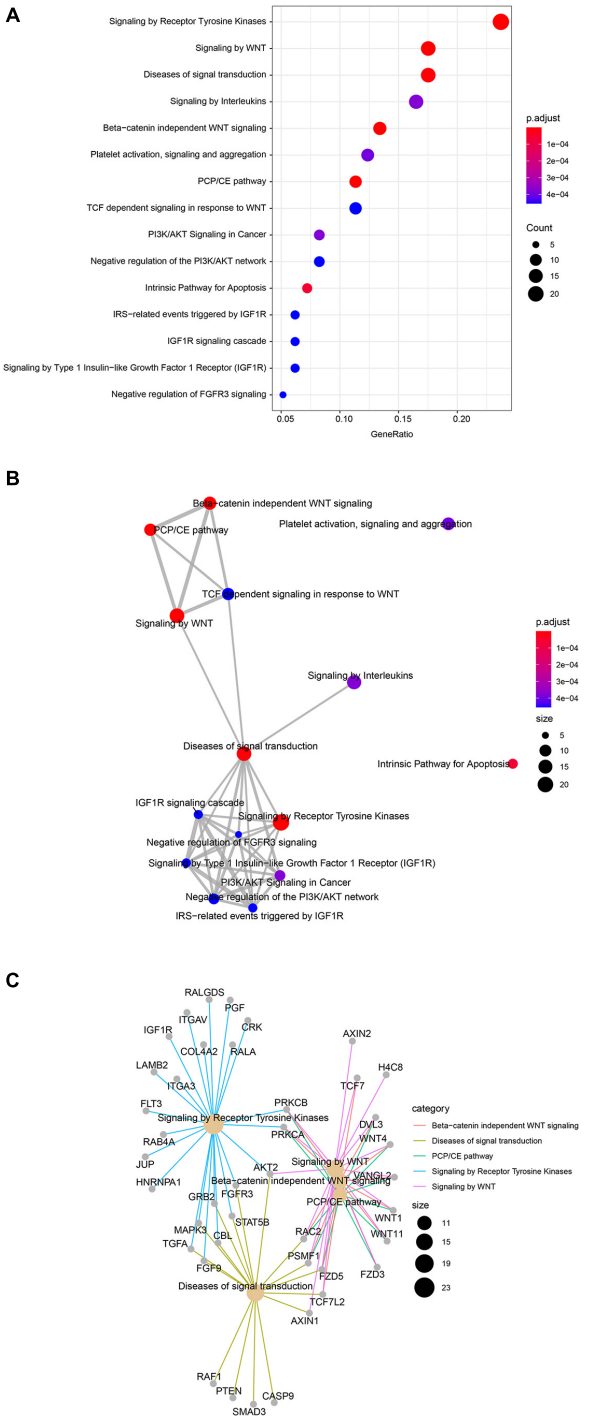

FIGURE 2 | Reactome pathway enrichment in miR-1291 targeted genes. **(A)** Dot plot graph shows the 15 most significant enriched pathways and the ratio between the genes that belongs to that category and the total number of genes in the pathway. The color of the nodes indicates the *p*-value and the size of the nodes the number of genes in the data that belong to that pathway **(B)** Enrichment map of the top 15 signaling pathways grouped by similarity. Nodes are colored by *p*-value and their size reflects the number of genes found in that term. **(C)** Network plot of the genes found in the top five enriched signaling pathways showing the linkages between genes and pathways. The size of nodes indicates the number of genes found in that pathway.

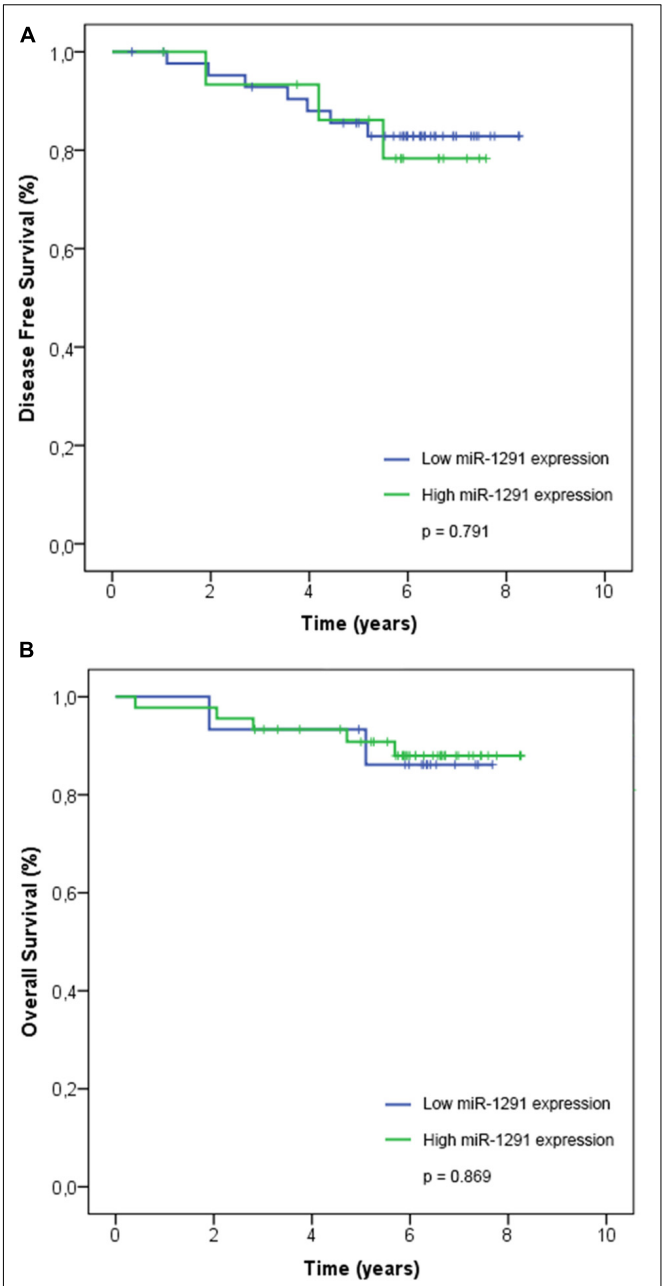

FIGURE 3 | Association of miR-1291 expression with patient outcome. The expression levels of log2-transformed miR-1291 was correlated with **(A)** disease-free survival and **(B)** overall survival. Kaplan–Meier curves and log-rank tests are shown.

expression of miR-1291 was significantly higher compare to patient with negative aLNs. However, these results must be interpreted with caution since only five patients fulfilled those characteristics in our dataset and thus, further clinical validation in a larger cohort of patients will be required to establish these observations. If confirmed, these results would suggest that the degree of down-regulation might have an effect on how miR-1291 modulates its target genes and affect various signaling

pathways. Nonetheless, our study did not correlate the miRNA expression with the transcriptional expression in the same samples. Therefore we could not determine if miR-1291 function as an oncogene as it has been described for hepatocellular carcinomas and liver cirrhosis patients (Hagag et al., 2020) or in microvesicles derived from medulloblastoma cell lines with stem-like properties (Kaid et al., 2020). Nevertheless, our data suggest the contrary and agrees with previous studies showing that the expression of miR-1291 is significantly down-regulated in various human cancers, including pancreatic cancer (Bi et al., 2014; Tu et al., 2016) esophageal squamous cell carcinoma (ESCC) (Luo et al., 2015), PCa (Cai et al., 2019) and renal cell carcinoma (RCC) (Yamasaki et al., 2013). In these studies, loss of tumor-suppressive miR-1291 have been reported to enhanced cell proliferation, migration, and invasion through up-regulation of various miR-1291 targets, whereas restoration of miR-1291 expression in cancer cell lines and animal models reverted those effects. For example, miR-1291 restoration repressed tumorigenesis in prostate and pancreatic xenograft tumor models via inhibition of Mediator of RNA polymerase II transcription subunit 1 (MED1) (Cai et al., 2019), N-methylnicotinamide (NMN) (Bi et al., 2014). In addition, miR-1291 reduced the protein levels of target genes including ATP Binding Cassette Subfamily C Member 11 (ABCC1), Forkhead box protein A2 (FOXA2), Anterior Gradient 2 (AGR2), methyl CpG binding protein 2 (MeCP2) and carnitine palmitoyltransferase 1C (CPT1C) resulting in the suppression of growth and tumorigenesis of human breast and pancreatic cell lines (Bi et al., 2014; Li et al., 2015; Tu et al., 2016; Chen et al., 2020). Similar effects have been reported in RCC through targeting SLC2A1/GLUT1 (Yamasaki et al., 2013) which it has also been reported to be overexpressed in human breast carcinomas (da Cunha et al., 2013), in esophageal carcinomas by inhibiting mucin 1 (MUC1) (Luo et al., 2015), whereas miR-1291 acts upstream of the Rho GTPase-activating protein 29 (ArhGAP29) to negatively regulate the RhoA/ROCK1 epithelial mesenchymal transition (EMT) pathway, ultimately leading to endometrial fibrosis (Xu et al., 2017). Interestingly, our target prediction analysis identified the reported genes as miR-1291 targets, whereas our enrichment analysis pointed out various signaling pathways and several miR-1291 target genes that are commonly overexpressed in breast cancer, suggesting that in fact miR-1291 acts as a tumor suppressor gene (TSG). For instance, our data shows that miR-1291 targets genes that are involved in both canonical and non-canonical WNT signaling and regulate many aspects of cell polarity, morphogenesis, and development (Saito-Diaz et al., 2013). In some contexts, both the canonical and non-canonical WNT signaling contribute to tumor formation by promoting cell migration, invasiveness, and metastasis (van Amerongen, 2012; VanderVorst et al., 2019). One of the non-canonical WNT pathways we identified in relation to miR-1291 is the PCP pathway, which contains core Wnt/PCP components that are overexpressed in a variety of solid tumors and have been directly implicated in promoting tumor cell migration and metastasis (VanderVorst et al., 2018). Our data shows that two of the Wnt/PCP genes (WNT11 and VANGL2) are targets of miR-1291 and both have been reported to promote cell motility and metastasis in breast cancer (Luga et al., 2012). The Frizzled-5 (FDZ5) and FDZ3 are another non-canonical WNT members targeted by miR-1291 that bind to the FZD receptor and leads to activation of small Rho GTPases (RAC2) and JNK, which regulate the cytoskeleton and coordinate cell migration and polarity (Schlessinger et al., 2009). Overexpression of FDZ5 and FDZ3 has been associated with cancer aggressiveness in human breast carcinomas (da Cunha et al., 2013; Kazanietz and Caloca, 2017). Furthermore, we showed that miR-1291 is associated with negative regulation of various RTKs pathways, such as the PI3K/AKT, IGFR1, and FGFR3 signaling pathways that have a pivotal role in the regulation of cancer proliferation, angiogenesis, and metastasis in breast cancer (Butti et al., 2018). We have identified several well-known oncogenes that participate in these pathways, including MAPK3, RAF1, AKT2, or TGFA as validated targets of miR-1291. Collectively, our data supports the role of miR-1291 as a TSG in the onset of locoregional metastasis in early breast cancer patients. Further research is warranted to investigate the expression of miR-1291 as a potential surrogate marker of SLN involvement in early breast cancer patients.

Our results complement other studies that have compared the miRNA expression profile between primary tumor and LNs from the same patients (Smeets et al., 2011; Gravgaard et al., 2012; Cascione et al., 2013). Gravgaard et al. (2012) found a differential expression of miRNA-200 family and miRNA-9 in LN associated with the metastatic process. Cascione et al. (2013) reported the analysis of 173 FFPE tumors and 53 matched LN in TNBC. They found two miRNA signatures (miR-16, 155, 125b, 374a and miR-16, 125b, 374a, 374b, 421, 655, 497) that were independent predictors for OS and distant-DFS, respectively. In another study, Smeets et al. (2011) identified eight miRNAs (miR-195, miR-191, miR-132, miR-203, miR-431, miR-16, miR-30c, miR-30a) with measurable differences in gene and miRNA expression between N0 and N+ patients. However, the deregulated miRNAs described here is different from previously published (Smeets et al., 2011; Gravgaard et al., 2012; Cascione et al., 2013). Various reasons might explain this discrepancy. First, the different molecular breast cancer subtypes investigated in each study. Our cohort of patients comprised mostly luminal tumors, whereas others analyzed only TNs tumors (Cascione et al., 2013) or a mix of molecular subtypes (Smeets et al., 2011; Gravgaard et al., 2012). Second, the type of starting biological material used in previous studies, which was either archive FFPE specimens (Gravgaard et al., 2012; Cascione et al., 2013) or frozen sections (Smeets et al., 2011). In contrast, we used the entire fresh SLN homogenate employed in the OSNA assay (Cserni, 2012). The OSNA assay was developed as an alternative to intra-operative microscopy-based pathological assessment of the SLN because such methods (frozen sections, touch imprints scrapes, or a combination of these) have a limited ability to accurately measure total metastatic volume in a LN (Cserni, 2012). Thus, our study was not limit by sampling bias that might have occurred from using different parts of the LN. Moreover, the OSNA technique allows for the immediate sample processing required to preserve RNA for research purposes. Therefore, we did not encounter the limitations of using RNA extracted from archival FFPE tissues, which has often suffered chemical modification,

cross-linking, and degradation over time as a result of the fixation and archiving methods (Esteve-Codina et al., 2017). Finally, a recent study has shown that differences on the protocol used for the preparation of the NGS libraries results in over- or underrepresentation of miRNAs in the sequencing library (Coenen-Stass et al., 2018). The use of different NGS library preparation kits and protocols could result on an inefficient size-selection or in an enrichment of other RNA species therefore decreasing the number of usable miRNAs reads. That could be our case, given that the library kit used to generate our RNA-seq data (NEBNext) has been recently reported to produce the least mirRNA with the greatest coefficient of variation (Coenen-Stass et al., 2018). In fact, a preliminary analysis of our RNA-seq data in tumor samples shows an enrichment in small nucleolar RNAs (snoRNAs) (data not published). Interestingly, several of the miRNAs described in this study appear to be derived from snoRNAs small ncRNAs (Thorenoor and Slaby, 2015). Other studies have reported human snoRNAs fragments that are annotated as miRNAs, including miR-1291 (SNORA2C), miR-1259 (SNORD12B), miR-1248 (SNORA81), miR-3535, and miR3653 (SNORD125), all of which are reported in our study. Similar reports have described other human snoRNAs showed miRNA-like processing signatures that can inhibit the expression of targets genes (Ender et al., 2008; Brameier et al., 2011; Makarova et al., 2013; Thorenoor and Slaby, 2015). Further studies are warranted to understand the role of snoRNAs in the mechanisms of invasion in breast cancer.

CONCLUSION

To summarize, in this study we have analyzed the miRNA expression profile in paired tumor and SLN from patients with early breast cancer. We found that miR-1291 is differentially expressed in breast cancer patients with positive SLN and further involvement of aLN metastases. Our prediction and enrichment analyses provides a comprehensive understanding of the involvement of miR-1291 in promoting metastasis. Altogether, our data supports the role of miR-1291 as a TSG. Given the increasing importance of miRNAs in regulating cellular processes and the clinical success of the OSNA assay in detecting LN metastases, further research is warranted to investigate the direct implications of our results (Condrat et al., 2020) in treatment decision-making and (Janssen et al., 2010) in developing more effective blood-based clinical tests.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. The datasets can be found in the Sequence Research Archive under ID PRJNA663033 and are available for download here http://www.ncbi.nlm.nih.gov/bioproject/663033.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Comité Ètic d'Investigació Clínica (CEIC) Hospital de la Santa Creu i Sant Pau Sant Quintí 77-79, 08041 Barcelona, Catalunya, Spain. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DE, LL-V, and AB were involved in the conceptualization of the project. DE, LL-V, OB, JM, JP, AM, CA, TR, and EL were involved in resources, investigation, and methodology. DE, LL-V, EL, and AB were involved in analysis and interpretation of data. All authors have contributed to the writing of the manuscript and have critically reviewed it.

FUNDING

Instituto de Salud Carlos III and European Regional Development Fund (PI13/00110 to DE); Centro de Investigación Biomédica en Red Cáncer (CB 16/12/00471 to AB); and Red Temática de Investigación Cooperativa de Cáncer (RTICC; RD12/0036/0076 to all authors).

ACKNOWLEDGMENTS

We thank the patients and their caregivers for their participation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.562114/full#supplementary-material

REFERENCES

Andersen, C. L., Jensen, J. L., and Orntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 64, 5245–5250. doi: 10.1158/0008-5472.can-04-0496

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bi, H. C., Pan, Y. Z., Qiu, J. X., Krausz, K. W., Li, F., Johnson, C. H., et al. (2014). N-methylnicotinamide and nicotinamide N-methyltransferase are associated with microRNA-1291-altered pancreatic carcinoma cell metabolome and suppressed tumorigenesis. *Carcinogenesis* 35, 2264–2272. doi: 10.1093/carcin/bgu174

Blenkiron, C., Goldstein, L. D., Thorne, N. P., Spiteri, I., Chin, S. F., Dunning, M. J., et al. (2007). MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol.* 8:R214.

Brameier, M., Herwig, A., Reinhardt, R., Walter, L., and Gruber, J. (2011). Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. *Nucleic Acids Res.* 39, 675–686. doi: 10.1093/nar/gkq776

Butti, R., Das, S., Gunasekaran, V. P., Yadav, A. S., Kumar, D., and Kundu, G. C. (2018). Receptor tyrosine kinases (RTKs) in breast cancer: signaling, therapeutic implications and challenges. *Mol. Cancer* 17:34.

Cai, Q., Zhao, A., Ren, L., Chen, J., Liao, K., Wang, Z., et al. (2019). MicroRNA-1291 mediates cell proliferation and tumorigenesis by downregulating MED1 in prostate cancer. *Oncol. Lett.* 17, 3253–3260.

Cascione, L., Gasparini, P., Lovat, F., Carasi, S., Pulvirenti, A., Ferro, A., et al. (2013). Integrated MicroRNA and mRNA signatures associated with survival in triple negative breast cancer. *PLoS One* 8:e55910. doi: 10.1371/journal.pone.0055910

Chen, Y., Zhou, Y., Han, F., Zhao, Y., Tu, M., Wang, Y., et al. (2020). A novel miR-1291-ERRα-CPT1C axis modulates tumor cell proliferation, metabolism and tumorigenesis. *Theranostics* 10, 7193–7210. doi: 10.7150/thno.44877

Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., and Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* 38, 1767–1771. doi: 10.1093/nar/gkp1137

Coenen-Stass, A. M. L., Magen, I., Brooks, T., Ben-Dov, I. Z., Greensmith, L., Hornstein, E., et al. (2018). Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA Biol.* 15, 1133–1145.

Condrat, C. E., Thompson, D. C., Barbu, M. G., Bugnar, O. L., Boboc, A., Cretoiu, D., et al. (2020). miRNAs as biomarkers in disease: latest findings regarding their role in diagnosis and prognosis. *Cells* 9:276. doi: 10.3390/cells9020276

Cserni, G. (2012). Intraoperative analysis of sentinel lymph nodes in breast cancer by one-step nucleic acid amplification. *J. Clin. Pathol.* 65, 193–199. doi: 10.1136/jclinpath-2011-200301

da Cunha, J. P., Galante, P. A., de Souza, J. E., Pieprzyk, M., Carraro, D. M., Old, L. J., et al. (2013). The human cell surfaceome of breast tumors. *Biomed Res. Int.* 2013:976816.

Dai, R., and Ahmed, S. A. (2011). MicroRNA, a new paradigm for understanding immunoregulation, inflammation, and autoimmune diseases. *Transl. Res.* 157, 163–179. doi: 10.1016/j.trsl.2011.01.007

Elston, C. W., and Ellis, I. O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 19, 403–410. doi: 10.1111/j.1365-2559.1991.tb00229.x

Ender, C., Krek, A., Friedlander, M. R., Beitzinger, M., Weinmann, L., Chen, W., et al. (2008). A human snoRNA with microRNA-like functions. *Mol. Cell* 32, 519–528. doi: 10.1016/j.molcel.2008.10.017

Esteve-Codina, A., Arpi, O., Martinez-Garcia, M., Pineda, E., Mallo, M., Gut, M., et al. (2017). A comparison of RNA-seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples. *PLoS One* 12:e0170632. doi: 10.1371/journal.pone.0170632

Giuliano, A. E., McCall, L., Beitsch, P., Whitworth, P. W., Blumencranz, P., Leitch, A. M., et al. (2010). Locoregional recurrence after sentinel lymph node dissection with or without axillary dissection in patients with sentinel lymph node metastases: the American College of Surgeons Oncology Group Z0011 randomized trial. *Ann. Surg.* 252, 426–432; discussion 432–433.

Gravgaard, K. H., Lyng, M. B., Laenkholm, A. V., Sokilde, R., Nielsen, B. S., Litman, T., et al. (2012). The miRNA-200 family and miRNA-9 exhibit differential expression in primary versus corresponding metastatic tissue in breast cancer. *Breast Cancer Res. Treat.* 134, 207–217. doi: 10.1007/s10549-012-1969-9

Hagag, N. A., Ali, Y. B. M., Elsharawy, A. A., and Talaat, R. M. (2020). Clinical impact of circulated miR-1291 in plasma of patients with liver cirrhosis (LC) and hepatocellular carcinoma (HCC): implication on glypican-3 expression. *J. Gastrointest. Cancer.* 51, 234–241. doi: 10.1007/s12029-019-00234-9

Jagsi, R., Chadha, M., Moni, J., Ballman, K., Laurie, F., Buchholz, T. A., et al. (2014). Radiation field design in the ACOSOG Z0011 (Alliance) trial. *J. Clin. Oncol.* 32, 3600–3606. doi: 10.1200/jco.2014.56.5838

Janssen, E. A., Slewa, A., Gudlaugsson, E., Jonsdottir, K., Skaland, I., Soiland, H., et al. (2010). Biologic profiling of lymph node negative breast cancers by means of microRNA expression. *Mod. Pathol.* 23, 1567–1576. doi: 10.1038/modpathol.2010.177

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al. (2020). The reactome pathway knowledgebase. *Nucleic Acids Res.* 48, D498–D503.

Kaid, C., Assoni, A., Marçola, M., Semedo-Kuriki, P., Bortolin, R. H., Carvalho, V. M., et al. (2020). Proteome and miRNome profiling of microvesicles derived from medulloblastoma cell lines with stem-like properties reveals biomarkers of poor prognosis. *Brain Res.* 1730:146646. doi: 10.1016/j.brainres.2020.146646

Kazanietz, M. G., and Caloca, M. J. (2017). The Rac GTPase in cancer: from old concepts to new paradigms. *Cancer Res.* 77, 5445–5451. doi: 10.1158/0008-5472.can-17-1456

Lei, R., Tang, J., Zhuang, X., Deng, R., Li, G., Yu, J., et al. (2014). Suppression of MIM by microRNA-182 activates RhoA and promotes breast cancer metastasis. *Oncogene* 33, 1287–1296. doi: 10.1038/onc.2013.65

Li, M. M., Addepalli, B., Tu, M. J., Chen, Q. X., Wang, W. P., Limbach, P. A., et al. (2015). Chimeric microRNA-1291 biosynthesized efficiently in *Escherichia coli* is effective to reduce target gene expression in human carcinoma cells and improve chemosensitivity. *Dis. Markers* 43, 1129–1136. doi: 10.1124/dmd.115.064493

Luga, V., Zhang, L., Viloria-Petit, A. M., Ogunjimi, A. A., Inanlou, M. R., Chiu, E., et al. (2012). Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration. *Cell* 151, 1542–1556. doi: 10.1016/j.cell.2012.11.024

Luo, H., Guo, W., Wang, F., You, Y., Wang, J., Chen, X., et al. (2015). miR-1291 targets mucin 1 inhibiting cell proliferation and invasion to promote cell apoptosis in esophageal squamous cell carcinoma. *Oncol. Rep.* 34, 2665–2673. doi: 10.3892/or.2015.4206

Makarova, J. A., Ivanova, S. M., Tonevitsky, A. G., and Grigoriev, A. I. (2013). New functions of small nucleolar RNAs. *Biochem. Biokhim.* 78, 638–650. doi: 10.1134/s0006297913060096

Nana-Sinkam, S. P., and Croce, C. M. (2011). MicroRNAs as therapeutic targets in cancer. *Transl. Res.* 157, 216–225.

Nicoloso, M. S., Spizzo, R., Shimizu, M., Rossi, S., and Calin, G. A. (2009). MicroRNAs–the micro steering wheel of tumour metastases. *Nat. Rev. Cancer* 9, 293–302. doi: 10.1038/nrc2619

Peg, V., Espinosa-Bravo, M., Vieites, B., Vilardell, F., Antunez, J. R., de Salas, M. S., et al. (2013). Intraoperative molecular analysis of total tumor load in sentinel lymph node: a new predictor of axillary status in early breast cancer patients. *Breast Cancer Res. Treat.* 139, 87–93. doi: 10.1007/s10549-013-2524-z

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna: The R Foundation for Statistical Computing.

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

Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25.

Ru, Y., Kechris, K. J., Tabakoff, B., Hoffman, P., Radcliffe, R. A., Bowler, R., et al. (2014). The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. *Nucleic Acids Res.* 42:e133. doi: 10.1093/nar/gku631

Saini, S., Majid, S., Shahryari, V., Tabatabai, Z. L., Arora, S., Yamamura, S., et al. (2014). Regulation of SRC kinases by microRNA-3607 located in a frequently deleted locus in prostate cancer. *Mol. Cancer Ther.* 13, 1952–1963. doi: 10.1158/1535-7163.mct-14-0017

Saito-Diaz, K., Chen, T. W., Wang, X., Thorne, C. A., Wallace, H. A., Page-McCaw, A., et al. (2013). The way Wnt works: components and mechanism. *Growth Factors* 31, 1–31. doi: 10.3109/08977194.2012.752737

Schlessinger, K., Hall, A., and Tolwinski, N. (2009). Wnt signaling pathways meet Rho GTPases. *Genes Dev.* 23, 265–277. doi: 10.1101/gad.1760809

Shek, L. L., and Godolphin, W. (1988). Model for breast cancer survival: relative prognostic roles of axillary nodal status, TNM stage, estrogen receptor concentration, and tumor necrosis. *Cancer Res.* 48, 5565–5569.

Smeets, A., Daemen, A., Vanden Bempt, I., Gevaert, O., Claes, B., Wildiers, H., et al. (2011). Prediction of lymph node involvement in breast cancer from primary tumor tissue using gene expression profiling and miRNAs. *Breast Cancer Res. Treat.* 129, 767–776. doi: 10.1007/s10549-010-1265-5

Thorenoor, N., and Slaby, O. (2015). Small nucleolar RNAs functioning and potential roles in cancer. *Tumour Biol.* 36, 41–53. doi: 10.1007/s13277-014-2818-8

Tsujimoto, M., Nakabayashi, K., Yoshidome, K., Kaneko, T., Iwase, T., Akiyama, F., et al. (2007). One-step nucleic acid amplification for intraoperative detection of lymph node metastasis in breast cancer patients. *Clin. Cancer Res.* 13, 4807–4816.

Tu, M. J., Pan, Y. Z., Qiu, J. X., Kim, E. J., and Yu, A. M. (2016). MicroRNA-1291 targets the FOXA2-AGR2 pathway to suppress pancreatic cancer cell

proliferation and tumorigenesis. *Oncotarget* 7, 45547–45561. doi: 10.18632/oncotarget.9999

Valastyan, S. (2012). Roles of microRNAs and other non-coding RNAs in breast cancer metastasis. *J. Mammary Gland Biol. Neoplasia* 17, 23–32. doi: 10.1007/s10911-012-9241-9

van Amerongen, R. (2012). Alternative Wnt pathways and receptors. *Cold Spring Harb. Perspect. Biol.* 4:a007914. doi: 10.1101/cshperspect.a007914

VanderVorst, K., Dreyer, C. A., Konopelski, S. E., Lee, H., Ho, H. H., and Carraway, K. L. III (2019). Wnt/PCP signaling contribution to carcinoma collective cell migration and metastasis. *Cancer Res.* 79, 1719–1729. doi: 10.1158/0008-5472.can-18-2757

VanderVorst, K., Hatakeyama, J., Berg, A., Lee, H., and Carraway, K. L. III (2018). Cellular and molecular mechanisms underlying planar cell polarity pathway contributions to cancer malignancy. *Semin. Cell Dev. Biol.* 81, 78–87. doi: 10.1016/j.semcdb.2017.09.026

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3:RESEARCH0034.

Webber, C., Gospodarowicz, M., Sobin, L. H., Wittekind, C., Greene, F. L., Mason, M. D., et al. (2014). Improving the TNM classification: findings from a 10-year continuous literature review. *Int. J. Cancer* 135, 371–378. doi: 10.1002/ijc.28683

Xu, Q., Duan, H., Gan, L., Liu, X., Chen, F., Shen, X., et al. (2017). MicroRNA-1291 promotes endometrial fibrosis by regulating the ArhGAP29-RhoA/ROCK1 signaling pathway in a murine model. *Mol. Med. Rep.* 16, 4501–4510. doi: 10.3892/mmr.2017.7210

Yamasaki, T., Seki, N., Yoshino, H., Itesako, T., Yamada, Y., Tatarano, S., et al. (2013). Tumor-suppressive microRNA-1291 directly regulates glucose transporter 1 in renal cell carcinoma. *Cancer Sci.* 104, 1411–1419.

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 December 2020
doi: 10.3389/fgene.2020.597888

Identifying Biomarkers to Predict the Progression and Prognosis of Breast Cancer by Weighted Gene Co-expression Network Analysis

Gengsheng Shi[1†], Zhenru Shen[2†], Yi Liu[1] and Wenqin Yin[1†]*

[1] Department of Clinical and Public Health, School of Health and Rehabilitation, Jiangsu College of Nursing, Jiangsu, China,
[2] Department of Cardiothoracic Surgery, The Second People's Hospital of Huai'an, Huai'an, China

Breast cancer (BC) is the leading cause of cancer death among women worldwide. The molecular mechanisms of its pathogenesis are still to be investigated. In our study, differentially expressed genes (DEGs) were screened between BC and normal tissues. Based on the DEGs, a weighted gene co-expression network analysis (WGCNA) was performed in 683 BC samples, and eight co-expressed gene modules were identified. In addition, by relating the eight co-expressed modules to clinical information, we found the blue module and pathological stage had a significant correlation (r = 0.24, p = 1e–10). Validated by multiple independent datasets, using one-way ANOVA, survival analysis and expression level revalidation, we finally screened 12 hub genes that can predict BC progression and prognosis. Functional annotation analysis indicated that the hub genes were enriched in cell division and cell cycle regulation. Importantly, higher expression of the 12 hub genes indicated poor overall survival, recurrence-free survival, and disease-free survival in BC patients. In addition, the expression of the 12 hub genes showed a significantly positive correlation with the expression of cell proliferation marker Ki-67 in BC. In summary, our study has identified 12 hub genes associated with the progression and prognosis of BC; these hub genes might lead to poor outcomes by regulating the cell division and cell cycle. These hub genes may serve as a biomarker and help to distinguish different pathological stages for BC patients.

Keywords: breast cancer, WGCNA, progression, cell cycle, prognosis

OPEN ACCESS

Edited by:
Dongqing Wei,
Shanghai Jiao Tong University, China

Reviewed by:
Liang Chen,
Zhongnan Hospital of Wuhan University, China
Jianxiang Shi,
Zhengzhou University, China

****Correspondence:***
Wenqin Yin
Yin.wenqin@jscn.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *22 August 2020*
Accepted: *23 November 2020*
Published: *17 December 2020*

Citation:
Shi G, Shen Z, Liu Y and Yin W (2020) Identifying Biomarkers to Predict the Progression and Prognosis of Breast Cancer by Weighted Gene Co-expression Network Analysis. Front. Genet. 11:597888. doi: 10.3389/fgene.2020.597888

INTRODUCTION

In 2018, there are approximately 2.1 million new cases of breast cancer (BC) and 630,000 deaths worldwide (Bray et al., 2018). Although adjuvant therapies have reduced BC-related mortality, up to 25% of patients will develop tumor relapse (Early Breast Cancer Trialists' Collaborative Group et al., 2012; Howlader et al., 2014). The mortality of BC is largely due to recurrent tumors (Berry et al., 2005). BC patients with higher clinical stage are more likely to recurrence and have worse prognosis (Garcia-Murillas et al., 2015). Genetic mutations have a key role in the progression of BC. About 20% of triple-negative BC patients have BRCA mutation, while BRCA mutations are rarely found in the healthy population (Trainer et al., 2011). Over 30% of BC patients have overexpressed HER2. Ki-67 was reported to be associated with disease-free survival of BC (Kontzoglou et al., 2013). BRAF mutations were present in over 3% of metastatic BC patients (Cejalvo et al., 2016). Although there have been great advances in the treatment of BC, the ability to treat advanced BC is still

limited due to the lack of precise molecular targets (Meng et al., 2016). Therefore, more novel candidate genes are needed to improve early diagnosis and treatment decisions.

Many studies have suggested that genes were involved in tumor progression and prognosis (van Kessel et al., 2018; McFaline-Figueroa et al., 2019). Gene expression profiles such as microarray and RNA-sequencing are common ways to determine biomarkers related to progression of various cancers (Dahinden et al., 2010; Gerlinger et al., 2014). However, most published studies have focused on the screening of differentially expressed genes (DEGs), ignoring the high connection between genes, although genes with similar expression patterns may be functionally related (Tavazoie et al., 1999). Therefore, it is very limited to merely focus on DEGs between normal and tumor cells, and more attention should be paid to combination of gene expression pattern and clinical features, such as tumor stage, histological grade, invasiveness, etc.

Co-expression networks are widely used to decipher disease mechanisms and provide a systematic view of dysregulation pathways (Nayak et al., 2009). The basic theory of co-expression analysis is that co-expressed genes may be functionally related. Weighted gene co-expression network analysis (WGCNA) is an open source tool to perform co-expression analysis based on the theory. WGCNA integrates the expression differences between samples into a higher-order network structure, and clarifies the relationship between genes based on their co-expression profiles. The WGCNA algorithm has been used to investigate the associations between gene modules and clinical indicators in the field of cancer research (Langfelder and Horvath, 2008). Specifically, WGCNA was applied to identify key genes significantly associated with clinical indicators of tumor progression including tumor stage, grade, and metastasis (Chen L. et al., 2017; Chen P. et al., 2017). WGCNA has been used to identify biomarkers related to BC progression in recent publications. Tang et al. screened several prognostic genes including CCNB2, FBXO5, KIF4A, MCM10, and TPX2 using WGCNA (Tang et al., 2018). Another recent study suggested that four hub genes (FAM171A1, NDFIP1, SKP1, and REEP5) were identified as candidate biomarkers for BC (Tian et al., 2020). WGCNA was also used to identify key modules and pathways in BC. Our study intends to use this algorithm to identify biomarkers associated with progression of BC. We try to construct a co-expression network of genes and identify the network hub genes related to the clinical characteristics of BC, and use various databases (GEO, TCGA, and STRING) to verify our results.

MATERIALS AND METHODS

Data Collection

Normalized gene expression data and corresponding clinical information were downloaded from Gene Expression Omnibus (GEO)[1]. Datasets GSE42568 (Cheng et al., 2017) performed on the platform Affymetrix Human Genome U133 Plus 2.0 Array included gene expression profiling of 104 BC and 17 normal breast biopsies. GSE42568 was analyzed to screen differential expressed genes (DEGs). Dataset GSE102484 (Kao et al., 2011) also performed on Affymetrix Human Genome U133 Plus 2.0 Array included 683 BC samples, which was used to perform weighted gene co-expression networks. Dataset GSE20685 (Sabatier et al., 2011b) had gene expression profiles of 327 BC samples. Dataset GSE21653 (Sabatier et al., 2011a) had 266 samples and was used for Ki-67 correlation analysis and module preservation analysis. In addition, 992 BC samples with RNA-seq data were obtained from the Cancer Genome Atlas (TCGA) database. GSE20685 and TCGA were both used for stage validation.

Screening for DEGs

Normalized gene expression matrix and corresponding annotation files were obtained from GEO database. Firstly, we used the annotation files to annotate the probes. DEGs between normal and tumor breast samples were identified by R package "limma" (Ritchie et al., 2015). The cut-off criteria were the FDR (false discovery rate) < 0.01 and $|\log2(\text{fold change})| \geq 1$.

Weighted Gene Co-expression Network Construction

Based on the expression values of all DEGs of 683 BC samples and the corresponding clinical information (GSE102484), the "WGCNA" (Langfelder and Horvath, 2008) R package was used for the co-expression network (Langfelder and Horvath, 2008). Before constructing the co-expression network, outlier samples should be excluded by sample clustering using Pearson's method. According to the tutorial of WGCNA, we firstly verified the qualification of genes and samples. Then we construct the Pearson correlation matrix, and use the formula $amn = |cmn|^{\beta}$ (cmn represents the Pearson correlation between genes, amn represents the adjacency between genes, β parameter can amplify the correlation between genes) to obtain the weighted adjacency matrix. The soft threshold power β is determined based on the standard scale-free network. Subsequently, we converted the adjacency relationship into a topological overlap matrix (TOM) (Yip and Horvath, 2007), and hierarchically clustered genes to identify modules containing similar genes. In this study, we selected the minimum size as 30 for the gene dendrogram, selected the cutting line (0.25) for the modular dendrogram, to merge some similar modules.

Identify Modules With Clinical Significance and Functional Annotations

It is of great biological significance to identify modules most significantly associated with clinical features. Based on the similarity expressed in samples, gene modules with clinical significance were identified by correlation analysis. We selected the gene modules most relevant to clinical features as the modules of interest, and analyzed their correlation with clinical features. In addition, in order to further clarify the potential mechanism of module of interest affecting clinical features, the genes were uploaded to DAVID (Dennis et al., 2003) (The Database

[1]http://www.ncbi.nlm.nih.gov/geo/

for Annotation, Visualization, and Integrated Discovery) for GO function enrichment analysis. False transmission rate (FDR) < 0.01 was considered statistically significant.

Module Preservation Analysis

Module preservation analysis was conducted to ensure the identified gene modules could also be found in the test network (Langfelder et al., 2011). To evaluate the module preservation, we applied median rank and Zsummary via permutation testing in the "WGCNA" package. Considering the computational complexity involved in the size of our network, the permutation was executed 200 times. According to the threshold set in a previous study (Langfelder et al., 2011), modules with ZSummary scores > 10 indicate preservation, 2–10 indicate weak to moderate preservation, and < 2 indicate no preservation in the permutation. The dataset GSE21653 was used for preservation analysis including 266 BC samples.

Hub Gene Identification and Validation

Hub genes have a significant correlation with clinical characteristics (Gene significance, GS), and have a high module characterization (Module Membership, MM) in the module. Hub genes, also called key genes, are a group of genes with the highest connectivity, and determine the characteristics of the gene module. There are two ways to identify the key genes in the module according to the official tutorial of WGCNA (Langfelder and Horvath, 2008). The first is to directly determine the key genes based on GS and MM greater than a certain threshold. Specifically, the screening criterion with GS greater than 0.2 and MM of more than 0.8 are often used. The second is to use the "networkScreening" function, which can be used to calculate the weighted *P*-value p.weighted of each gene. Our study chose the first way to identify hub genes. In order to ensure the reliability of the hub genes, we used other independent datasets to validate the expression of hub genes in different tumor stages. We used BC samples from other independent datasets to compare hub gene expression at different pathological stages. We also obtained prognostic data for hub genes and analyzed the survival time of each gene.

RESULTS

Screening DEGs in BC Tissue Samples

The flow chart of the study is shown in **Supplementary Figure S1**. When the cut-off criteria is FDR < 0.05 and |log2 (FC)| ≥ 1, 3046 DEGs were screened between 104 BC and 17 normal breast biopsies from dataset GSE42568. The heatmap of all DEGs was shown in **Figure 1A**. Pathway and functional enrichment analysis showed that the upregulated DEGs were significantly enriched in cell proliferation and migration related pathways, including cell division, positive regulation of cell proliferation, cell–cell adhesion, cell migration etc. The downregulated DEGs were associated with metabolism related pathways, such as metabolic process, glucose homeostasis, and fatty acid beta-oxidation.

Identifying Co-expression Network and Module Preservation Analysis

Co-expression analysis included 683 BC samples and their complete clinical data and 3046 differential gene expression data. Four outlier samples were excluded after the samples were clustered by correlation analysis (**Figure 1C** and **Supplementary Figure S2A**). We used WGCNA R package and classified differential genes with similar expression patterns into different modules by average link clustering. When the soft threshold β was selected as 8, the genes in the network was scale-free (**Supplementary Figures S2B–E**). Different modules were identified, and the genes in the same module had a similar co-expression trend. A total of 8 modules were identified after the modules with a similar co-expression trend were combined (**Figure 2A**). The genes in the gray module were not co-expressed (**Figures 2A,B**). We did module preservation analysis by comparing the identified gene modules above with the test dataset GSE21653 to ensure the repeatability of the modules. As shown in **Figures 3A,B**, since the Zsummary statistic of the blue module was higher than 10 and the median rank statistic was close to the minimum in the test dataset, the module showed considerable stability.

Identifying Clinically Significant Gene Modules

The main purpose of WGCNA is calculating the correlation between different modules and clinical features, and identifying the modules most relevant to clinical features, which has important biological significance. We used Pearson's correlation analysis to calculate the correlation coefficients between different gene modules and clinical features, and found that the blue module and tumor stage (R = 0.24, p = 1e–10) has the highest correlation, and it also has a significant correlation with the tumor T stage (t.stage, R = 0.23, p = 2e–9) **Figure 1B**. The bar plot in **Figure 2C** also showed that the blue module had the highest gene significance across all modules. Therefore, the blue module is identified as a clinically significant module for further analysis. To investigate the functional role of the 504 genes of the blue module, we performed GO enrichment analysis and found that the biological process mainly focused on cell cycle and cell division (all p < 0.01, **Figure 3C**). Under the threshold of p < 0.01, **Figure 3C** showed the genes included in the three top biological processes including chromosome segregation, mitotic nuclear division, and organelle fission.

Identification and Validation of Hub Genes

Different methods were used to identify the hub gene from the hub module. Firstly, 504 genes in the stage-related module (blue module) are screened by module membership (MM) and gene significance (GS). As mentioned in section "Materials and Methods," when the absolute value of MM is greater than 0.8 and the absolute value of GS is greater than 0.2, 36 hub genes were identified (**Figure 3E**). The PPI network showed that 62 genes with the top connectivity degree were identified as hub genes under the cutoff of confidence > 0.4

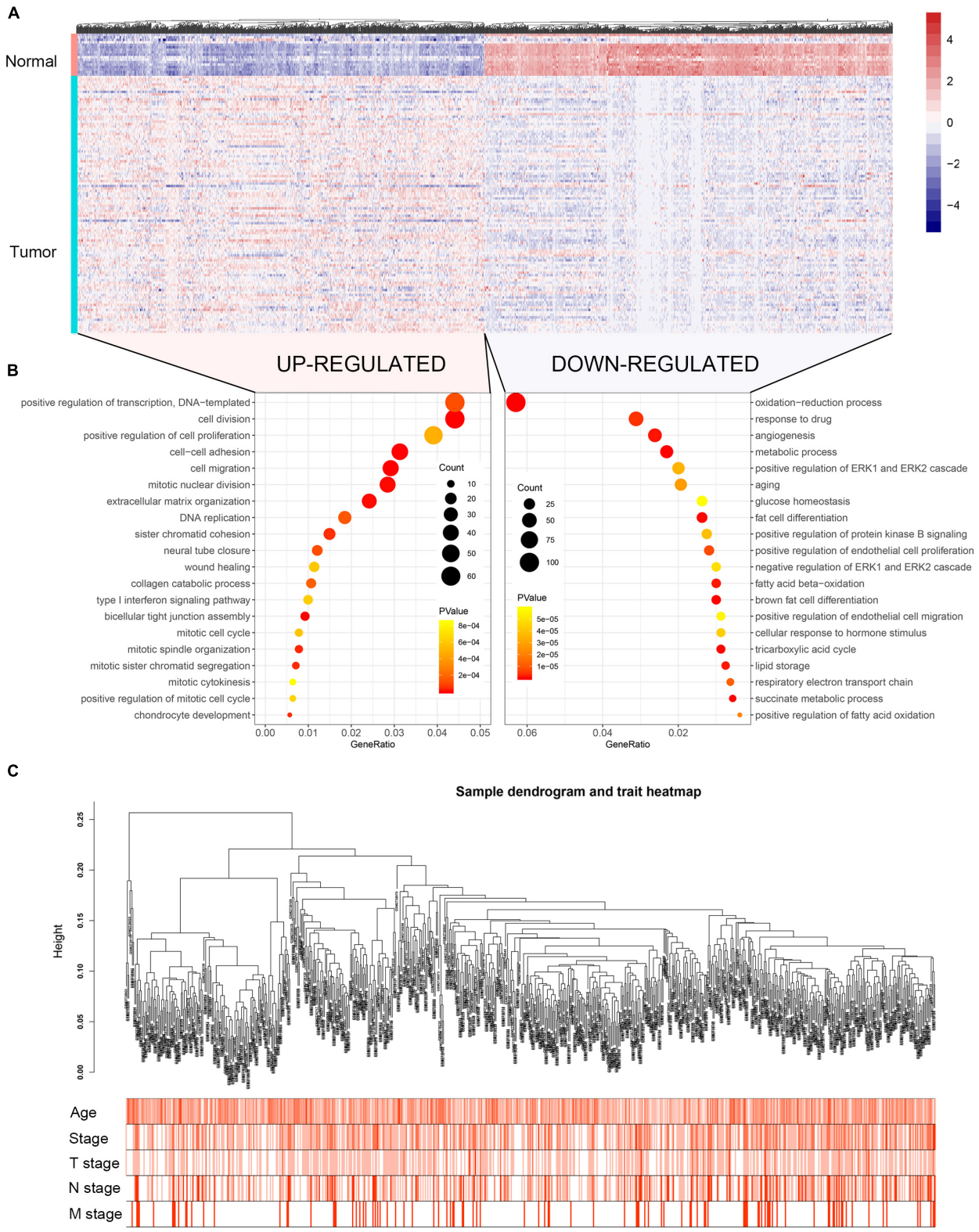

FIGURE 1 | Heatmap of the DEGs, GO functional annotation and clustering dendrogram of 683 tumor samples. **(A)** The heatmap shows the DEGs between 104 BC and 17 normal breast samples based on the dataset GSE42568. **(B)** The bubble plot shows the enriched biological processes of the upregulated genes and downregulated genes. **(C)** The clustering of 683 BC samples based on all DEGs expression and clinical features. The color intensity represents older age and higher stage.

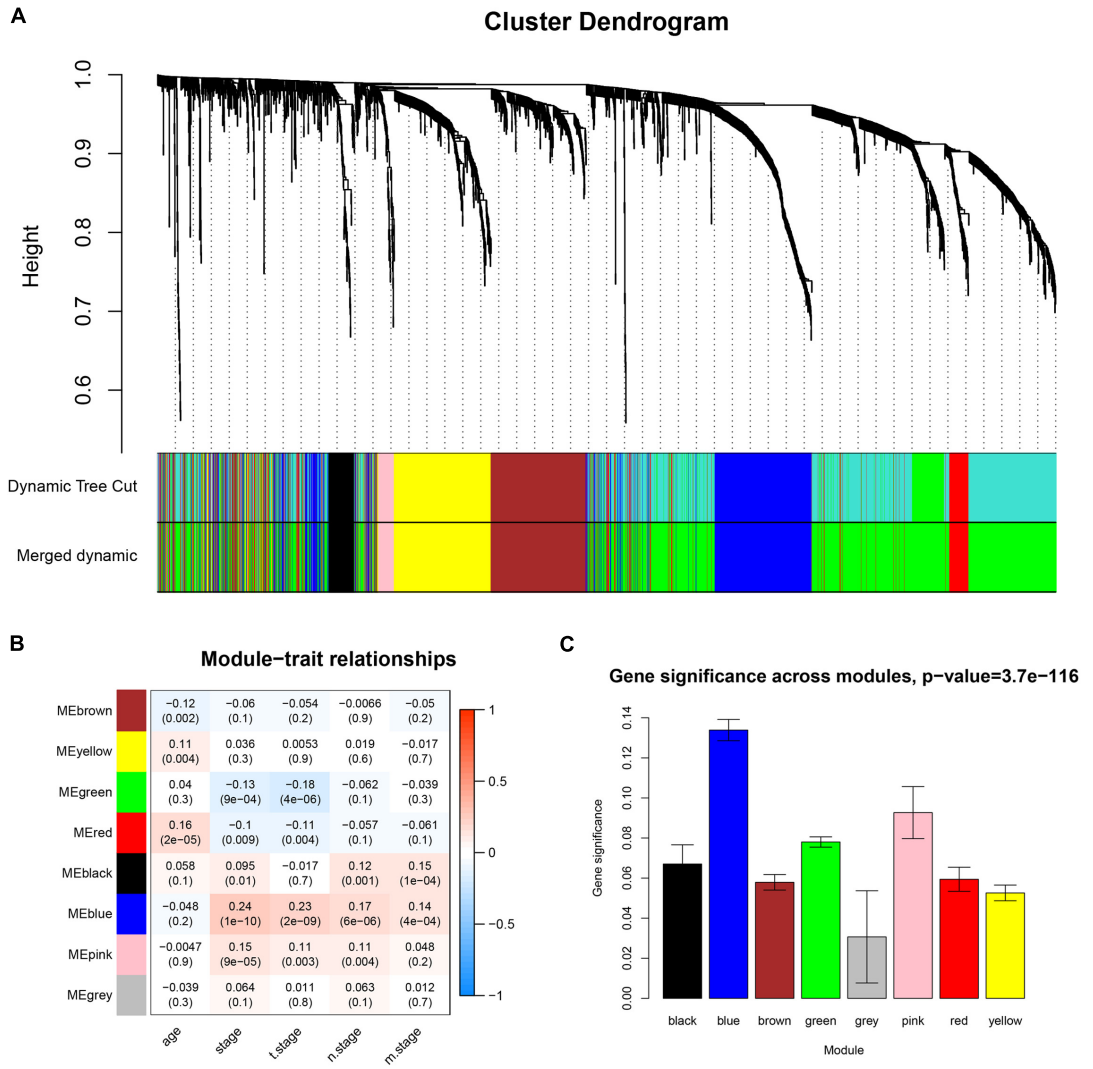

FIGURE 2 | Identifying co-expressed modules and the correlation between modules and the clinical features. **(A)** Dendrogram of all DEGs clustered based on a dissimilarity measure (1-TOM). **(B)** Correlation heatmap showed the correlation coefficients between gene modules and clinical features of BC. **(C)** The bar plot showed the distribution of average gene significance at different modules.

and connectivity degree of $\geq$ 4 (node/edge) (**Figure 3D** and **Supplementary Figure S6**). One-way ANOVA analyses were performed to validate candidate hub genes in the datasets GSE102484, GSE20685, and TCGA-BRCA, and 30 of 36 genes could be verified. As tumor prognosis was always affected by tumor progression, the candidate hub genes were validated by overall survival analysis (OS), recurrence-free survival analysis (RFS), and metastasis-free survival analysis (MFS), which showed that most of the hub genes had significant *P*-values in different test sets (**Figures 3F**, **6**, **7** and **Supplementary Figures S4, S5**). We regarded the common genes with statistical significance in different methods to the candidate hub genes, and 12 genes were finally screened (AURKA, BUB1B, CCNB2, CDK1, CDT1, HJURP, KIF20A, KIF2C, KIF4A, MELK, TPX2, UBE2C) (**Figures 4A–C**). As we all know, MKi67 is a cell proliferation marker, and the correlation coefficient between the candidate hub gene and MKi67 was calculated by Pearson correlation. The results showed that the expression of 12 candidate hub genes was highly positively correlated with MKi67. In addition, BC samples with stronger KI67 IHC staining showed higher gene expression of hub genes (**Figures 5A–C**).

DISCUSSION

By 2019, about 268,600 invasive BC and 48,100 DCIS cases were diagnosed among American women, and 41,760 women will die of the disease. About 13% of women will be diagnosed with invasive BC (DeSantis et al., 2019). From 2009 to 2015, the 5 years survival rate for women diagnosed with BC was stage I: 98%, stage II: 92%, stage III: 75%, and stage IV: 27% (Marinac et al., 2016). Because TNM staging focuses on the anatomical information of

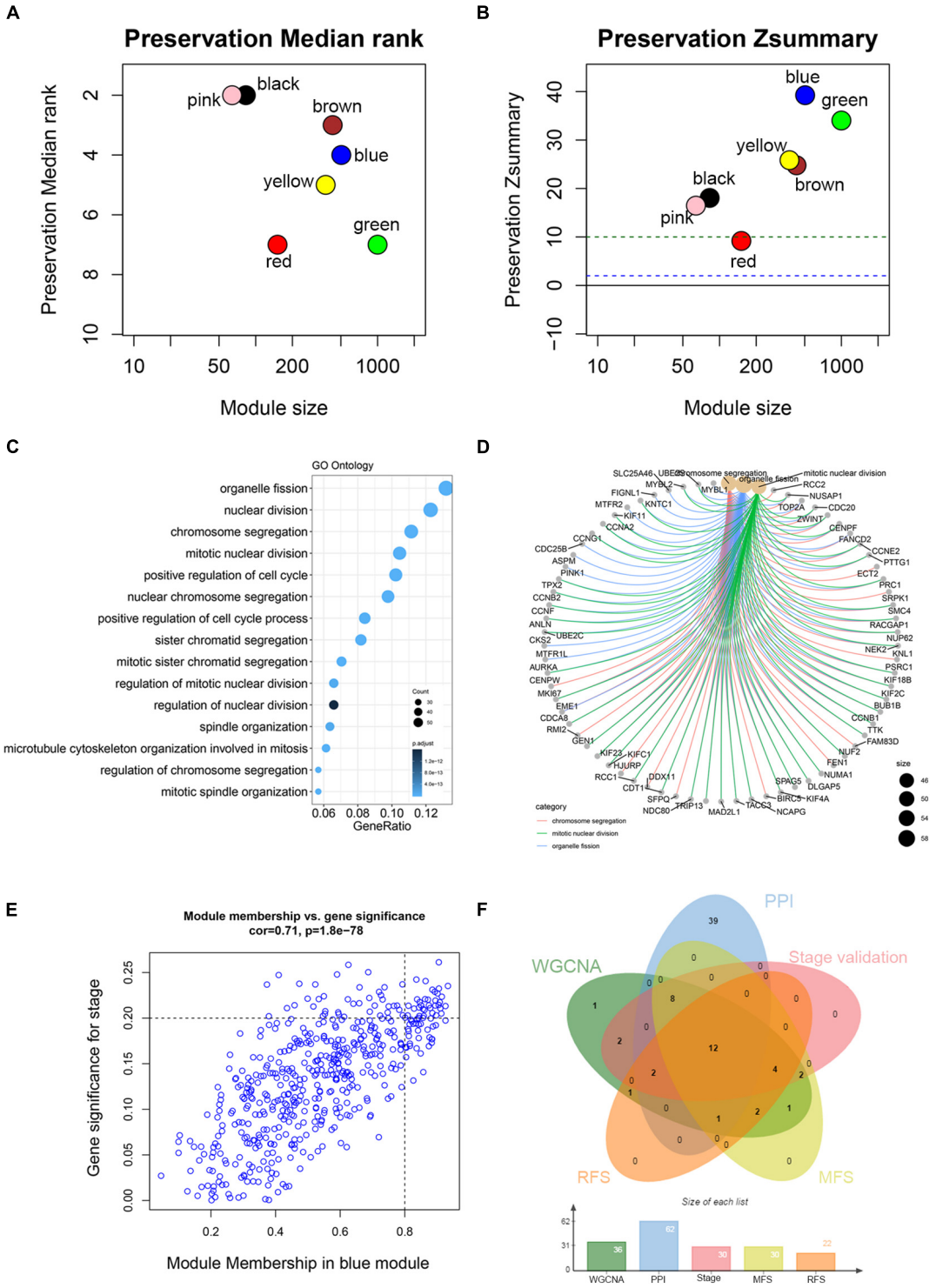

FIGURE 3 | Module preservation analysis and functional enrichment of clinically significant module. **(A)** The medianRank statistics of the module preservation using independent dataset GSE21653. The medianRank of the modules close to zero indicates a high degree of module preservation. **(B)** The Zsummary statistics of the module preservation using independent dataset GSE21653. These horizontal lines represent the Zsummary threshold ($Z = 2$ and $Z = 10$), which is used to indicate strong evidence of preservation (above 10) and low to moderate preservation (above 2). **(C)** Biological processes of genes in the blue module. The *x*-axis is the gene ratio, which is the ratio of enriched genes to the total number of genes in the term. **(D)** The three top-ranked biological processes and the corresponding genes of the blue module. **(E)** Scatter plot shows the gene module membership and gene significance of the blue module. **(F)** Common genes with statistical significance in different methods, including survival, one-way ANOVA, Pearson's correlation, and co-expression analysis.

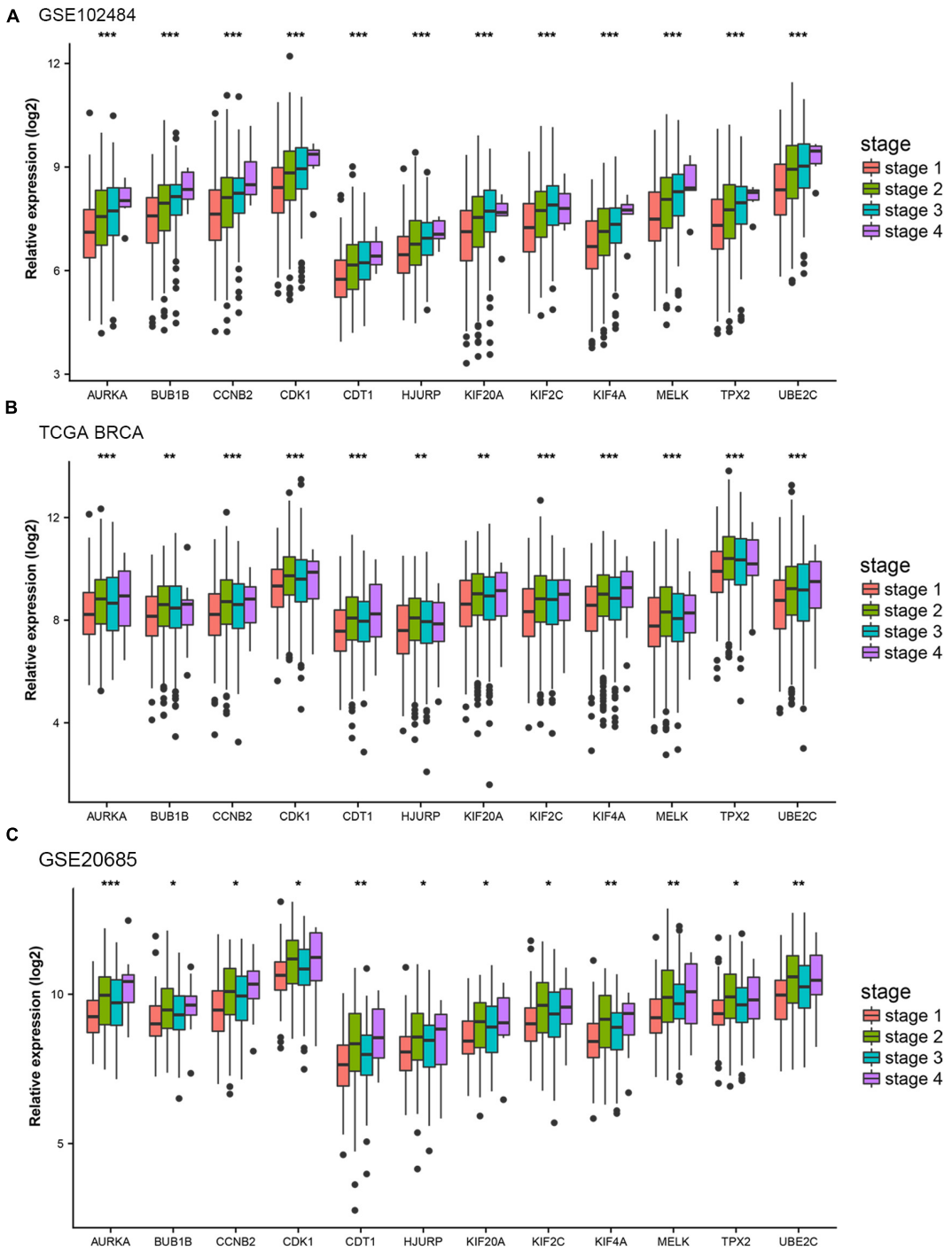

FIGURE 4 | Stage validation of hub gene. **(A)** Relative expression of the 12 hub genes at different stages in the GSE102484. **(B)** Relative expression of the 12 hub genes at different stages in TCGA BRCA. **(C)** Relative expression of the 12 hub genes at different stages in the GSE20685. The medians and dispersions are shown in the boxplot. One-way ANOVA is used to test statistical significance. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

the tumor, the disease progression and prognosis of BC patients cannot be fully evaluated. So, our study aimed to find biomarkers that could adequately predict BC progression and prognosis.

WGCNA has been widely used in the screening of biomarkers that predict disease progression (Chen L. et al., 2017). WGCNA is an algorithm for mining gene module information from expression profile analysis chips, and it has been widely used in gene expression profile data analysis (Langfelder and Horvath, 2008). In this method, a module is defined as a set of genes, where genes have similar expression trends in different physiological

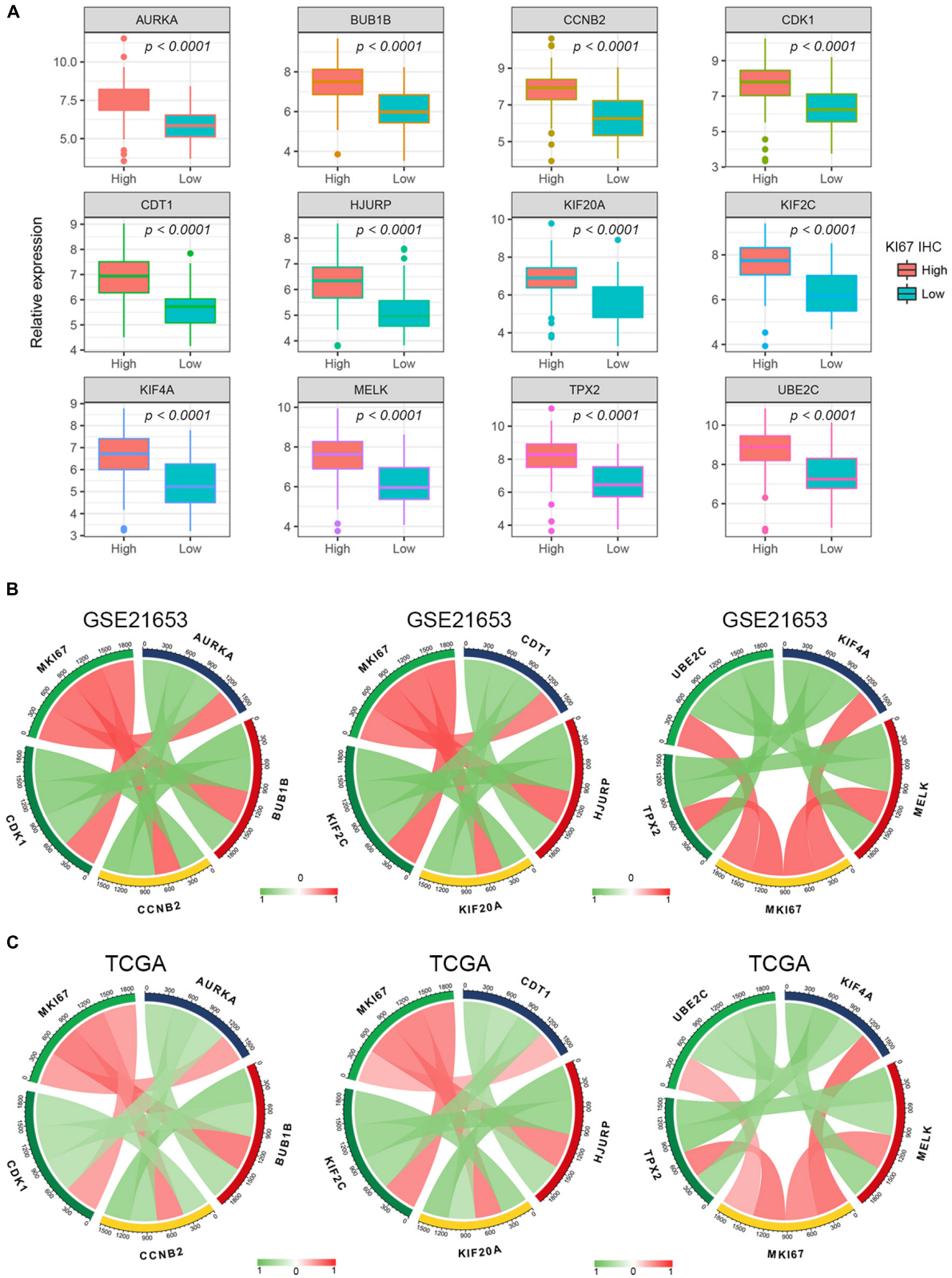

FIGURE 5 | Pearson's correlation analysis for the expression of Ki67 and the candidate hub genes. **(A)** The gene expression of the hub genes in Ki-67 IHC staining high and low BC samples. **(B,C)** The Pearson's correlation between the expression of MKi67 and the hub genes.

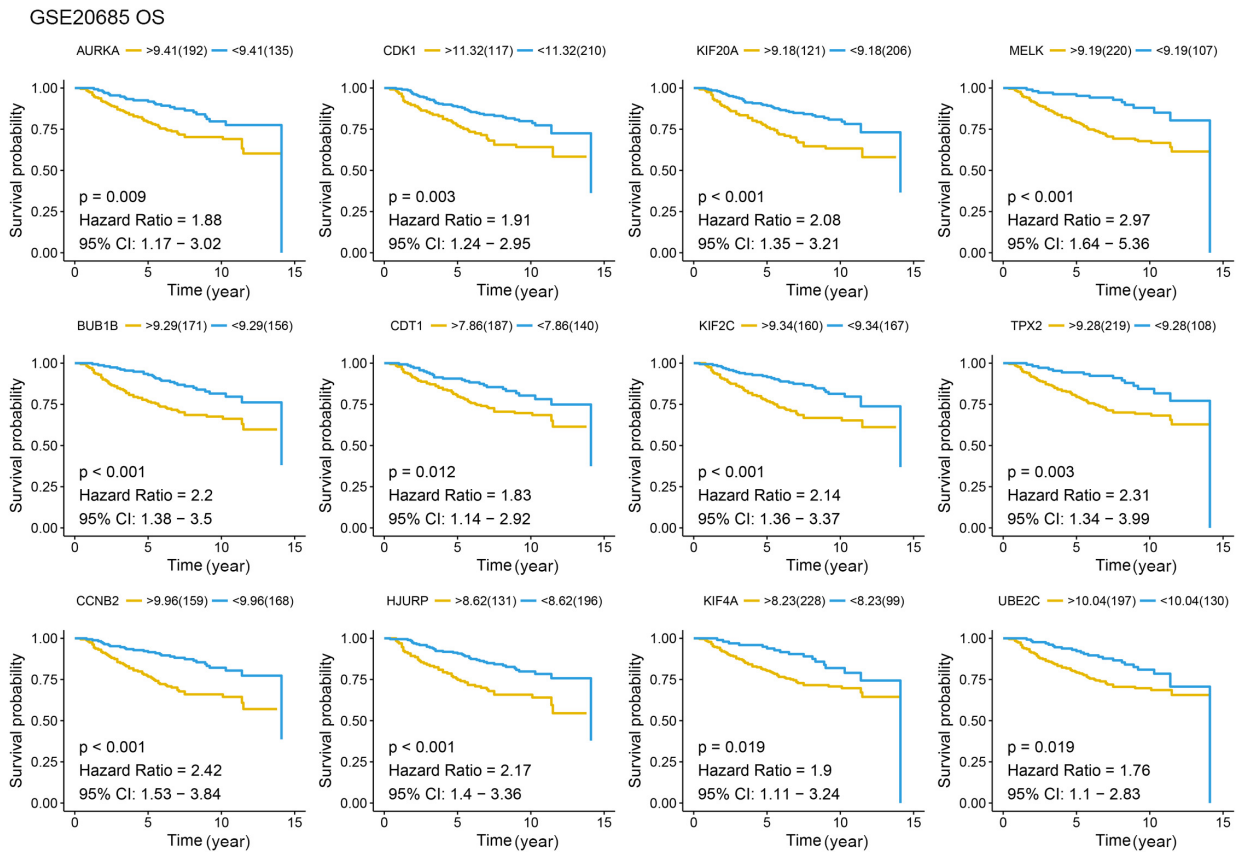

FIGURE 6 | Overall survival (OS) analysis of the candidate hub genes. Overall survival analysis of the candidate hub genes based on GSE20685. The red line is the high gene expression group, and blue line is the low gene expression group. The unit of time is year.

processes or different samples. After identifying gene modules with WGCNA, the correlation between gene modules and clinical characteristics such as tumor stage and grade is calculated. In this way, the gene modules most relevant to clinical features can be used to explore the main causes of tumor development. The characteristic of the scale-free network is that there are some nodes, the connectivity of these nodes is much higher than that of ordinary nodes, and the "few" nodes genes are defined as central genes (Niemira et al., 2019). Therefore, the study on the correlation between the module of interest and certain clinical features can be simplified to the correlation between the module of interest and the hub genes, so as to provide an important molecular basis for studying the mechanism of disease (Chen L. et al., 2019; Tian et al., 2019; Wang et al., 2019). By comparing different histological levels of BC, molecular targets have been identified to distinguish different stages of BC (Tian et al., 2020). We use systematic biology methods to identify specific biomarkers of BC based on a large number of samples. In our study, eight co-expression gene modules were determined by the dynamic tree cutting method. Correlation analysis shows that the blue module has the highest correlation with tumor staging, identifying the hub gene with the highest connectivity from the hub module. The functional annotations of clinical related modules suggest focusing on the cell proliferation related pathways, such as organelle fission, nuclear division, and chromosome segregation. WGCNA has been used to identify biomarkers related to BC progression. In comparation with recent studies, Tang et al. used WGCNA to screen several prognostic genes, including CCNB2, FBXO5, KIF4A, MCM10 and TPX2. These five genes are all related to cell division, which is consistent with our results. Among them, three genes are consistent with the results we found. However, the difference is that we have more BC samples to discover and validate, and our results are complementary to their findings. In addition, we have included more methods, including module preservation analysis and protein–protein interaction (PPI) to make our findings convincing.

In this study, 12 pathological hub genes (AURKA, BUB1B, CCNB2, CDK1, CDT1, HJURP, KIF20A, KIF2C, KIF4A, MELK, TPX2, and UBE2C) that are significantly related to the pathological stage were identified and verified, and significant differences can also be found in the expression value of each hub gene between different tumor stages and grades. Further verification also confirmed that the 12 hub genes were positively correlated with the progression of BC, and their expression was also related to the prognosis of BC patients. Aurora kinase A (AURKA), a member of the serine/threonine kinase family, plays an important role in mitotic cell division and genetic instability

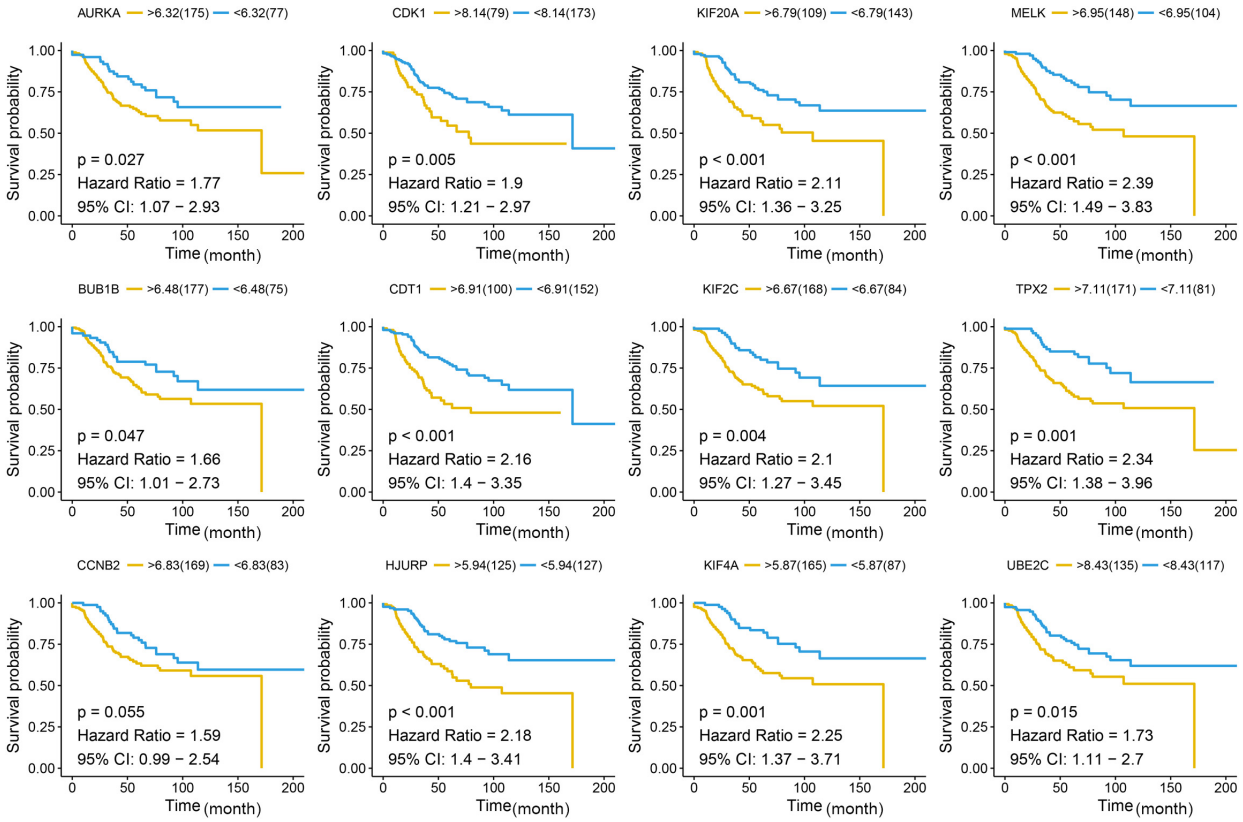

FIGURE 7 | Disease-free survival (DFS) analysis of the candidate hub genes. Disease-free survival of the candidate hub genes based on GSE21653. The red line is the high gene expression group, and blue line is the low gene expression group. The unit of time is month.

(Wu et al., 2018). It has been reported to stabilize FOXM1 by attenuating its ubiquitination in triple-negative breast cancer (TNBC), thus promoting proliferation of TNBC cells (Gartel, 2017). It has also been found to inhibit autophagy induction, suggesting that it may be the mechanism of drug-resistant BC cell apoptosis26. BUB1 mitotic checkpoint serine/threonine kinase B (BUB1B) encodes is a kinase involved in spindle testing. This protein plays a key role in the cell cycle (Lee et al., 2017). Its mRNA level has been found to be associated with intrachromosomal instability (Lee et al., 2017). Cyclin-dependent kinase 1 (CDK1) is a mitotic kinase, it mainly mediates tumor-related cell cycle defects, misregulated CDK1 may cause tumor cell proliferation and genome instability (Prevo et al., 2018). It has been reported that CDK1 could directly phosphorylate AMPK and promote the progress of BC (Galindo-Moreno et al., 2017). Other hub genes also play an important role in promoting cancer in BC.

There are still some limitations to our research. First, all data used in our study were based on publicly available datasets without validating in prospective clinical trials. Moreover, several important clinical factors, such as tumor size and lymph node metastasis, were not provided in these datasets. Finally, the mechanism between these gene signatures and BC recurrence still needs further experimental verification. In conclusion, through high-throughput screening and further screening by the WGCNA algorithm, we finally identified 12 hub genes that were significantly related to the progress and prognosis of BC after layers of validation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Jiangsu College of Nursing. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GS and WY conceived and designed the study and contributed to the writing of the manuscript. GS, ZS, and YL performed the analysis procedures, analyzed the results, and contributed analysis tools. All authors reviewed the manuscript.

FUNDING

The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We are very grateful for the NCBI, GEO, and TCGA databases for their free use.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.597888/full#supplementary-material

Supplementary Figure 1 | The flow chart of the study.

Supplementary Figure 2 | Sample clustering and determining the soft threshold power in WGCNA. **(A)** Sample clustering before removing the outlier samples. Four outlier samples could be found with red marks and were removed from further analysis. **(B)** Scale-free fitting index analysis of different soft threshold power (β). **(C)** The average connectivity analysis of various soft threshold powers. **(D)** The distribution of connectivity when β is 4. **(E)** Check the scale-free topology when β is 4.

Supplementary Figure 3 | Boxplots of hub genes at different grades in the GSE42568. The relative expression of hub genes at different tumor grades. One-way ANOVA was used to test statistical significance at different tumor grades. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Supplementary Figure 4 | Metastasis-free survival (MFS) analysis of the candidate hub genes. Metastasis-free survival analysis of the candidate hub genes based on GSE21653. The unit of time is year.

Supplementary Figure 5 | Recurrence-free survival (RFS) analysis of the candidate hub genes. Recurrence-free survival of the candidate hub genes based on GSE42568. The unit of time is day.

Supplementary Figure 6 | Protein–protein interaction (PPI) network of the genes in clinically significant modules.

REFERENCES

Berry, D. A., Cronin, K. A., Plevritis, S. K., Fryback, D. G., Clarke, L., Zelen, M., et al. (2005). Effect of screening and adjuvant therapy on mortality from breast cancer. *N. Engl. J. Med.* 353, 1784–1792. doi: 10.1056/NEJMoa050518

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 Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Cejalvo, J. M., Pérez-Fidalgo, J. A., Ribas, G., Burgués, O., Mongort, C., Alonso, E., et al. (2016). Clinical implications of routine genomic mutation sequencing in PIK3CA/AKT1 and KRAS/NRAS/BRAF in metastatic breast cancer. *Breast Cancer Res. Treat.* 160, 69–77. doi: 10.1007/s10549-016-3980-z

Chen, L., Peng, T., Luo, Y., Zhou, F., Wang, G., Qian, K., et al. (2019). ACAT1 and metabolism-related pathways are essential for the progression of clear cell renal cell carcinoma (ccRCC), as determined by co-expression network analysis. *Front. Oncol.* 9:957. doi: 10.3389/fonc.2019.00957

Chen, L., Yuan, L., Wang, Y., Wang, G., Zhu, Y., Cao, R., et al. (2017). Co-expression network analysis identified FCER1G in association with progression and prognosis in human clear cell renal cell carcinoma. *Int. J. Biol. Sci.* 13, 1361–1372. doi: 10.7150/ijbs.21657

Chen, P., Wang, F., Feng, J., Zhou, R., Chang, Y., Liu, J., et al. (2017). Co-expression network analysis identified six hub genes in association with metastasis risk and prognosis in hepatocellular carcinoma. *Oncotarget* 8, 48948–48958. doi: 10.18632/oncotarget.16896

Cheng, S. H., Huang, T. T., Cheng, Y. H., Tan, T. B. K., Horng, C. F., Wang, Y. A., et al. (2017). Validation of the 18-gene classifier as a prognostic biomarker of distant metastasis in breast cancer. *PLoS One* 12:e0184372. doi: 10.1371/journal.pone.0184372

Dahinden, C., Ingold, B., Wild, P., Boysen, G., Luu, V. D., Montani, M., et al. (2010). Mining tissue microarray data to uncover combinations of biomarker expression patterns that improve intermediate staging and grading of clear cell renal cell cancer. *Clin. Cancer Res.* 16, 88–98. doi: 10.1158/1078-0432.ccr-09-0260

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4:P3.

DeSantis, C. E., Ma, J., Gaudet, M. M., Newman, L. A., Miller, K. D., Goding Sauer, A., et al. (2019). Breast cancer statistics, 2019. *CA Cancer J. Clin.* 69, 438–451. doi: 10.3322/caac.21583

Early Breast Cancer Trialists' Collaborative Group, Peto, R., Davies, C., Godwin, J., Gray, R., Pan, H. C., et al. (2012). Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet* 379, 432–444. doi: 10.1016/S0140-6736(11)61625-5

Galindo-Moreno, M., Giráldez, S., Sáez, C., Japón, M. Á, Tortolero, M., and Romero, F. (2017). Both p62/SQSTM1-HDAC6-dependent autophagy and the aggresome pathway mediate CDK1 degradation in human breast cancer. *Sci. Rep.* 7:10078. doi: 10.1038/s41598-017-10506-8

Garcia-Murillas, I., Schiavon, G., Weigelt, B., Ng, C., Hrebien, S., Cutts, R. J., et al. (2015). Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci. Transl. Med.* 7:302ra133. doi: 10.1126/scitranslmed.aab0021

Gartel, A. L. (2017). FOXM1 in cancer: interactions and vulnerabilities. *Cancer Res.* 77, 3135–3139. doi: 10.1158/0008-5472.CAN-16-3566

Gerlinger, M., Horswell, S., Larkin, J., Rowan, A. J., Salm, M. P., Varela, I., et al. (2014). Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* 46, 225–233. doi: 10.1038/ng.2891

Howlader, N., Altekruse, S. F., Li, C. I., Chen, V. W., Clarke, C. A., Ries, L. A., et al. (2014). US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *J. Natl. Cancer Inst.* 106:dju055. doi: 10.1093/jnci/dju055

Kao, K. J., Chang, K. M., Hsu, H. C., and Huang, A. T. (2011). Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC Cancer* 11:143. doi: 10.1186/1471-2407-11-143

Kontzoglou, K., Palla, V., Karaolanis, G., Karaiskos, I., Alexiou, I., Pateras, I., et al. (2013). Correlation between Ki67 and breast cancer prognosis. *Oncology* 84, 219–225. doi: 10.1159/000346475

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Langfelder, P., Luo, R., Oldham, M. C., and Horvath, S. (2011). Is my network module preserved and reproducible? *PLoS Comput. Biol.* 7:e1001057. doi: 10.1371/journal.pcbi.1001057

Lee, E., Pain, M., Wang, H., Herman, J. A., Toledo, C. M., DeLuca, J. G., et al. (2017). Sensitivity to BUB1B inhibition defines an alternative classification of glioblastoma. *Cancer Res.* 77, 5518–5529. doi: 10.1158/0008-5472.CAN-17-0736

Marinac, C. R., Nelson, S. H., Breen, C. I., Hartman, S. J., Natarajan, L., Pierce, J. P., et al. (2016). Prolonged nightly fasting and breast cancer prognosis. *JAMA Oncol.* 2, 1049–1055. doi: 10.1001/jamaoncol.2016.0164

McFaline-Figueroa, J. L., Hill, A. J., Qiu, X., Jackson, D., Shendure, J., and Trapnell, C. (2019). A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nat. Genet.* 51, 1389–1398. doi: 10.1038/s41588-019-0489-5

Meng, L., Xu, Y., Xu, C., and Zhang, W. (2016). Biomarker discovery to improve prediction of breast cancer survival: using gene expression profiling, meta-analysis, and tissue validation. *Onco Targets Ther.* 9, 6177–6185. doi: 10.2147/ott.s113855

Nayak, R. R., Kearns, M., Spielman, R. S., and Cheung, V. G. (2009). Coexpression network based on natural variation in human gene expression reveals gene interactions and functions. *Genome Res.* 19, 1953–1962. doi: 10.1101/gr.097600.109

Niemira, M., Collin, F., Szalkowska, A., Bielska, A., Chwialkowska, K., Reszec, J., et al. (2019). Molecular signature of subtypes of non-small-cell lung cancer by large-scale transcriptional profiling: identification of key modules and genes by weighted gene co-expression network analysis (WGCNA). *Cancers* 12:37. doi: 10.3390/cancers12010037

Prevo, R., Pirovano, G., Puliyadi, R., Herbert, K. J., Rodriguez-Berriguete, G., O'Docherty, A., et al. (2018). CDK1 inhibition sensitizes normal cells to DNA damage in a cell cycle dependent manner. *Cell Cycle* 17, 1513–1523. doi: 10.1080/15384101.2018.1491236

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:e47. doi: 10.1093/nar/gkv007

Sabatier, R., Finetti, P., Adelaide, J., Guille, A., Borg, J.-P., Chaffanet, M., et al. (2011a). Down-regulation of ECRG4, a candidate tumor suppressor gene, in human breast cancer. *PLoS One* 6:e27656. doi: 10.1371/journal.pone.0027656

Sabatier, R., Finetti, P., Cervera, N., Lambaudie, E., Esterni, B., Mamessier, E., et al. (2011b). A gene expression signature identifies two prognostic subgroups of basal breast cancer. *Breast Cancer Res. Treat.* 126, 407–420. doi: 10.1007/s10549-010-0897-9

Tang, J., Kong, D., Cui, Q., Wang, K., Zhang, D., Gong, Y., et al. (2018). Prognostic genes of breast cancer identified by gene co-expression network analysis. *Front. Oncol.* 8:374. doi: 10.3389/fonc.2018.00374

Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J., and Church, G. M. (1999). Systematic determination of genetic network architecture. *Nat. Genet.* 22, 281–285. doi: 10.1038/10343

Tian, A., Pu, K., Li, B., Li, M., Liu, X., Gao, L., et al. (2019). Weighted gene coexpression network analysis reveals hub genes involved in cholangiocarcinoma progression and prognosis. *Hepatol. Res.* 49, 1195–1206. doi: 10.1111/hepr.13386

Tian, Z., He, W., Tang, J., Liao, X., Yang, Q., Wu, Y., et al. (2020). Identification of important modules and biomarkers in breast cancer based on WGCNA. *Onco Targets Ther.* 13, 6805–6817. doi: 10.2147/OTT.S258439

Trainer, A. H., Thompson, E., and James, P. A. (2011). BRCA and beyond: a genome-first approach to familial breast cancer risk assessment. *Discov. Med.* 12, 433–443.

van Kessel, K. E. M., van der Keur, K. A., Dyrskjot, L., Algaba, F., Welvaart, N. Y. C., Beukers, W., et al. (2018). Molecular markers increase precision of the European association of urology non-muscle-invasive bladder cancer progression risk groups. *Clin. Cancer Res.* 24, 1586–1593. doi: 10.1158/1078-0432.ccr-17-2719

Wang, Y., Chen, L., Ju, L., Qian, K., Liu, X., Wang, X., et al. (2019). Novel biomarkers associated with progression and prognosis of bladder cancer identified by co-expression analysis. *Front. Oncol.* 9:1030. doi: 10.3389/fonc.2019.01030

Wu, C., Lyu, J., Yang, E. J., Liu, Y., Zhang, B., and Shim, J. S. (2018). Targeting AURKA-CDC25C axis to induce synthetic lethality in ARID1A-deficient colorectal cancer cells. *Nat. Commun.* 9:3212. doi: 10.1038/s41467-018-05694-4

Yip, A. M., and Horvath, S. (2007). Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics* 8:22. doi: 10.1186/1471-2105-8-22

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fgene.2020.625659

Adverse Drug Reaction Discovery Using a Tumor-Biomarker Knowledge Graph

*Meng Wang[1], Xinyu Ma[1], Jingwen Si[2], Hongjia Tang[3], Haofen Wang[4], Tunliang Li[3], Wen Ouyang[3], Liying Gong[5], Yongzhong Tang[3], Xi He[3], Wei Huang[6] and Xing Liu[3]**

[1] School of Computer Science and Engineering, Southeast University, Nanjing, China, [2] Department of Pharmaceutical Sciences, Tsinghua University, Beijing, China, [3] Department of Anesthesiology, Third Xiangya Hospital, Central South University, Changsha, China, [4] College of Design and Innovation, Tongji University, Shanghai, China, [5] Department of Intensive Care Unit, Third Xiangya Hospital, Central South University, Changsha, China, [6] Department of Cardiology, Third Xiangya Hospital, Central South University, Changsha, China

Adverse drug reactions (ADRs) are a major public health concern, and early detection is crucial for drug development and patient safety. Together with the increasing availability of large-scale literature data, machine learning has the potential to predict unknown ADRs from current knowledge. By the machine learning methods, we constructed a Tumor-Biomarker Knowledge Graph (TBKG) which contains four types of node: Tumor, Biomarker, Drug, and ADR using biomedical literatures. Based on this knowledge graph, we not only discovered potential ADRs of antitumor drugs but also provided explanations. Experiments on real-world data show that our model can achieve 0.81 accuracy of three cross-validation and the ADRs discovery of Osimertinib was chosen for the clinical validation. Calculated ADRs of Osimertinib by our model consisted of the known ADRs which were in line with the official manual and some unreported rare ADRs in clinical cases. Results also showed that our model outperformed traditional co-occurrence methods. Moreover, each calculated ADRs were attached with the corresponding paths of "tumor-biomarker-drug" in the knowledge graph which could help to obtain in-depth insights into the underlying mechanisms. In conclusion, the tumor-biomarker knowledge-graph based approach is an explainable method for potential ADRs discovery based on biomarkers and might be valuable to the community working on the emerging field of biomedical literature mining and provide impetus for the mechanism research of ADRs.

Keywords: adverse drug reaction, biomarker, knowledge graph, antitumor drugs, explainable model

OPEN ACCESS

Edited by:
Xiaofeng Dai,
Jiangnan University, China

Reviewed by:
Zequn Sun,
Nanjing University, China
Yanjun Zhang,
The University of Queensland, Australia
Jinghua Liu,
Huaqiao University, China

****Correspondence:***
Xing Liu
xingxingmail@csu.edu.cn

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *03 November 2020*
Accepted: *09 December 2020*
Published: *12 January 2021*

Citation:
Wang M, Ma X, Si J, Tang H, Wang H, Li T, Ouyang W, Gong L, Tang Y, He X, Huang W and Liu X (2021) Adverse Drug Reaction Discovery Using a Tumor-Biomarker Knowledge Graph. Front. Genet. 11:625659. doi: 10.3389/fgene.2020.625659

INTRODUCTION

Adverse drug reactions (ADRs) are a cause of significant morbidity and mortality in patients and a source of financial burden in the healthcare system (Patton and Borshoff, 2018). In tumor patients, pharmacokinetic parameters can be altered by the disease itself, or hepatic or renal impairment, or reduction of serum-binding proteins due to malnutrition. They experience a relatively high rate of ADRs from antitumor drugs and more easily experience rare and severe ADRs, which could seriously impact the quality of life (Shrestha et al., 2017).

The identification of rare and serious ADRs during the premarket period is limited due to the limited sample size and generalizability of clinical trials. Exploring the potential ADRs is critical to decrease the incidence. Therefore, great efforts have been devoted to detecting potential ADRs based on the data mining of literature databases or electronic health records (Bean et al., 2017; Lee and Chen, 2020). However, there are still two challenges to achieving good performance: (1) the unstructured biomedical literature contains many irrelevant words and contexts, and how to extract ADR-related entities and fully explore their relations (e.g., tumor-biomarker-drug) is difficult; and (2) some predicted unseen ADRs are unexpected and cause confusion, which means explainability and validation become critically important for automatic detection.

A knowledge graph (Wang et al., 2017a,b) is a data model that represents facts as nodes and relations between the nodes. Under a general medical information network, objects such as diseases, drugs, biomarkers, or treatments can all be linked together through different types of referential relationships, which enable the discovery of knowledge on a scale and at a speed that traditional pharmacologic experiments or clinical trials cannot approach. Recently, in addition to diagnosis and prognostication, biomarker (Califf, 2018; Carr and Pirmohamed, 2018) has been widely in tumor treatment to offer the opportunity to accurately and specifically predict therapeutic efficacy and safety during the course of antitumor therapy, which can provide oncologists with the opportunity to quickly modify a therapeutic regimen in ways that would provide the best therapy for their patients. However, there is little biomarker beginning to be applied to assess the ADRs. The aim of this study is to discover potential ADRs of anti-tumor drugs and provide explanations by constructing knowledge graph using literature data source. **Figure 1** summarizes the workflow of our study.

METHODS

Data Source

The biomedical database employed in this study was the MEDLINE database, which consists of more than 22 million journal citations and abstracts. The database is maintained by the National Library of Medicine (NLM). The MEDLINE corpus can be acquired -in XML format from http://www.nlm.nih.gov/bsd/licensee/access/medline_pubmed.html. Each citation contains the bibliographical information of an article, such as the article ID (PubMed Unique Identifier, PMID), article title, author list, journal title, venue, publication type, and indexed MeSH terms.

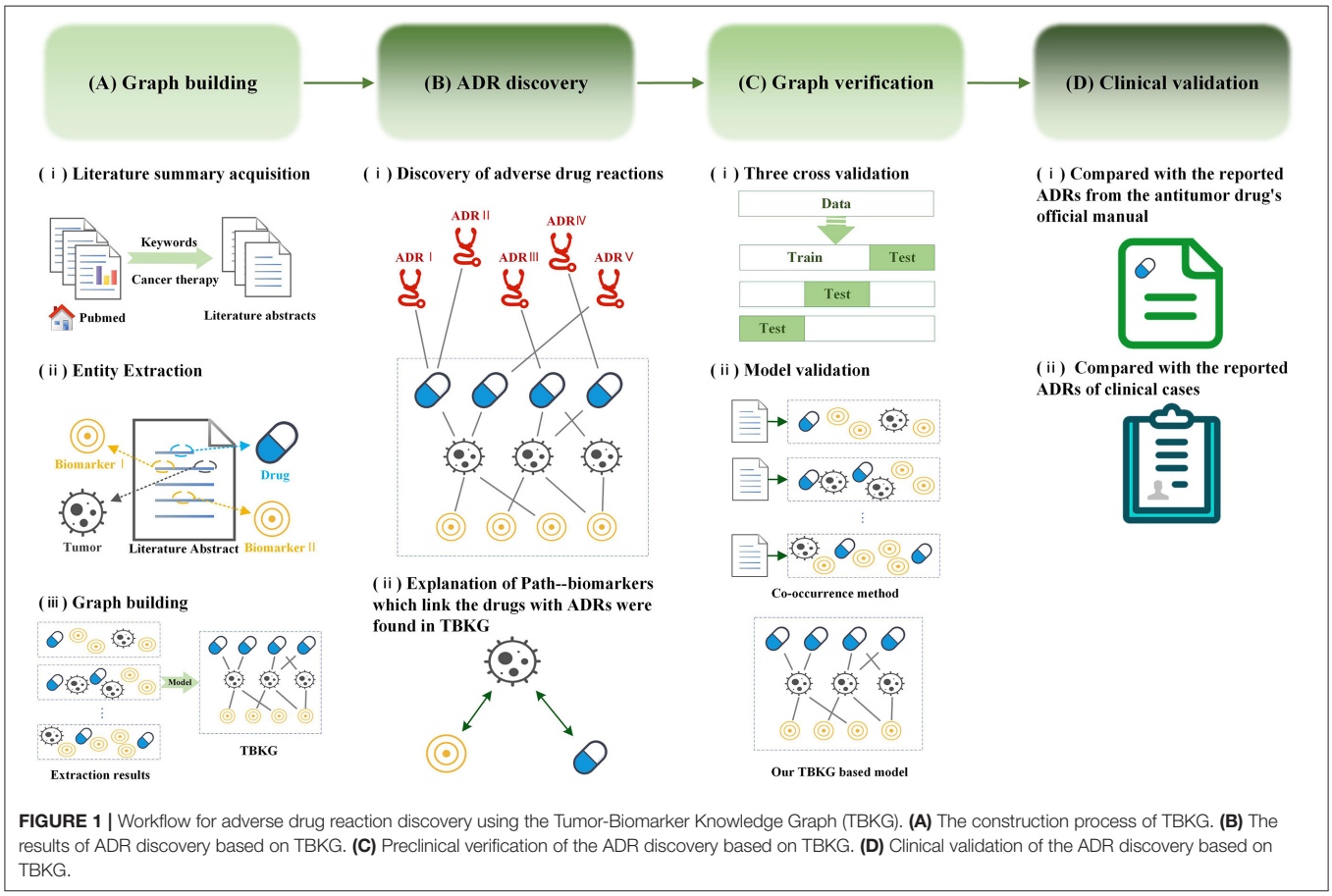

FIGURE 1 | Workflow for adverse drug reaction discovery using the Tumor-Biomarker Knowledge Graph (TBKG). **(A)** The construction process of TBKG. **(B)** The results of ADR discovery based on TBKG. **(C)** Preclinical verification of the ADR discovery based on TBKG. **(D)** Clinical validation of the ADR discovery based on TBKG.

TABLE 1 | The concept category used to build the dictionary.

Specific categories	Category meaning	Type in TBKG
T109	Organic Chemical	Biomarker-type
T114	Nucleic Acid, Nucleoside, or Nucleotide	Biomarker-type
T116	Amino Acid, Peptide, or Protein	Biomarker-type
T121	Pharmacologic Substance	Drug-type
T123	Biologically Active Substance	Biomarker-type
T125	Hormone	Biomarker-type
T126	Enzyme	Biomarker-type
T129	Immunologic Factor	Biomarker-type
T130	Indicator, Reagent, or Diagnostic Aid	Biomarker-type
T191	Neoplastic Process	Tumor-type
T192	Receptor	Biomarker-type
T195	Antibiotic	Biomarker-type
T200	Clinical Drug	Drug-type

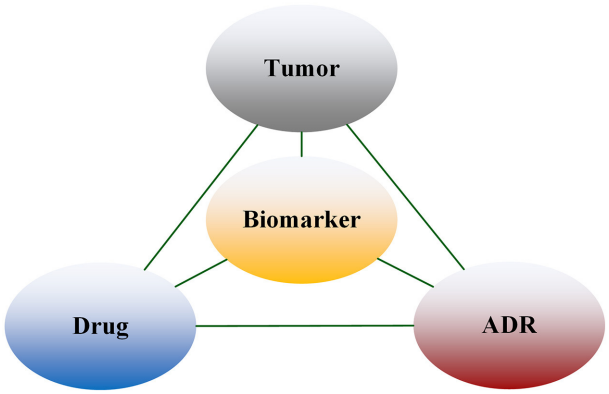

FIGURE 2 | Structure of TBKG.

MEDLINE is used as a surrogate for full-text articles. Permission to access the data were acquired by the 3rd Xiangya Hospital in China in May 2016.

The Unified Medical Language System (UMLS) Metathesaurus integrates the information of 216 source vocabularies and brings together many different types of biomedical vocabularies, mainly including 25% diagnosis, 25% procedures and supplies, 19% diseases and 14% drugs. Metathesaurus refines these categories to 127 different categories. As shown in **Table 1**, the category description of T191 is "Neoplastic Process," therefore, we fit it into tumor-type nodes. The category descriptions of T121 and T200 are "Pharmacologic Substance" and "Clinical Drug," respectively, we hence fit it into drug-type nodes. For ADR-type nodes, we use the WHO source dictionary in UMLS because WHO is used for coding clinical information related to adverse drug reactions (**Supplementary Table 1**). Finally, for biomarker-type nodes, there are many types of biomarker referring to the definition of biomarker (Carr and Pirmohamed, 2018), including genomic, immunogenetic, circulating protein, nucleic acid and so on. Therefore, we fit the corresponding categories T109, T114, T116, T123, T125, T126, T129, T130, T192, and T195 in **Table 1** into biomarker-type nodes.

Tumor-Biomarker Knowledge Graph-based ADR Discovery

Entity Extraction

From MEDLINE, we downloaded the abstracts of papers from 1928 to 2020 with "cancer therapy" as the key word. The article number, title, author, author unit, publication time, MeSH word, journal title and publication type were saved, and short abstracts were removed.

Four kinds of entities were extracted from the abstracts: tumors, biomarkers, drugs, and ADRs. We used the Metathesaurus 2020AA version provided by UMLS as the dictionary for entity extraction. Apache's open source tool cTAKES, which is a natural language processing system for extracting information from medical free texts, was utilized to extract entities. An entity mentioned positively was seen as an entity that was related to each abstract. Here, we limited the minimum frequency of entities to 50.

Relation Discovery

We used the entities extracted above to construct the Tumor-Biomarker Knowledge Graph (TBKG). TBKG is defined as G = (V, E), where G stands for TBKG, V is the set of vertices in G, and E is the set of edges in G. V contains the vertices of four entity types, namely tumor, biomarker, drug, and ADR. E contains undirected weighted edges. Each edge connects two different types of vertices. The weight on the edge represents the correlation (distance) between the two vertices. A basic schema is shown in **Figure 2**. Entities were transformed into matrix form. Each row of the matrix represented a summary file, and whether the entity appeared in the summary was represented by 0 and 1; this matrix data was taken as the input of the model.

A naive Bayesian model (Murphy, 2012) was utilized to explore correlations. The naive Bayesian model combined the prior probability and the posterior probability at the same time when building the graph, which could avoid the subjective bias from using only the prior probability and avoided the overfitting phenomenon from using the sample information alone at the same time. The calculation method for each relationship was the same, and we mainly used the calculation of the relationship between tumor and biomarker to illustrate the principle of this model. The parameter was learned by Maximum Likelihood Estimation. We learned a model for each tumor, as shown in **Figure 3**. An importance measure was used to determine whether there was a relationship between the tumor and biomarker:

$$IMPT_{NB} = \log\left(p\left(x_i = 1 \middle| y_j = 1\right)\right) - \log\left(p\left(x_i = 1 \middle| y_j = 0\right)\right), (1)$$

where x_i is 0 or 1 to indicate the presence of biomarker i and y_j is 0 or 1 to indicate the presence of tumor j. The reason why we use the importance measure is that if the presence of a biomarker makes it more likely that a tumor will be observed, we are more

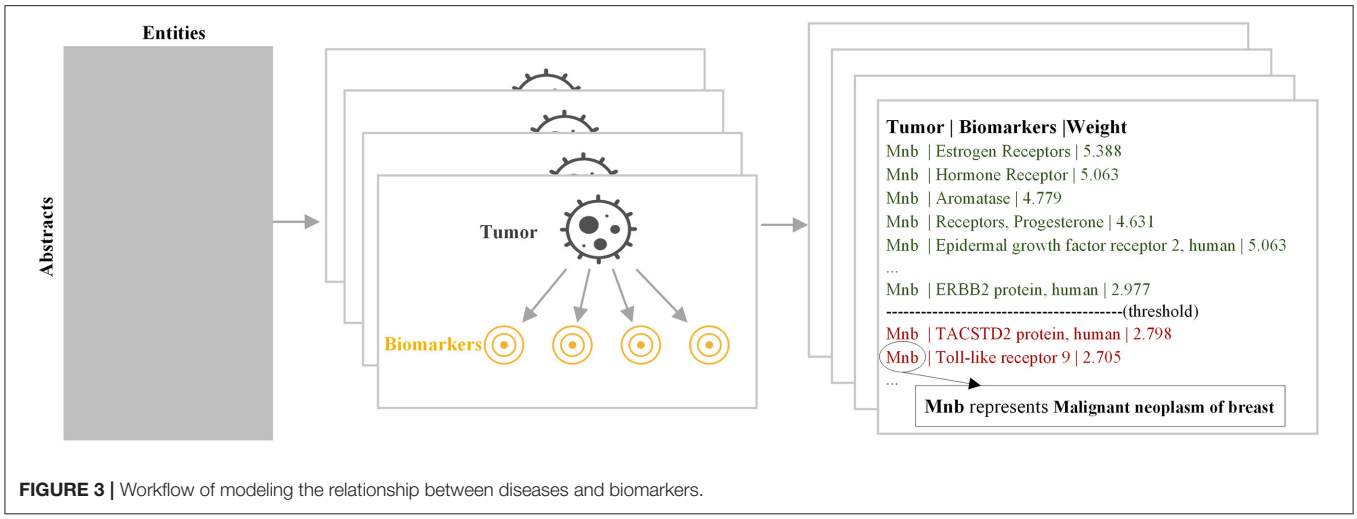

FIGURE 3 | Workflow of modeling the relationship between diseases and biomarkers.

confident that there is a relationship between them. Relationship whose importance was greater than a certain threshold was considered to exist.

ADR Discovery

For the ADR discovery based on TBKG, we collected all drugs and determined the corresponding ADR to form drug-ADR pairs as the calculated ADRs. The Depth First Search (DFS) algorithm was utilized to find every path between the drug and ADR, such as (drug, biomarker, ADR). Each output of ADR discovery contains a drug-ADR pair and all corresponding paths.

Experimental Settings

Accuracy With Cross-Validation

Three-fold cross-validation, which is mainly used to prevent overfitting caused by the model being too complex, was utilized to verify the effect of graph construction. The basic idea is to divide the original data into a training set to train the model and a test set to test the training results. The original data were randomly divided into three groups and each time, two groups were selected as the training set and the remaining group was used as the test set. This validation was repeated three times, and we took the average accuracy as the evaluation of the model.

Comparison With Co-Occurrence Analysis

We conducted a co-occurrence analysis (Callon et al., 1986) on the summaries and performed clinical verification on this result, which was compared with ADR discovery based on TBKG. The basic principle of co-occurrence analysis is to reflect the correlation strength between keywords by counting the co-occurrence of word pairs or noun phrases in the literature. According to this principle, we conducted frequency statistics and sorted the word pairs for all entities. Through clinical verification, we can compare the difference between the two results.

Clinical Validation

Clinical validation was performed to validate the efficacy of our model. Osimertinib is a third-generation epidermal growth factor receptor tyrosine kinase inhibitor that is used to treat non-small-cell lung carcinomas with specific mutations (Odogwu et al., 2018). This medication was approved as an antitumor treatment in 2017 by both the Food and Drug Administration and the European Commission. As a novel antitumor drug, neither clinical trials nor real-world studies had enough data to provide an early warning of ADRs after marketing. Thus, we chose osimertinib as an example.

TABLE 2 | Numbers of entities and relationships in TBKG.

Relationships	Entity numbers	Edge numbers
Tumor-Biomarker	1,179–2,550	30,065
Tumor-Drug	1,179–1,806	21,293
Tumor-ADR	1,179–756	8,913
Drug-Biomarker	1,806–2,550	46,052
Drug-ADR	1,806–756	13,653
Biomarker-ADR	2,550–756	19,278

First, the results of our model was compared with the reported ADRs from the official manual (**Supplementary Table 2**) and EGFR-TKI ADR Management Chinese Expert Consensus (Anti-Cancer Association, 2019). The ADRs of osimertinib were calculated for different literature quantities from 10,667 literature abstracts on "cancer therapy" since osimertinib appears in the literature in 2014. The ADRs ranked in the top 5% according to our model were defined as important ADRs, and those ranked in the bottom 5% were defined as unlikely ADRs. Kappa index, sensitivity and specificity were used to determine the reliability of our model with the official manual of osimertinib. The difference in ADR discovery was evaluated among different literature quantities, and the difference between our model and co-occurrence analysis was also measured. All analyses were performed using SPSS (version 23.0) statistical software.

Second, we also compared the results of our model with the reported ADRs of all clinical cases from the 3rd Xiangya Hospital. The clinical data were retrospectively extracted

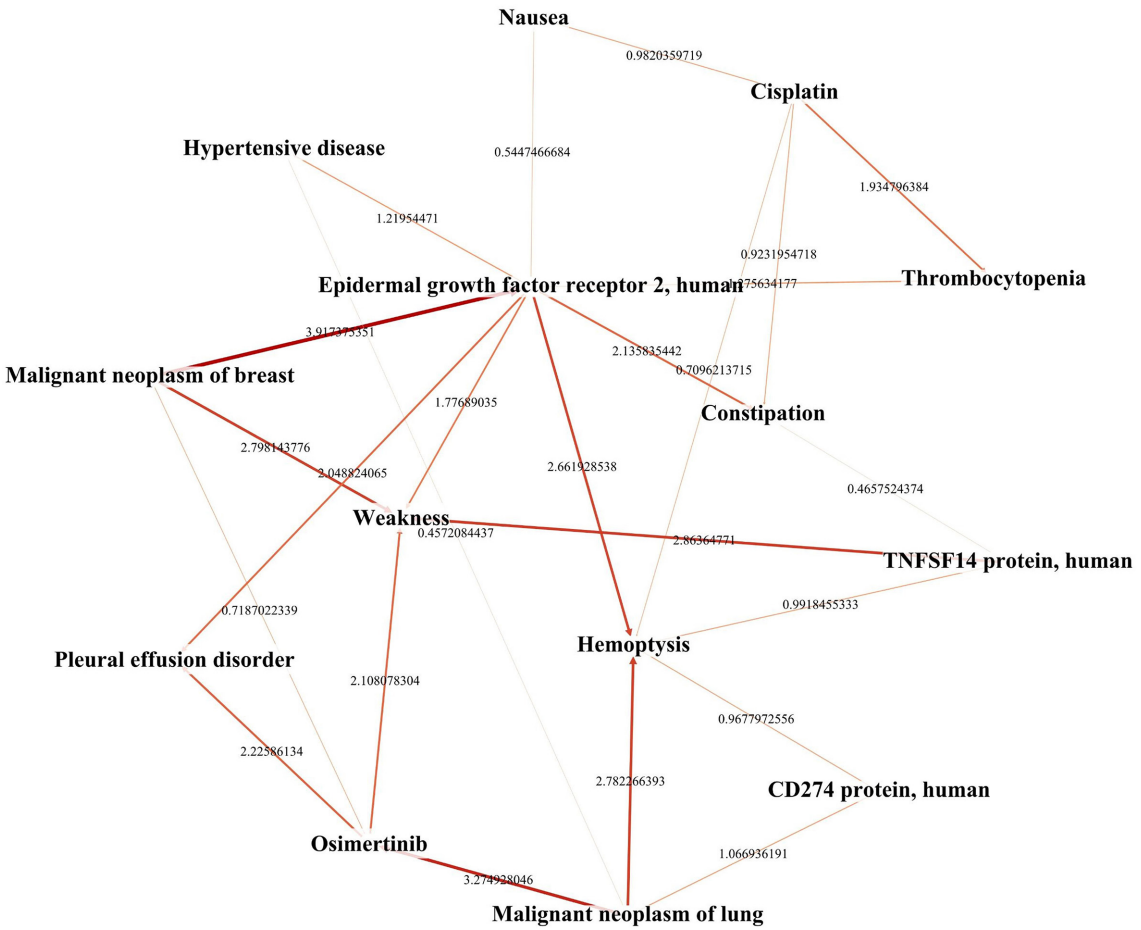

FIGURE 4 | A portion of the network extracted from TBKG.

from the structured hospital information system (HIS) of the 3rd Xiangya Hospital, Central South University (Changsha, China), which provides patient health record information, e.g., information regarding the Enterprise Master Patient Index (EMPI), laboratory tests, International Classification of Disease (ICD-10) clinical diagnosis, medical records and so on. All patients treated with osimertinib in our hospital ($n = 8$) from May 2017 to September 2020 were included in this study. ADRs (Edwards and Aronson, 2000) refer to adverse medical events that occur after a patient receives a drug but that do not necessarily have a causal relationship with the experimental drug. ADRs that meet the definition of Common Terminology Criteria for Adverse Events v4.0 (CTCAE) (US Department of Health Human Services, 2009) include the following: (1) existing exacerbations of chronic or intermittent diseases, including increased frequency and/or increased disease severity; (2) new diseases detected or diagnosed after the administration of the experimental drug, although they may have existed before the study began; (3) signs, symptoms or clinical sequelae due to suspected drug interactions; and (4) signs, symptoms, or clinical sequelae resulting from suspected overdose of an experimental drug or combination of drugs (overdose itself is not reported as an adverse event/serious adverse event). All ADRs of osimertinib during hospitalization were recorded. The Institutional Review Board of the 3rd Xiangya Hospital approved this study (No. 2020-S662).

In addition, the assumption in our study is that drugs have effects on some biomarkers and that these biomarkers are associated with the specific ADRs. And we applied TBKG to discover the adverse reactions of osimertinib and tried to find relative biomarkers that link the drugs with ADRs.

RESULTS

We constructed the TBKG, which is a weighted heterogeneous graph with four types of objects extracted from the MEDLINE corpus: tumor, biomarker, drug, and ADR. Six relationships were built between them (**Table 2**). Then, the naive Bayes model was

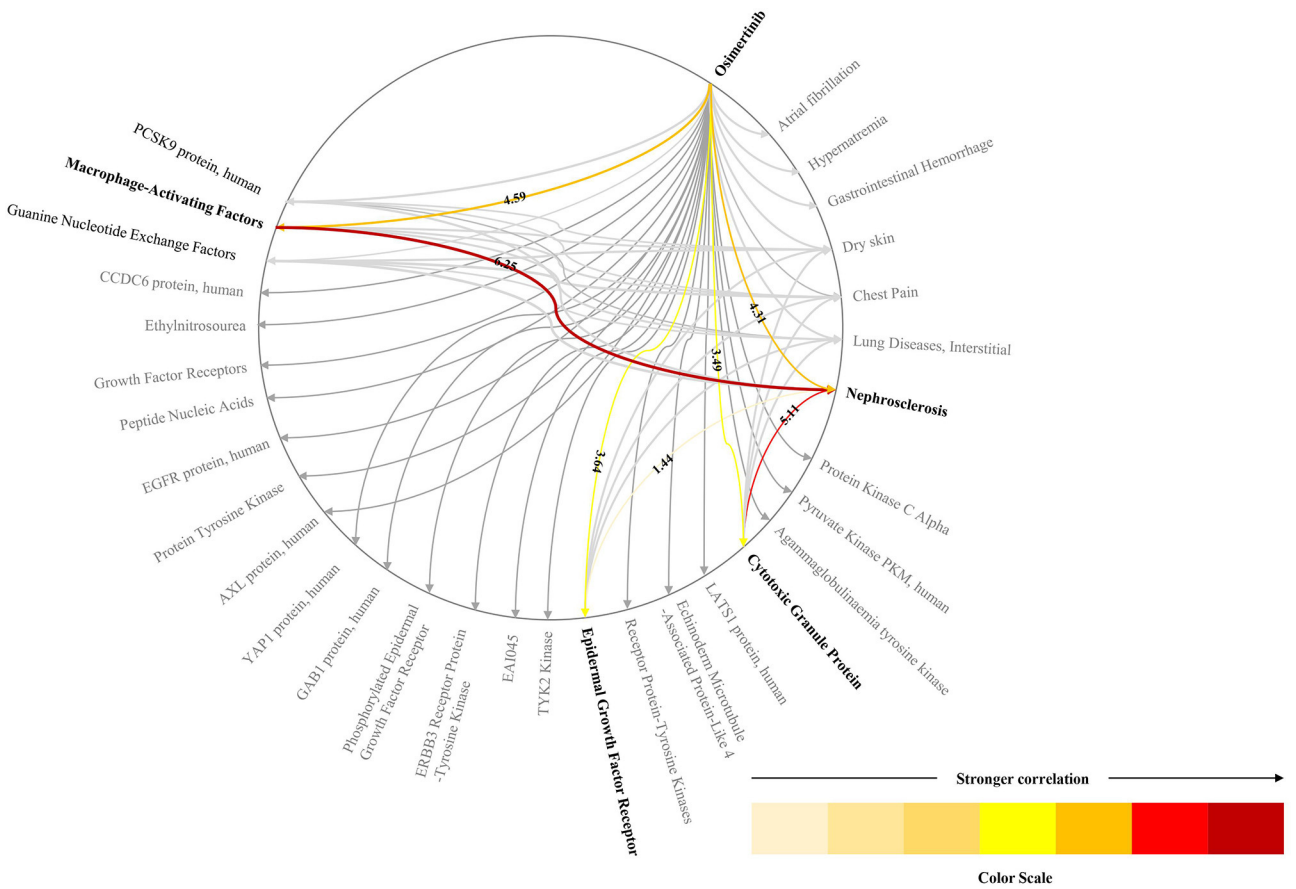

FIGURE 5 | The explainable pathway between osimertinib and one of its ADRs. This figure shows the pathways between osimerinib, top 1% related biomarkers and parts of ADRs. We only highlights the pathways between osimerinib, nephrosclerosis and three biomarkers: "Cytotoxic Granule Protein," "Epidermal Growth Factor Receptor," and "Macrophage-Activating Factors".

used to determine whether relationship exist. We show a part of the TBKG in **Figure 4**.

We use the biomarkers and ADRs related to the drug "osimertinib" as an example to show the TBKG results. The correlation between "osimertinib" and one of its ADRs, "Nephrosclerosis" is 4.31. The correlation result means that in the case of "osimertinib" appearing, the probability of "Nephrosclerosis" is 10.4%, while in the case of "osimertinib" not appearing, the probability of "Nephrosclerosis" is 0.5%. The greater the correlation, the more likely it is that "osimertinib" but no other factors will cause ADRs.

Biomarkers which link the drugs with ADRs were found in TBKG. As shown in **Figure 5**, the correlations between "osimertinib" and the biomarkers "Cytotoxic Granule Protein," "Epidermal Growth Factor Receptor," and "Macrophage-Activating Factors" are 3.49, 3.64, and 4.59, respectively. The corresponding correlations between these three tumor factors and the adverse reaction "Nephrosclerosis" are 5.11, 1.44, and 6.25. Compared with epidermal growth factor receptor and cytotoxic granule protein, macrophage activating factors seems to mediate the incident of osimertinib induced nephrosclerosis.

According to the calculations of our model, 775 ADRs were included in the current study, and the most important ADRs for osimertinib were ordered as follows: dry skin, paronychia inflammation, visual field defects, interstitial lung diseases, and so on. **Supplementary Table 3** lists the calculation results. Our model had moderate consistency with the reports in the official manual (Kappa = 0.68, **Figure 6**), and better than co-occurrence (Kappa = 0.4). And compared with co-occurrence, our model had better specificity than sensitivity.

Furthermore, our model could find rare and serious ADRs that had not been reported in the official manual. Eight lung adenocarcinoma patients had received osimertinib treatment in our hospital since 2017. The characteristics of the included patients are shown in **Table 3**. The mean age of the total population was 61 years, and 50% of the patients were female. The median follow-up time was 6 months. The most common adverse reactions in the clinical cases were lymphocytopenia (3/8), anemia (3/8), and constipation (3/8). From **Figure 7**, it

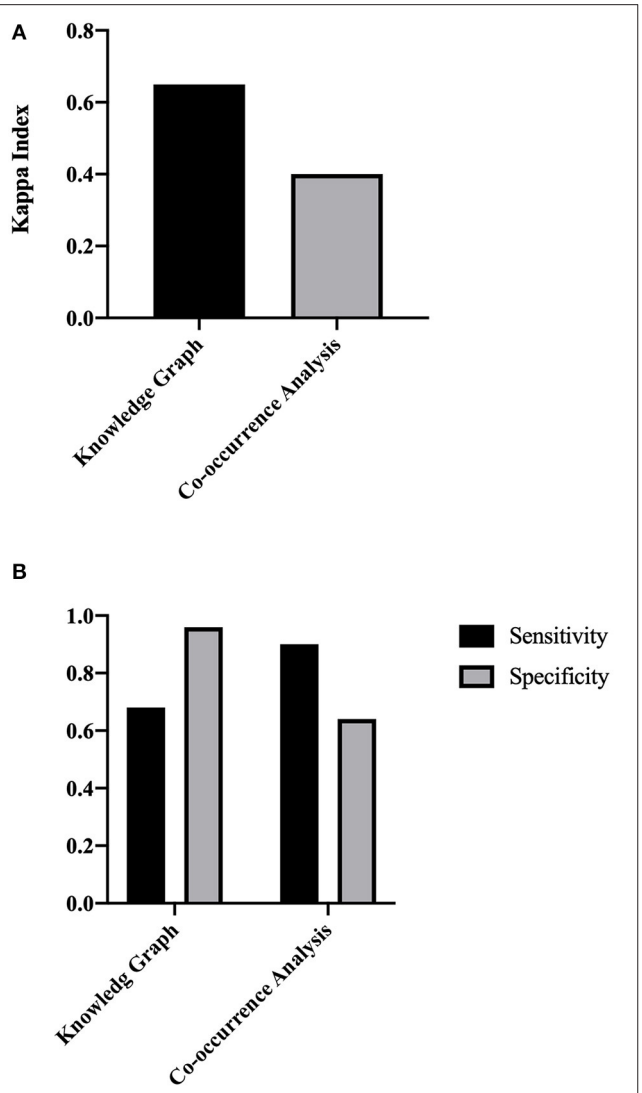

FIGURE 6 | The concordance evaluation of the ADR discovery based on TBKG. **(A)** The Kappa index of TBKG compared with the co-occurrence analysis. **(B)** The sensitivity and specificity of TBKG compared with the co-occurrence analysis.

is worth noting that some serious ADRs that has never been reported before and could be calculated by our model, for example, patient 8 developed renal failure and needed dialysis 1 week after taking osimertinib.

DISCUSSION

Our study opens up a new direction for ADR discovery that combines the following features.

- This is the first knowledge graph-based approach to discover potential adverse reactions of antitumor drugs. By exploring the relations among tumors, biomarkers, and drugs in the knowledge graph, our approach is able to provide explanations for the potential of supervised machine learning methods.
- We contribute a dataset to study knowledge graphs for ADRs by entity extraction and relation building. We verified the efficacy of this approach with clinical data and released the data and the codes that might be valuable to the community working on emerging fields of biomedical literature mining.

Multiple scientific disciplines have been addressing the ADRs discovery problems from different perspectives (Tan et al., 2016). Not only clinical trials before marketing and reports of adverse reactions after marketing, but the detection of metabolic enzyme-related genes has also been used to discover ADRs with the development of pharmacogenomics (Phillips et al., 2001). In the data mining area (Rastegar-Mojarad et al., 2016; Santiso et al., 2018; Shen et al., 2018; Zhang et al., 2019), leveraged the existing information of the drugs and ADRs as the input for a machine learning classifier (e.g., logistic regression, decision tree, and support vector machine), which outputs a binary prediction. Recently, several methods have employed deep learning approaches (Fan et al., 2020) to detect possible ADRs with an effective integration of heterogeneous and multidimensional drug data sources. However, ADR discovery should not be as narrow as a simple true or false question. The reasons behind the ADR in the real world also exist in rich and variant biomedical literature.

Here, we proposed and verified a knowledge graph method based on literature data, that can calculate potential ADRs that never reported before. A knowledge graph is a data model that represents facts as nodes (e.g., drugs, diseases, tumors, side effects, and biomarkers) and relations between the nodes (e.g., drug-biomarker relations). Graph representations are not only able to reveal how individual semantic entities are related to each other but also appealing for human conceptual understanding. This graph structure opens up a new approach to model the abundance of ADR-related information and introduces structural information to determine whether an ADR exists for a tumor drug. It's proved to be feasible that knowledge source extraction and knowledge discovery (Rotmensch et al., 2017; Li et al., 2019; Malas et al., 2019) (e.g., drug prioritization, drug interaction, rare disease classification) by constructing health knowledge graphs. And the experimental results in this study showed that the naive Bayesian model combined building the knowledge graph could outperformed the co-occurrence analysis. Moreover, similar to a few of studies (Guney et al., 2016; Bean et al., 2017), we calculated the potential ADRs by measuring the distance in the graph between the drug, biomarkers, diseases and ADRs (i.e., the drug that cures a disease that are associated with an ADR).The underlying assumption was that a short distance between drugs and ADRs meant that the drug was likely to cause the ADRs.

Our findings not only uncover simple ADRs but also provide explanations, i.e., the paths of "tumor-biomarker-drug" in the knowledge graph. Capturing detailed ADR information may help to obtain in-depth insights into the underlying mechanisms. Previous studies focused on the integration of drug structure or chemical features (Frid and Matthews, 2010; Pauwels et al., 2011)

TABLE 3 | Clinical characteristics of patients who received osimertinib at the 3rd Xiangya Hospital.

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8
Age at osimertinib treatment (years)	73	49	57	71	48	56	81	55
Sex	Male	Female	Male	Female	Male	Male	Female	Female
Diagnosis	Adenocarcinoma of right lung	Bronchial adenocarcinoma	Adenocarcinoma of right lung	Adenocarcinoma of left lung	Adenocarcinoma of right lung	Lung cancer*	Lung cancer*	Lung cancer*
Duration of cancer history	1 month	4 months	12 months	12 months	20 months	1 months	10 months	8 months
Complications	Benign prostatic hyperplasia	/	COPD	Hypertension	CKD; hypertension	/	Postoperative colon cancer; gallstone; remote cerebral infarction	/
Drug combination	Tiotropium bromide, tamsulosin, finasteride	/	Morphine, ampeptide elemente	Tramadol, celecoxib	Ulbenemax, mosapride, calcium malate, montelukast	/	Mecobalamin, trimetazidine, magnesium potassium aspartate, atorvastatin, aspirin	/
Relative gene	Undetected	EGFR (–)	EGFR (+)	EGFR (+)	EGFR (+)	Undetected	EGFR (+)	EGFR (–)
Metastasis site	Adrenal gland; bone; mediastinal lymph nodes; pleura; right subclavian lymph nodes	Lung; pleura	Mediastinal lymph nodes; right subclavian lymph nodes	Bone; liver	Lymph nodes; pericardium; pleura; enterocoelia	Bone; lung	No metastasis	Adrenal gland; bone; brain
Relative gene	Undetected	EGFR (–)	EGFR (+)	EGFR (+)	EGFR (+)	Undetected	EGFR (+)	EGFR (–)
Start date of osimertinib treatment	2016/12/28	2019/1/29	2019/12/10	2019/6/20	2019/10/28	2020/4/15	2020/4/22	2019/10/13
Follow-up time	5 months	6 months	2 months	14 months	7 months	1 months	5 months	7 months

** = Without pathological diagnosis; / = Not mentioned in the electronic health records.*

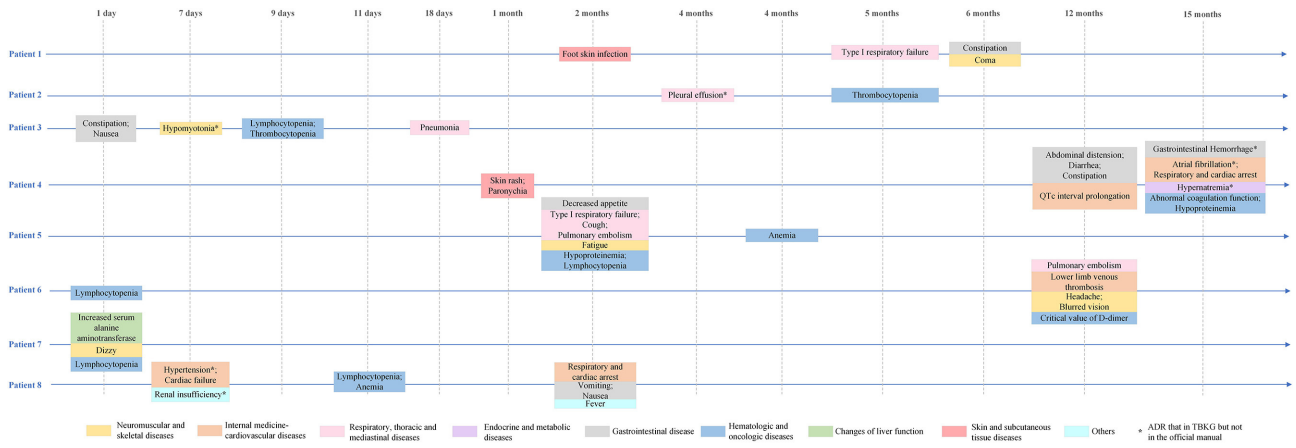

FIGURE 7 | Adverse drug reactions of clinical patients receiving osimertinib. *Not mentioned in the official manual of osimertinib, but mentioned in TBKG.

or gene expression (Wang et al., 2016) or drug target (Párez-Nueno et al., 2015) to optimize ADR prediction models. Risk factors of drug-induced organ damage include drug overdose, drug-drug interactions and drug-related adverse effects, and the discovery of the early biomarkers and development of accurate diagnostic methods are effective prevention strategies for organ damage (Wu and Huang, 2018).

Few studies focus on molecular mechanisms' interpretability in ADR discovery (Hristovski et al., 2016). The spread through relevance in the knowledge graph provides convenience for interpretability. The assumption in our study is that drugs have effects on some biomarkers (proteins, enzymes, and so on) and that these biomarkers are associated with the specific ADRs. Therefore, we try to find biomarkers link the drugs with ADRs in TBKG. The quality of the explanations for the ADRs provided in our approach largely depends on the precision of this knowledge graph, which have been validated whether from the model performance or clinical perspective in this study. Thus, our method can offer better understanding of the biomarkers of ADRs, which could not only significantly predict the potential ADRs before the drug development, but can also provide oncologists with the opportunity to quickly predict patients' sensitivity to the ADRs.

Although important discoveries have been revealed by the current study, there are also limitations. First, the calculated drug-biomarker combinations cannot distinguish between a drug-treatment relationship, a drug-ADR relationship, or a "is not a target drug" type of relationship. However, this is one of major shifts of the big data mindset–a growing emphasis on correlations rather than a continuing quest for elusive causality. In traditional clinical trials, both causal investigations and correlation analysis begin with a hypothesis that is tested to be falsified or verified with little data available. In the age of big data, this type of noncausal analysis will help us understand the "what" rather than the "why." Nonetheless, the construction of relation extraction templates based on the domain knowledge graph (e.g., increasing risk, causing) is encouraging. Second, this study lacks the attention to drug-drug interaction. The use of antitumor drugs often results in the use of other agents to reduce or prevent ADRs and cancer itself increases the need for more medications, which could increase the risk of ADRs. Third, this study focused on ADR discovery based on medical literature. Although compared with the detection of ADRs using clinical data alone, the ADR discovery based on literature have the potential to find the unreported ADR as in our study. For future research, we will improve this knowledge graph-based approach by data fusion and knowledge representation. Finally, these types of recommendations should be assessed by studying the following questions: How many clinicians read them? Are they applied? How effective have they been in reducing the incidence of the complications of hypertension and adverse drug effects? Further prospective clinical trials evaluating the effectiveness of this type of decision support will be explored as the next steps.

CONCLUSION

In summary, the approach described a reliable method for ADR discovery of antitumor drugs and provided explanations of predicted ADRs by exploring the relations among tumors, biomarkers, and drugs in the knowledge graph. This study contributes a dataset to study knowledge graphs for ADRs by entity extraction and relation building; and releases the data and the codes that might be valuable to the community working biomedical literature mining. These findings also provide impetus for the mechanism research of ADRs and therefore offer biomarkers to predict ADRs.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The IRB of Thira Xiangya Hospital, Central South University. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

HT, TL, WO, LG, YT, XH, and WH collected and analysis the clinical data. XL, MW, HW, JS, and XM conceived and designed the experiments, performed the experiments, and wrote the paper. All authors contributed to the conceptual design of the study, the data collection planning and the initial drafting of the manuscript.

FUNDING

The study was supported by the National Natural Science Foundation of China (81901842, 61906037), China Primary Health Care Foundation (YLGX-WS-2020003), CCF-Tencent Open Fund, CCF-Baidu (OF2020003), Scientific research project of Hunan health and Family Planning Commission (20201802) and the National Key R&D Program of China (2018YFC2001800).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.625659/full#supplementary-material

REFERENCES

Anti-Cancer Association (2019). EGFR-TKI ADR management chinese expert consensus. *Chin. J. Lung Cancer* 22:57. doi: 10.3779/j.issn.1009-3419.2019.02.01

Bean, D. M., Wu, H., Iqbal, E., Dzahini, O., Ibrahim, Z. M., Broadbent, M., et al. (2017). Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. *Sci. Rep.* 7, 1–11. doi: 10.1038/s41598-017-16674-x

Califf, R. M. (2018). Biomarker definitions and their applications. *Exp. Biol. Med.* 243, 213–221. doi: 10.1177/1535370217750088

Callon, M., Rip, A., and Law, J. (1986). *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*. Houndmills: Springer.

Carr, D. F., and Pirmohamed, M. (2018). Biomarkers of adverse drug reactions. *Exp. Biol. Med.* 243, 291–299. doi: 10.1177/1535370217733425

Edwards, I. R., and Aronson, J. K. (2000). Adverse drug reactions: definitions, diagnosis, and management. *Lancet* 356, 1255–1259. doi: 10.1016/S.0140-6736(00)02799-9

Fan, B., Fan, W., and Smith, C. (2020). Adverse drug event detection and extraction from open data: a deep learning approach. *Infor. Proces. Manage.* 57:102131. doi: 10.1016/j.ipm.2019.102131

Frid, A. A., and Matthews, E. J. (2010). Prediction of drug-related cardiac adverse effects in humans-B: use of QSAR programs for early detection of drug-induced cardiac toxicities. *Regul. Toxicol. Pharmacol.* 56, 276–289. doi: 10.1016/j.yrtph.2009.11.005

Guney, E., Menche, J., Vidal, M., and Barábasi, A. L. (2016). Network-based in silico drug efficacy screening. *Nat. Commun.* 7, 1–13. doi: 10.1038/ncomms10331

Hristovski, D., Kastrin, A., Dinevski, D., Burgun, A., Žiberna, L., and Rindflesch, T. C. (2016). Using literature-based discovery to explain adverse drug effects. *J. Med. Syst.* 40:185. doi: 10.1007/s10916-016-0544-z

Lee, C. Y., and Chen, Y. P. P. (2020). Prediction of drug adverse events using deep learning in pharmaceutical discovery. *Brief. Bioinform.* bbaa040. doi: 10.1093/bib/bbaa040

Li, X., Wang, Y., Wang, D., Yuan, W., Peng, D., and Mei, Q. (2019). Improving rare disease classification using imperfect knowledge graph. *BMC Med. Inform. Decis. Mak.* 19:238. doi: 10.1186/s12911-019-0938-1

Malas, T. B., Vlietstra, W. J., Kudrin, R., Starikov, S., Charrout, M., Roos, M., et al. (2019). Drug prioritization using the semantic properties of a knowledge graph. *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-42806-6

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. London: MIT press.

Odogwu, L., Mathieu, L., Goldberg, K. B., Blumenthal, G. M., Larkins, E., Fiero, M. H., et al. (2018). FDA benefit-risk assessment of osimertinib for the treatment of metastatic non-small cell lung cancer harboring epidermal growth factor receptor T790M mutation. *Oncologist* 23:353. doi: 10.1634/theoncologist.2017-0425

Párez-Nueno, V. I., Souchet, M., Karaboga, A. S., and Ritchie, D. W. (2015). GESSE: predicting drug side effects from drug-target relationships. *J. Chem. Inform. Model.* 55, 1804–1823. doi: 10.1021/acs.jcim.5b00120

Patton, K., and Borshoff, D. C. (2018). Adverse drug reactions. *Anaesthesia* 73, 76–84. doi: 10.1111/anae.14143

Pauwels, E., Stoven, V., and Yamanishi, Y. (2011). Predicting drug side-effect profiles: a chemical fragment-based approach. *BMC Bioinform.* 12, 1–13. doi: 10.1186/1471-2105-12-169

Phillips, K. A., Veenstra, D. L., Oren, E., Lee, J. K., and Sadee, W. (2001). Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. *JAMA* 286, 2270–2279. doi: 10.1001/jama.286.18.2270

Rastegar-Mojarad, M., Elayavilli, R. K., Wang, L., Prasad, R., and Liu, H. (2016). "Prioritizing adverse drug reaction and drug repositioning candidates generated by literature-based discovery," in *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 289–296.

Rotmensch, M., Halpern, Y., Tlimat, A., Horng, S., and Sontag, D. (2017). Learning a health knowledge graph from electronic medical records. *Sci. Rep.* 7, 1–11. doi: 10.1038/s41598-017-05778-z

Santiso, S., Pérez, A., and, Casillas, A. (2018). Exploring joint ab-lstm with embedded lemmas for adverse drug reaction discovery. *IEEE J. Biomed. Health Inform.* 23, 2148–2155. doi: 10.1109/JBHI.2018.2879744

Shen, Y., Zhang, L., Zhang, J., Yang, M., Tang, B., Li, Y., et al. (2018). CBN: Constructing a clinical Bayesian network based on data from the electronic medical record. *J. Biomed. Inform.* 88, 1–10. doi: 10.1016/j.jbi.2018.10.007

Shrestha, S., Shakya, R., Shrestha, S., and Shakya, S. (2017). Adverse drug reaction due to cancer chemotherapy and its financial burden in different hospitals of Nepal. *Int. J. Pharmacovigil.* 2, 1–7. doi: 10.15226/2476-2431/2/1/00114

Tan, Y., Hu, Y., Liu, X., Yin, Z., Chen, X. W., and Liu, M. (2016). Improving drug safety: from adverse drug reaction knowledge discovery to clinical implementation. *Methods* 110, 14–25. doi: 10.1016/j.ymeth.2016.07.023

US Department of Health and Human Services (2009). *Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0*. National Institutes of Health, National Cancer Institute, 4.

Wang, M., Liu, M., Liu, J., Wang, S., Long, G., and Qian, B. (2017a). Safe medicine recommendation via medical knowledge graph embedding.

Wang, M., Zhang, J., Liu, J., Hu, W., Wang, S., Li, X., et al. (2017b). "Pdd graph: Bridging electronic medical records and biomedical knowledge graphs via entity linking," in *International Semantic Web Conference*. (Cham: Springer), 219–227.

Wang, Z., Clark, N. R., and Ma'ayan, A. (2016). Drug-induced adverse events prediction with the LINCS L1000 data. *Bioinformatics* 32, 2338–2345. doi: 10.1093/bioinformatics/btw168

Wu, H., and Huang, J. (2018). Drug-induced nephrotoxicity: pathogenic mechanisms, biomarkers and prevention strategies. *Curr. Drug Metab.* 19, 559–567. doi: 10.2174/1389200218666171108154419

Zhang, T., Lin, H., Ren, Y., Yang, L., Xu, B., Yang, Z., et al. (2019). Adverse drug reaction detection via a multihop self-attention mechanism. *BMC Bioinform.* 20:479. doi: 10.1186/s12859-019-3053-5

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.

frontiers in Genetics

MINI REVIEW
published: 28 January 2021
doi: 10.3389/fgene.2020.614726

Alpha-Enolase: Emerging Tumor-Associated Antigen, Cancer Biomarker, and Oncotherapeutic Target

*Frankis A. Almaguel[1,2], Tino W. Sanchez[1†], Greisha L. Ortiz-Hernandez[1] and Carlos A. Casiano[1,3]**

[1] Center for Health Disparities and Molecular Medicine, Department of Basic Sciences, Loma Linda University School of Medicine, Loma Linda, CA, United States, [2] Loma Linda University Cancer Center, Loma Linda, CA, United States, [3] Department of Medicine, Division of Rheumatology, Loma Linda University Health, Loma Linda, CA, United States

OPEN ACCESS

Edited by:
Jianying Zhang,
The University of Texas at El Paso,
United States

Reviewed by:
Jitian Li,
Luoyang Orthopedic Traumatological Hospital, China
Liping Dai,
Zhengzhou University, China
Paola Cappello,
University of Turin, Italy

***Correspondence:**
Carlos A. Casiano
ccasiano@llu.edu

†Present address:
Tino W. Sanchez,
National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD, United States

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: 06 October 2020
Accepted: 29 December 2020
Published: 28 January 2021

Citation:
Almaguel FA, Sanchez TW, Ortiz-Hernandez GL and Casiano CA (2021) Alpha-Enolase: Emerging Tumor-Associated Antigen, Cancer Biomarker, and Oncotherapeutic Target. Front. Genet. 11:614726. doi: 10.3389/fgene.2020.614726

Alpha-enolase, also known as enolase-1 (ENO1), is a glycolytic enzyme that "moonlights" as a plasminogen receptor in the cell surface, particularly in tumors, contributing to cancer cell proliferation, migration, invasion, and metastasis. ENO1 also promotes other oncogenic events, including protein-protein interactions that regulate glycolysis, activation of signaling pathways, and resistance to chemotherapy. ENO1 overexpression has been established in a broad range of human cancers and is often associated with poor prognosis. This increased expression is usually accompanied by the generation of anti-ENO1 autoantibodies in some cancer patients, making this protein a tumor associated antigen. These autoantibodies are common in patients with cancer associated retinopathy, where they exert pathogenic effects, and may be triggered by immunodominant peptides within the ENO1 sequence or by posttranslational modifications. ENO1 overexpression in multiple cancer types, localization in the tumor cell surface, and demonstrated targetability make this protein a promising cancer biomarker and therapeutic target. This mini-review summarizes our current knowledge of ENO1 functions in cancer and its growing potential as a cancer biomarker and guide for the development of novel anti-tumor treatments.

Keywords: alpha-enolase, autoantibodies, cancer biomarker, therapeutic target, ENO1

INTRODUCTION

Alpha-enolase (ENO1, 47 kD) has recently emerged as a major driver of tumor metabolism and progression and is considered a rising cancer biomarker and therapeutic target (Capello et al., 2011; Hsiao et al., 2013; Principe et al., 2017; Cappello et al., 2018). ENO1 is one of three enolase isoforms encoded by different genes: ENO1, expressed in most human tissues and upregulated in cancer cells; gamma-enolase (ENO2), expressed in neuronal cells and neuroendocrine differentiated tumors; and beta-enolase (ENO3), expressed in muscles (Pancholi, 2001; Isgrò et al., 2015; Ji et al., 2016). These isoforms show high sequence conservation and similar size, and combine to catalyze the dehydration of 2-phosphoglycerate to phosphoenolpyruvate during glycolysis. In cancer cells, this reaction occurs under both aerobic and anaerobic glycolysis, contributing to the Warburg Effect,

which increases glucose uptake, proliferation, and tumor growth (Pancholi, 2001; Liberti and Locasale, 2016; Qian et al., 2017).

Alpha-enolase is overexpressed in multiple human cancer types, contributing to increased glycolysis and tumor growth (Altenberg and Greulich, 2004; Chang et al., 2006; He et al., 2007; Tsai et al., 2010; Capello et al., 2011; Song et al., 2014; Fu et al., 2015; Sun et al., 2017, 2019; Zhan et al., 2017; Yin et al., 2018; Zhang et al., 2018, 2020; Cheng et al., 2019; Ji et al., 2019; Qiao et al., 2019; Xu et al., 2019; Chen et al., 2020). ENO1 overexpression is often associated with anti-ENO1 autoantibody responses and may have prognostic and diagnostic value in certain cancers (**Table 1**; Adamus et al., 1998; Tomaino et al., 2011; Pranay et al., 2013; Hsiao et al., 2015; Griggio et al., 2017; Zhang et al., 2020). ENO1 is also localized on the surface of cancer cells where it enhances plasmin formation (Miles et al., 1991; Redlitz et al., 1995) to promote extracellular matrix degradation, cell migration, invasion, and metastasis (Hsiao et al., 2013; Didiasova et al., 2014; Principe et al., 2015, 2017; Zakrzewicz et al., 2018). These properties make ENO1 a tumor-associated antigen (TAA) and promising cancer biomarker and therapeutic target. Below we summarize ENO1's functions in cancer, growing potential as a cancer biomarker, and rising opportunities for targeting this enzyme for cancer treatment.

MULTIFUNCTIONAL ONCOPROTEIN

Alpha-enolase mRNA gives rise to an alternative translation product of 37 kD called *c-MYC* promoter binding protein 1 (MBP1) (**Figure 1A**; Subramanian and Miller, 2000). Although MBP1 does not have glycolytic activity, it regulates the cellular response to altered glucose concentration (Sedoris et al., 2007). ENO1 is upregulated by the c-MYC oncoprotein (Osthus et al., 2000), and is localized in the cytoplasm and the cell surface, playing multiple roles (**Figure 1B**; Diaz-Ramos et al., 2012; Didiasova et al., 2019). In contrast, MBP1 is a nuclear protein that represses c-MYC transcription under cellular stress and low glucose conditions, leading to decreased cell proliferation (Feo et al., 2000; Subramanian and Miller, 2000; Sedoris et al., 2007; Maranto et al., 2015). The ratio of ENO1/MBP1 expression in cancer cells is regulated by glucose, with c-MYC-driven elevated ENO1 expression under high glucose conditions, and elevated MBP1 expression under low glucose conditions (Sedoris et al., 2007). Cancer cells adapt to hypoxia by overexpressing c-MYC, which stimulates glycolysis and cell proliferation *via* ENO1 upregulation and MBP-1 downregulation (Sedoris et al., 2010). The ENO1/MBP-1 ratio influences cancer aggressiveness, as demonstrated in human breast tumors where overexpression of ENO1 and extracellular matrix metalloproteinases MMP-2 and MMP-9, concomitant with MBP-1 downregulation, correlates with worse prognosis (Cancemi et al., 2019).

Alpha-enolase also plays important roles as a plasminogen receptor, component of exosomal vesicles, cytoskeleton reorganizing protein, stabilizer of mitochondrial membrane, and modulator of oncogenic signaling pathways (**Figure 1B**; Diaz-Ramos et al., 2012; Didiasova et al., 2019). These functions allow overexpressed ENO1 to promote cancer cell proliferation, survival, clonogenicity, epithelial-mesenchymal transition (EMT), chemoresistance, extracellular matrix degradation, migration, invasion, and metastasis. These functions can be inhibited in cancer cells by ENO1 depletion (Georges et al., 2011; Song et al., 2014; Fu et al., 2015; Zhu et al., 2015; Capello et al., 2016; Principe et al., 2017; Qian et al., 2017; Zhan et al., 2017; Qiao et al., 2018a, 2019; Ji et al., 2019; Sun et al., 2019; Wang et al., 2019; Xu et al., 2019; Santana-Rivera et al., 2020), or targeting with antibodies (Hsiao et al., 2013; Principe et al., 2015), microRNA (miR) (Liu et al., 2018), or long non-coding RNAs (lncRNAs) (Yu et al., 2018). ENO1 also regulates oncogenic signaling pathways, including PI3K/Akt (Fu et al., 2015; Sun et al., 2019; Chen et al., 2020; Zang et al., 2020), v/β-3 integrin (Principe et al., 2017), β-catenin (Ji et al., 2019), transforming growth factor beta (Xu et al., 2019), AMPK/mTOR (Zhan et al., 2017; Dai et al., 2018), and others (Huang et al., 2019).

Acting as a plasminogen receptor, ENO1 "moonlights" on the surface of tumor cells to facilitate plasminogen conversion into plasmin (Miles et al., 1991; Redlitz et al., 1995; Capello et al., 2011; Diaz-Ramos et al., 2012; Ceruti et al., 2013; Hsiao et al., 2013; Didiasova et al., 2014, 2019). During inflammatory conditions, plasmin activation leads to fibrinolysis and facilitates extracellular matrix degradation, a function linked to ENO1's ability to promote cancer cell migration, invasion, and metastasis (Hsiao et al., 2013; Kumari and Malla, 2015). Bacteria and immune cells take advantage of ENO1's plasminogen receptor functions to facilitate tissue invasion (Wygrecka et al., 2009; Bergmann et al., 2013).

The plasminogen binding activity of ENO1 has been mapped to the C-terminal peptide $_{422}$KFAGRNFRNPLAK$_{434}$, Miles et al. (1991) and Redlitz et al. (1995), with another putative plasminogen binding site located at $_{250}$FFRSGK$_{256}$ (**Figure 1A**; Kang et al., 2008). ENO1 surface localization is guided by post-translational modifications (PTMs), particularly methylation of arginine 50 (Zakrzewicz et al., 2018). Other PTMs, including citrullination (Lundberg et al., 2008), acetylation, and phosphorylation (Zhou et al., 2010; Capello et al., 2011; Tomaino et al., 2011; Sanchez et al., 2016), are also likely to influence ENO1 functions, localization, and immunogenicity (Didiasova et al., 2019). ENO1 exteriorization is promoted by lipopolysaccharide (Zakrzewicz et al., 2018), calcium influx (Didiasova et al., 2015), and interaction with caveolin 1, annexin 2, and heat shock protein 70 (Zakrzewicz et al., 2014; Perconti et al., 2017).

Alpha-enolase interacts in the cell surface with B7-H3, an immune co-stimulatory molecule with oncoprotein functions, to promote glycolysis (Zuo et al., 2018). It also interacts with granulin A (GRN-A), a 6 kDa peptide derived from progranulin that inhibits ENO1's ability to promote cancer cell proliferation, migration, and invasion (Chen et al., 2017). GRN-A synergizes with cisplatin to induce apoptosis in hepatocellular carcinoma cells (Qiao et al., 2018b). Overexpressed ENO1 promotes resistance to cisplatin and other anti-tumor drugs in cancer cells by increasing glycolysis and cell proliferation (Tu et al., 2010; Qian et al., 2017; Qiao et al., 2018a; Wang et al., 2019; Santana-Rivera et al., 2020), interaction with microtubules (Georges et al., 2011), and cell adhesion (Zhu et al., 2015; Principe et al., 2017).

TABLE 1 | Potential prognostic and diagnostic value of ENO1 expression in tumors and cancer-associated anti-ENO1 autoantibodies.

Cancer type	Molecule	Prognostic/diagnostic value	References
Bladder	ENO1	Prognostic	Ji et al., 2019
Breast	ENO1	Prognostic	Tu et al., 2010; Cancemi et al., 2019
Cancer-associated retinopathy	Autoantibodies	Prognostic (progressive blinding)	Adamus et al., 1998
Chronic lymphocytic leukemia	Autoantibodies	Prognostic	Griggio et al., 2017
Colorectal	ENO1	Prognostic	Zhan et al., 2017
Gastric cancer	ENO1	Prognostic	Qian et al., 2017; Qiao et al., 2019; Sun et al., 2019; Xu et al., 2019
Glioma	ENO1	Prognostic	Song et al., 2014
Head and Neck	Both ENO1 and autoantibodies	Prognostic	Tsai et al., 2010; Pranay et al., 2013
Liver	Both ENO1 and autoantibodies	Prognostic/diagnostic	Takashima et al., 2005; Hamaguchi et al., 2008; Zhang et al., 2020
Lung Cancer	Both ENO1 and autoantibodies	Prognostic/Diagnostic	Chang et al., 2006; He et al., 2007; Shih et al., 2010; Hsiao et al., 2015; Dai et al., 2017; Zhang et al., 2018; Zang et al., 2019
Multiple myeloma	ENO1	Prognostic	Ray et al., 2020
Non-Hodgkin's Lymphoma	ENO1	Prognostic	Zhu et al., 2015
Pancreatic cancer	Both ENO1 and autoantibodies	Prognostic	Tomaino et al., 2011; Sun et al., 2017; Yin et al., 2018; Wang et al., 2019

Alpha-enolase has also been implicated in the regulation of T cell effector functions, including the suppressive functions of induced regulatory T cells (De Rosa et al., 2015), T cell activation La Rocca et al. (2017), and the diabetogenic functions of islet-specific CD4+ T cells (Berry et al., 2015). Gemta et al. (2019) recently reported that downregulation of ENO1 activity represses the glycolytic activity of tumor infiltrating CD8+ lymphocytes (CD8+ TILs), leading to their functional exhaustion. This impaired ENO1 function is unrelated to its expression, suggesting the involvement of post-transcriptional regulatory mechanisms such as PTMs influencing ENO1 enzymatic activity or subcellular localization (Gemta et al., 2019).

TUMOR-ASSOCIATED ANTIGEN

The presence of anti-ENO1 autoantibodies is well documented in autoimmune diseases such as rheumatoid arthritis (RA) and autoimmune retinopathy (Adamus, 2017). In RA, these autoantibodies recognize an immunodominant citrullinated peptide within the ENO1 N-terminus (Lundberg et al., 2008), and are clinical diagnostic biomarkers.

Alpha-enolase autoantibodies are also present in cancer patients, often associated with cancer-associated retinopathy (CAR) (Adamus, 2017). Unlike in RA, ENO1 autoantibodies from CAR patients do not specifically target citrullinated peptides but rather recognize several epitopes, including an immunodominant N-terminal domain peptide, $_{56}$RYMGKGVS$_{63}$, and a C-terminal peptide implicated in plasminogen binding, $_{421}$AKFAGRNF$_{428}$ (Adamus et al., 1998). CAR-linked ENO1 autoantibodies promote retinopathy by inducing retinal cell apoptosis, leading to retinal dysfunction or degeneration (Adamus, 2018; Adamus et al., 2020). *In vitro* treatment of retinal cells with an anti-ENO1 monoclonal antibody significantly impaired glycolysis, reduced ATP production, and induced apoptosis (Magrys et al., 2007). ENO1 autoantibodies from patients with autoimmune retinopathy also target retinal ganglion cells and induce apoptosis in rats (Ren and Adamus, 2004). Further, the survival of retinal cells treated with ENO1 autoantibodies from patients with autoimmune retinopathy and CAR was impaired compared to retinal cells exposed to sera from healthy controls (Adamus et al., 1998). While ENO1 autoantibodies are known to trigger pathological effects through their internalization by retinal cells (Ren and Adamus, 2004), it cannot be ruled out that they also directly target cell surface ENO1, leading to glycolysis impairment and apoptosis. The study of ENO1 autoantibodies in CAR has uncovered a potential unintended consequence - i.e., antibody-induced retinal apoptosis- that requires careful consideration as ENO1-based cancer immunotherapies are developed.

Alpha-enolase autoantibodies are associated with either improved or poor tumor patient outcomes in different cancer types, suggesting a context-dependent clinical significance (**Table 1**; Shih et al., 2010; Tomaino et al., 2011; Pranay et al., 2013; Hsiao et al., 2015; Griggio et al., 2017). While these autoantibodies may occur in cancer patients independent of CAR, a recent study showed that vision loss and anti-retinal autoantibodies occur in at least 20 different human cancers, with ENO1 being the most frequent target of these antibodies (Adamus et al., 2020). Autoantibodies to other glycolytic enzymes have also been detected in CAR patients (Adamus et al., 2020), suggesting that they are induced by immune presentation of peptides from overexpressed metabolic proteins released from tumor cells (Adamus et al., 2020).

ENO1 autoantibodies have been included in TAA panels for cancer immunodiagnosis. For instance, Zang et al. (2019) examined a panel of four cancer biomarkers (carcinoembryonic antigen, cancer antigen 125, Annexin A1 autoantibodies, and ENO1 autoantibodies) for lung cancer detection that yielded high specificity, sensitivity, and diagnostic accuracy. Dai et al. (2017)

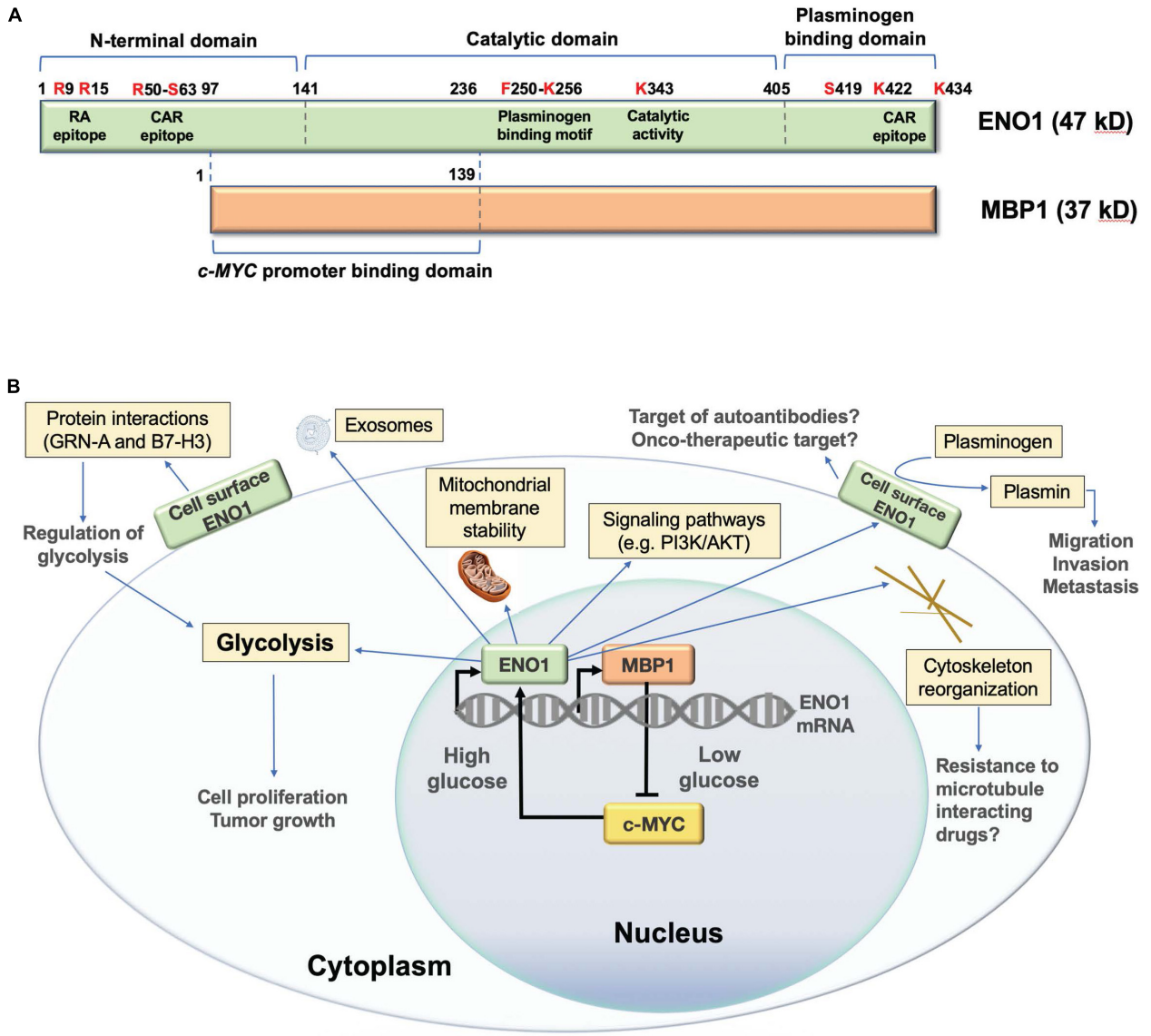

FIGURE 1 | Structure and functions of human ENO1. **(A)** Schematic representation of domain structure of ENO1 and its alternative translation variant c-MYC promoter binding protein 1 (MBP1). Several lysine residues (K256, K422, and K434) have been implicated in the plasminogen binding functions of ENO1, whereas K343 has been implicated in its catalytic activity, required for the conversion of 2-phosphoglycerate to phosphoenolpyruvate during glycolysis. Citrullination of arginines 9 (R9) and 15 (R15) generates an immunodominant peptide (residues 5–22) that is targeted by autoantibodies in patients with rheumatoid arthritis. Methylation of arginine 50 (R50) has been implicated in ENO1 externalization. R50 is also part of an immunodominant epitope recognized by autoantibodies from patients with cancer associated retinopathy (CAR). Another CAR epitope is located within the plasminogen binding domain. Phosphorylated serine 419 (S419) within the plasminogen binding domain is recognized of ENO1 autoantibodies in pancreatic cancer patients. While MBP1 shares the catalytic and plasminogen binding domains of ENO1, it lacks these functions due to its exclusive nuclear localization. MPB1 residues 1–139 (ENO1 96–236) comprise its DNA binding domain, required for binding to the *c-MYC* gene promoter, which results in repression of promoter activity and downregulation of c-MYC protein expression. **(B)** Schematic representation of the cellular functions of ENO1 and MBP1. MBP1 localizes primarily in the cell nucleus where it represses the *c-MYC* gene promoter, whose activity is essential for ENO1 upregulation. ENO1 is primarily localized in the cytoplasm, where it functions in glycolysis, promoting mitochondrial stability and cytoskeleton reorganization, and regulating oncogenic signaling pathways. This protein is also localized on the cell surface, where it acts as a plasminogen receptor and interacting partner of various proteins to regulate glycolysis, as well as cancer cell migration, invasion, and metastasis. ENO1 can also be secreted from cells as a component of exosomal vesicles.

also reported that combining ENO1 autoantibodies with carcinoembryonic antigen and cytokeratin 19 fragment in a diagnostic panel increased diagnostic sensitivity for non-small cell lung cancer. Another study detected ENO1 autoantibodies at higher frequencies in patients with early stage lung cancer compared to late stage patients (Zhang et al., 2018).

Post-translational modifications contribute to the generation of ENO1 autoantibodies, as evidenced by the observation

that patients with pancreatic ductal adenocarcinoma (PDA) produce antibodies that specifically target epitopes containing phosphorylated serine 419 within the plasminogen binding domain of ENO1 (**Figure 1**), and correlate with improved outcome in patients receiving chemotherapy, suggesting a protective role (Tomaino et al., 2011). It is not clear if, like in RA, citrullination triggers ENO1 autoantibodies in cancer patients, although citrullinated ENO1 was reported to elicit anti-tumor CD4+ T responses in murine tumor xenografts and in ovarian cancer patients (Cook et al., 2018; Brentville et al., 2020). Our group and others identified citrullinated ENO1 in cancer cells (Jiang et al., 2013; Sanchez et al., 2016), suggesting that this PTM could potentially trigger ENO1 autoantibodies in cancer patients.

We reported that ENO1 autoantibodies occur at higher frequency in prostate cancer (PCa) patients compared to controls, showing racial differences in reactivity (Sanchez et al., 2016). While autoantibodies from European American (EA) PCa patients reacted strongly with human recombinant ENO1 by ELISA but weakly by immunoblotting against endogenous ENO1 from PCa cells, autoantibodies from African American (AA) patients showed the opposite pattern. ENO1 autoantibodies from AA patients also displayed differential reactivity against endogenous ENO1 in a panel of PCa cell lines, reacting strongly with ENO1 in metastatic PCa cell lines by immunoblotting, whereas autoantibodies from EA patients reacted uniformly against this protein across the panel. Intriguingly, ENO1 autoantibodies from AA patients lost immunoreactivity in docetaxel-resistant cells, while autoantibodies from EA patients retained this reactivity. Proteomics analysis revealed differences in PTMs (e.g., acetylation, methylation, phosphorylation, and citrullination) within endogenous ENO1 between chemosensitive and chemoresistant PCa cells, suggesting that the observed racial differences in ENO1 autoantibody reactivity in these cell types might be influenced by PTMs.

In addition to ovarian cancer (Brentville et al., 2020), T cell responses targeting ENO1 have also been reported in patients with PDA. ENO1-specific CD8+ T cell responses were detected in 8 out of 12 PDA patients with circulating anti-ENO1 IgG autoantibodies, whereas patients without these autoantibodies lacked these responses, suggesting an integrated humoral and cellular anti-ENO1 response (Cappello et al., 2009). A later study reported that phosphorylated ENO1 also triggers CD4+ T cell responses in PDA patients (Capello et al., 2015).

CANCER BIOMARKER AND THERAPEUTIC TARGET

The need for new cancer-specific targets that can act as beacons to localize tumors with high efficiency is a key feature of a robust biomarker. As mentioned above, growing evidence suggests that ENO1 is upregulated in a broad range of human tumors, making it a candidate cancer biomarker. ENO1 localization on the surface of cancer cells also provides an excellent opportunity to develop small molecules with high affinity to this protein, which enables its direct targeting in the tumor surface for diagnostic imaging and therapeutics.

The diagnostic and prognostic value of ENO1 overexpression has been confirmed in several tumors (**Table 1**). For example, in breast cancer, enhanced ENO1 expression correlated with greater tumor size, poor nodal status, and a shorter disease-free interval (Tu et al., 2010). Patients with lung cancer overexpressing ENO1 also showed poor clinical outcomes, with shorter overall and progression-free survival, compared to low expressing patients (Chang et al., 2006; Hsiao et al., 2013). ENO1 overexpression in hepatocellular carcinoma increased with tumor de-differentiation and correlated positively with venous invasion (Takashima et al., 2005; Hamaguchi et al., 2008). These characteristics position ENO1 as a selective biomarker able to identify aggressive tumor types with high accuracy.

Alpha-enolase has several key characteristics of an ideal cancer biomarker: (1) localization in the cell surface where it can be targeted for imaging and treatment; (2) overexpression in cancer cells with low expression in normal tissues; and (3) overexpression correlating with prognosis and clinical outcomes. Thus, ENO1 can be envisioned as an excellent biomarker to guide patient management and alter disease timeline. Ultimately, ENO1 surface imaging could potentially be used to screen for occult cancers. This information could then be translated to improve prognosis and management of patients diagnosed with cancer by monitoring disease state, detecting recurrence and progression, or assessing response to therapy.

Alpha-enolase has a potent three punch combination to advance cancer progression: (1) promotes tumor glycolysis, (2) activates cancer signaling pathways, and (3) drives tumor migration, invasion, and metastasis. These unique characteristics make ENO1 a strong candidate to deliver targeted therapies to tumors overexpressing this protein, particularly those tumors expressing surface ENO1. For instance, molecular imaging of tumors guided by ENO1-specific small molecule probes could open the door to new strategies to target this protein in tumors, leading to early interventions and improved patient outcomes. Several reports have already provided pre-clinical data supporting ENO1 therapeutic targeting. As mentioned above, ENO1 depletion attenuates glycolysis, cell proliferation, EMT, migration, and invasion, and metastasis in several cancer types (Fu et al., 2015; Capello et al., 2016; Principe et al., 2017; Zhan et al., 2017; Ji et al., 2019; Sun et al., 2019). Targeting of ENO1 in combination with chemotherapy may be beneficial in patients with drug resistant cancers given, as mentioned earlier, its emerging role in chemoresistance.

There is a major need for small molecule inhibitors of ENO1. A promising inhibitor, ENOblock, has been used to target ENO1 in various disease contexts (Jung et al., 2013; Cho et al., 2017, 2019; Haque et al., 2017; Polcyn et al., 2020) but its specificity was disputed (Satani et al., 2016). Another ENO1 inhibitor was recently reported to enhance anti-multiple myeloma (MM) immunity in combination with immunotherapy in pre-clinical models (Ray et al., 2020). In addition, a novel nanoparticle-delivered peptide targeting ENO1 in combination with doxorubicin demonstrated strong antitumor activity in pre-clinical models of PCa (Wang et al., 2018).

The study of immune responses to ENO1 has sparked the development of novel immunotherapeutic strategies.

For instance, treatment of lung cancer cells with anti-ENO1 monoclonal antibodies *in vitro* suppressed cell-associated plasminogen and matrix metalloproteinase activation, collagen and gelatin degradation, and cell invasion (Hsiao et al., 2013). Interestingly, adoptive transfer of these antibodies to mice resulted in their accumulation in subcutaneous tumors and inhibition of lung and bone metastases. Principe et al. (2015) reported that *in vitro* and *in vivo* blockade of ENO1 with anti-human ENO1 monoclonal antibodies reduced PDA cell migration and invasion. Further, administration of adeno-associated virus (AAV) encoding an anti-ENO1 monoclonal antibody led to a reduction of lung metastasis in mouse PDA xenografts (Principe et al., 2015). The same group developed an ENO1 DNA vaccine that significantly inhibited, although did not eradicate, tumor growth in a mouse PDA model, suggesting that the effectiveness of this vaccine could be amplified in the context of combinatorial therapies (Cappello et al., 2013, 2018). Recently, Mandili et al. (2020) demonstrated that treatment of PDA mice with combined gemcitabine chemotherapy and ENO1 DNA vaccination induced a strong CD4+ T cell antitumor activity that impaired tumor progression, compared with mice that received vaccine or gemcitabine alone.

CONCLUSION

Alpha-enolase promotes cellular functions associated with tumor aggressiveness, including increased glycolysis, activation of oncogenic signaling pathways, chemoresistance, and cell proliferation, migration, invasion, and metastasis. Therefore, ENO1 can be considered an oncoprotein critical for maintaining several "hallmarks of cancer" (Hanahan and Weinberg, 2011), particularly sustained proliferative signaling, deregulated energy metabolism, apoptosis resistance, and activation of invasion and metastasis. ENO1 overexpression in a broad range of human cancers and targetability make it an attractive cancer biomarker candidate and therapeutic target. Its localization in the tumor surface, key metabolic functions, and ability to promote tumor aggressive properties could be exploited for the development of novel comprehensive cancer care modalities that combine ENO1 surface imaging with targeted therapeutic interventions.

AUTHOR CONTRIBUTIONS

GLO-H designed and prepared the figures and crosschecked the references for accuracy. CAC designed the overall organization of the manuscript and approved the final version. All authors contributed to the literature review and the writing and final editing of the manuscript.

FUNDING

Research on ENO1 and other tumor-associated antigens in the authors' laboratories was funded in part by National Institutes of Health (NIH) grants P20MD006988-Project 2 and R21CA226654-01A1 to CAC, Loma Linda University (LLU) Center for Health Disparities and Molecular Medicine, LLU School of Medicine, and LLU Cancer Center. GLO-H was supported by NIH grant R25GM060507 and the LLU-NIH Initiative for Maximizing Student Development (IMSD) graduate training program.

REFERENCES

Adamus, G. (2017). Impact of autoantibodies against glycolytic enzymes on pathogenicity of autoimmune retinopathy and other autoimmune disorders. *Front. Immunol.* 8:505. doi: 10.3389/fimmu.2017.00505

Adamus, G. (2018). Are anti-retinal autoantibodies a cause or a consequence of retinal degeneration in autoimmune retinopathies? *Front. Immunol.* 9:765. doi: 10.3389/fimmu.2018.00765

Adamus, G., Amundson, D., Seigel, G. M., and Machnicki, M. (1998). Anti-enolase-alpha autoantibodies in cancer-associated retinopathy: epitope mapping and cytotoxicity on retinal cells. *J. Autoimmun.* 11, 671–677. doi: 10.1006/jaut.1998.0239

Adamus, G., Champaigne, R., and Yang, S. (2020). Occurrence of major anti-retinal autoantibodies associated with paraneoplastic autoimmune retinopathy. *Clin. Immunol.* 210:108317. doi: 10.1016/j.clim.2019.108317

Altenberg, B., and Greulich, K. O. (2004). Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. *Genomics* 2004, 1014–1020. doi: 10.1016/j.ygeno.2004.08.010

Bergmann, S., Schoenen, H., and Hammerschmidt, S. (2013). The interaction between bacterial enolase and plasminogen promotes adherence of *Streptococcus pneumoniae* to epithelial and endothelial cells. *Int. J. Med. Microbiol.* 303, 452–462. doi: 10.1016/j.ijmm.2013.06.002

Berry, G. J., Frielle, C., Luu, T., Salzberg, A. C., Rainbow, D. B., Wicker, L. S., et al. (2015). Genome-wide transcriptional analyses of islet-specific CD4+ T cells identify Idd9 genes controlling diabetogenic T cell function. *J. Immunol.* 194, 2654–2663. doi: 10.4049/jimmunol.1401288

Brentville, V. A., Metheringham, R. L., Daniels, I., Atabani, S., Symonds, P., Cook, K. W., et al. (2020). Combination vaccine based on citrullinated vimentin and enolase peptides induces potent CD4-mediated anti-tumor responses. *J. Immunother. Cancer.* 8, e000560. doi: 10.1136/jitc-2020-000560

Cancemi, P., Buttacavoli, M., Roz, E., and Feo, S. (2019). Expression of alpha-Enolase (ENO1), Myc promoter-binding protein-1 (MBP-1) and matrix metalloproteinases (MMP-2 and MMP-9) reflect the nature and aggressiveness of breast tumors. *Int. J. Mol. Sci.* 20, 3952. doi: 10.3390/ijms20163952

Capello, M., Caorsi, C., Bogantes Hernandez, P. J., Dametto, E., Bertinetto, F. E., Magistroni, P., et al. (2015). Phosphorylated alpha-enolase induces autoantibodies in HLA-DR8 pancreatic cancer patients and triggers HLA-DR8 restricted T-cell activation. *Immunol. Lett.* 167, 11–16. doi: 10.1016/j.imlet.2015.06.008

Capello, M., Ferri-Borgogno, S., Cappello, P., and Novelli, F. (2011). α-Enolase: a promising therapeutic and diagnostic tumor target. *FEBS J.* 278, 1064–1074. doi: 10.1111/j.1742-4658.2011.08025.x

Capello, M., Ferri-Borgogno, S., Riganti, C., Chattaragada, M. S., Principe, M., Roux, C., et al. (2016). Targeting the Warburg effect in cancer cells through ENO1 knockdown rescues oxidative phosphorylation and induces growth arrest. *Oncotarget* 7, 5598–5612. doi: 10.18632/oncotarget.6798

Cappello, P., Curcio, C., Mandili, G., Roux, C., Bulfamante, S., and Novelli, F. (2018). Next generation immunotherapy for pancreatic cancer: DNA vaccination is seeking new combo partners. *Cancers* 10, 51. doi: 10.3390/cancers10020051

Cappello, P., Rolla, S., Chiarle, R., Principe, M., Cavallo, F., Perconti, G., et al. (2013). Vaccination with ENO1 DNA prolongs survival of genetically engineered mice with pancreatic cancer. *Gastroenterology* 144, 1098–1106. doi: 10.1053/j.gastro.2013.01.020

Cappello, P., Tomaino, B., Chiarle, R., Ceruti, P., Novarino, A., Castagnoli, C., et al. (2009). An integrated humoral and cellular response is elicited in pancreatic

cancer by alpha-enolase, a novel pancreatic ductal adenocarcinoma-associated antigen. *Int. J. Cancer* 125, 639–648. doi: 10.1002/ijc.24355

Ceruti, P., Principe, M., Capello, M., Cappello, P., and Novelli, F. (2013). Three are better than one: plasminogen receptors as cancer theranostic targets. *Exp. Hematol. Oncol.* 2, 12. doi: 10.1186/2162-3619-2-12

Chang, G. C., Liu, K. J., Hsieh, C. L., Hu, T. S., Charoenfuprasert, S., Liu, H. K., et al. (2006). Identification of alpha-enolase as an autoantigen in lung cancer: its overexpression is associated with clinical outcomes. *Clin. Cancer Res.* 12, 5746–5754. doi: 10.1158/1078-0432.CCR-06-0324

Chen, J. M., Chiu, S. C., Chen, K. C., Huang, Y. J., Liao, Y. A., and Yu, C. R. (2020). Enolase 1 differentially contributes to cell transformation in lung cancer but not in esophageal cancer. *Oncol. Lett.* 19, 3189–3196. doi: 10.3892/ol.2020.11427

Chen, X., Xu, H., Wu, N., Liu, X., Qiao, G., Su, S., et al. (2017). Interaction between granulin A and enolase 1 attenuates the migration and invasion of human hepatoma cells. *Oncotarget* 8, 30305–30316. doi: 10.18632/oncotarget.16328

Cheng, Z., Shao, X., Xu, M., Zhou, C., and Wang, J. (2019). ENO1 acts as a prognostic biomarker candidate and promotes tumor growth and migration ability through the regulation of Rab1A in colorectal cancer. *Cancer Manag. Res.* 11, 9969–9978. doi: 10.2147/CMAR.S226429

Cho, H., Lee, J. H., Um, J., Kim, S., Kim, Y., Kim, W. H., et al. (2019). ENOblock inhibits the pathology of diet-induced obesity. *Sci. Rep.* 9, 493. doi: 10.1038/s41598-018-36715-3

Cho, H., Um, J., Lee, J. H., Kim, W. H., Kang, W. S., Kim, S. H., et al. (2017). ENOblock, a unique small molecule inhibitor of the non-glycolytic functions of enolase, alleviates the symptoms of type 2 diabetes. *Sci. Rep.* 7, 44186. doi: 10.1038/srep44186

Cook, K., Daniels, I., Symonds, P., Pitt, T., Gijon, M., Xue, W., et al. (2018). Citrullinated α-enolase is an effective target for anti-cancer immunity. *Oncoimmunology* 7, e1390642.

Dai, J., Zhou, Q., Chen, J., Rexius-Hall, M. L., Rehman, J., and Zhou, G. (2018). Alpha-enolase regulates the malignant phenotype of pulmonary artery smooth muscle cells via the AMPK-Akt pathway. *Nat. Commun.* 9, 3850. doi: 10.1038/s41467-018-06376-x

Dai, L., Qu, Y., Li, J., Wang, X., Wang, K., Wang, P., et al. (2017). Serological proteome analysis approach-based identification of ENO1 as a tumor-associated antigen and its autoantibody could enhance the sensitivity of CEA and CYFRA 21-1 in the detection of non-small cell lung cancer. *Oncotarget* 8, 36664–36673. doi: 10.18632/oncotarget.17067

De Rosa, V., Galgani, M., Porcellini, A., Colamatteo, A., Santopaolo, M., Zuchegna, C., et al. (2015). Glycolysis controls the induction of human regulatory T cells by modulating the expression of FOXP3 exon 2 splicing variants. *Nat. Immunol.* 16, 1174–1184. doi: 10.1038/ni.3269

Diaz-Ramos, A., Roig-Borrellas, A., Garcia-Melero, A., and Lopez-Alemany, R. (2012). Alpha-Enolase, a multifunctional protein: its role on pathophysiological situations. *J. Biomed. Biotechnol.* 2012, 156795. doi: 10.1155/2012/156795

Didiasova, M., Schaefer, L., and Wygrecka, M. (2019). When place matters: shuttling of Enolase-1 across cellular compartments. *Front. Cell. Dev. Biol.* 7:61. doi: 10.3389/fcell.2019.00061

Didiasova, M., Wujak, L., Wygrecka, M., and Zakrzewicz, D. (2014). From plasminogen to plasmin: role of plasminogen receptors in human cancer. *Int. J. Mol. Sci.* 15, 21229–21252. doi: 10.3390/ijms151121229

Didiasova, M., Zakrzewicz, D., Magdolen, V., Nagaraj, C., Bálint, Z., Rohde, M., et al. (2015). M.STIM1/ORAI1-mediated Ca2+ influx regulates Enolase-1 exteriorization. *J. Biol. Chem.* 290, 11983–11999. doi: 10.1074/jbc.M114.598425

Feo, S., Arcuri, D., Piddini, E., Passantino, R., and Giallongo, A. (2000). ENO1 gene product binds to the c-myc promoter and acts as a transcriptional repressor: relationship with Myc promoter-binding protein 1 (MBP-1). *FEBS Lett.* 473, 47–52. doi: 10.1016/s0014-5793(00)01494-0

Fu, Q. F., Liu, Y., Fan, Y., Hua, S. N., Qu, H. Y., Dong, S. W., et al. (2015). Alpha-enolase promotes cell glycolysis, growth, migration, and invasion in non-small cell lung cancer through FAK-mediated PI3K/AKT pathway. *J. Hematol. Oncol.* 8, 22. doi: 10.1186/s13045-015-0117-5

Gemta, L. F., Siska, P. J., Nelson, M. E., Gao, X., Liu, X., Locasale, J. W., et al. (2019). Impaired enolase 1 glycolytic activity restrains effector functions of tumor-infiltrating $CD8^+$ T cells. *Sci. Immunol.* 4, eaa9520. doi: 10.1126/sciimmunol.aap9520

Georges, E., Bonneau, A. M., and Prinos, P. (2011). RNAi-mediated knockdown of α-enolase increases the sensitivity of tumor cells to antitubulin chemotherapeutics. *Int. J. Biochem. Mol. Biol.* 2, 303–308.

Griggio, V., Mandili, G., Vitale, C., Capello, M., Macor, P., Serra, S., et al. (2017). Humoral immune responses toward tumor-derived antigens in previously untreated patients with chronic lymphocytic leukemia. *Oncotarget* 8, 3274–3288. doi: 10.18632/oncotarget.13712

Hamaguchi, T., Iizuka, N., Tsunedomi, R., Hamamoto, Y., Miyamoto, T., Iida, M., et al. (2008). Glycolysis module activated by hypoxia-inducible factor 1alpha is related to the aggressive phenotype of hepatocellular carcinoma. *Int. J. Oncol.* 33, 725–731.

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

Haque, A., Capone, M., Matzelle, D., Cox, A., and Banik, N. L. (2017). Targeting enolase in reducing secondary damage in acute spinal cord injury in rats. *Neurochem. Res.* 42, 2777–2787. doi: 10.1007/s11064-017-2291-z

He, P., Naka, T., Serada, S., Fujimoto, M., Tanaka, T., Hashimoto, S., et al. (2007). Proteomics-based identification of alpha-enolase as a tumor antigen in non-small lung cancer. *Cancer Sci.* 98, 1234–1240. doi: 10.1111/j.1349-7006.2007.00509.x

Hsiao, K. C., Shih, N. Y., Chu, P. Y., Hung, Y. M., Liao, J. Y., Chou, S. W., et al. (2015). Anti-α-enolase is a prognostic marker in postoperative lung cancer patients. *Oncotarget* 6, 35073–35086. doi: 10.18632/oncotarget.5316

Hsiao, K. C., Shih, N. Y., Fang, H. L., Huang, T. S., Kuo, C. C., Chu, P. Y., et al. (2013). Surface α-enolase promotes extracellular matrix degradation and tumor metastasis and represents a new therapeutic target. *PLoS One* 8:e69354. doi: 10.1371/journal.pone.0069354

Huang, Z., Lin, B., Pan, H., Du, J., He, R., Zhang, S., et al. (2019). Gene expression profile analysis of ENO1 knockdown in gastric cancer cell line MGC-803. *Oncol. Lett.* 17, 3881–3889. doi: 10.3892/ol.2019.10053

Isgrò, M. A., Bottoni, P., and Scatena, R. (2015). Neuron-specific enolase as a biomarker: biochemical and clinical aspects. *Adv. Exp. Med. Biol.* 867, 125–143. doi: 10.1007/978-94-017-7215-0_9

Ji, H., Wang, J., Guo, J., Li, Y., Lian, S., Guo, W., et al. (2016). Progress in the biological function of alpha-enolase. *Anim. Nutr.* 2, 12–17. doi: 10.1016/j.aninu.2016.02.005

Ji, M., Wang, Z., Chen, J., Gu, L., Chen, M., Ding, Y., et al. (2019). Up-regulated ENO1 promotes the bladder cancer cell growth and proliferation via regulating β-catenin. *Biosci. Rep.* 39, BSR20190503. doi: 10.1042/BSR20190503

Jiang, Z., Cui, Y., Wang, L., Zhao, Y., Yan, S., and Chang, X. (2013). Investigating citrullinated proteins in tumour cell lines. *World J. Surg. Oncol.* 11, 260. doi: 10.1186/1477-7819-11-260

Jung, D. W., Kim, W. H., Park, S. H., Lee, J., Kim, J., Su, D., et al. (2013). A unique small molecule inhibitor of enolase clarifies its role in fundamental biological processes. *ACS Chem. Biol.* 8, 1271–1282. doi: 10.1021/cb300687k

Kang, H. J., Jung, S. K., Kim, S. J., and Chung, S. J. (2008). Structure of human alpha-enolase (hENO1), a multifunctional glycolytic enzyme. *Acta Crystallogr. D Biol. Crystallogr.* 64, 651–657. doi: 10.1107/S0907444908008561

Kumari, S., and Malla, R. (2015). New insight on the role of plasminogen receptor in cancer progression. *Cancer Growth Metastasis* 8, 35–42. doi: 10.4137/CGM.S27335

La Rocca, C., Carbone, F., De Rosa, V., Colamatteo, A., Galgani, M., Perna, F., et al. (2017). Immunometabolic profiling of T cells from patients with relapsing-remitting multiple sclerosis reveals an impairment in glycolysis and mitochondrial respiration. *Metabolism* 77, 39–46. doi: 10.1016/j.metabol.2017.08.011

Liberti, M. V., and Locasale, J. W. (2016). Metabolism: a new layer of glycolysis. *Nat. Chem. Biol.* 12, 577–578. doi: 10.1038/nchembio.2133

Liu, Y., Li, H., Liu, Y., and Zhu, Z. (2018). MiR-22-3p targeting alpha-enolase 1 regulates the proliferation of retinoblastoma cells. *Biomed. Pharmacother.* 105, 805–812. doi: 10.1016/j.biopha.2018.06.038

Lundberg, K., Kinloch, A., Fisher, B. A., Wegner, N., Wait, R., Charles, P., et al. (2008). Antibodies to citrullinated alpha-enolase peptide 1 are specific for rheumatoid arthritis and cross-react with bacterial enolase. *Arthritis Rheum.* 58, 3009–3019. doi: 10.1002/art.23936

Magrys, A., Anekonda, T., Ren, G., and Adamus, G. (2007). The role of anti-alpha-enolase autoantibodies in pathogenicity of autoimmune-mediated retinopathy. *J. Clin. Immunol.* 27, 181–192. doi: 10.1007/s10875-006-9065-8

Mandili, G., Curcio, C., Bulfamante, S., Follia, L., Ferrero, G., Mazza, E., et al. (2020). In pancreatic cancer, chemotherapy increases antitumor responses to tumor-associated antigens and potentiates DNA vaccination. *J. Immunother. Cancer* 8, e001071. doi: 10.1136/jitc-2020-001071

Maranto, C., Perconti, G., Contino, F., Rubino, P., Feo, S., and Giallongo, A. (2015). Cellular stress induces cap-independent alpha-enolase/MBP-1 translation. *FEBS Lett.* 589, 2110–2116. doi: 10.1016/j.febslet.2015.06.030

Miles, L. A., Dahlberg, C. M., Plescia, J., Felez, J., Kato, K., and Plow, E. F. (1991). Role of cell-surface lysines in plasminogen binding to cells: identification of alpha-enolase as a candidate plasminogen receptor. *Biochemistry* 30, 1682–1691. doi: 10.1021/bi00220a034

Osthus, R. C., Shim, H., Kim, S., Li, Q., Reddy, R., Mukherjee, M., et al. (2000). Deregulation of glucose transporter 1 and glycolytic gene expression by c-Myc. *J. Biol. Chem.* 275, 21797–21800. doi: 10.1074/jbc.C000023200

Pancholi, V. (2001). Multifunctional alpha-enolase: its role in diseases. *Cell Mol. Life Sci.* 58, 902–920. doi: 10.1007/pl00000910

Perconti, G., Maranto, C., Romancino, D. P., Rubino, P., Feo, S., Bongiovanni, A., et al. (2017). Pro-invasive stimuli and the interacting protein Hsp70 favour the route of alpha-enolase to the cell surface. *Sci. Rep.* 7, 3841. doi: 10.1038/s41598-017-04185-8

Polcyn, R., Capone, M., Matzelle, D., Hossain, A., Chandran, R., Banik, N. L., et al. (2020). Enolase inhibition alters metabolic hormones and inflammatory factors to promote neuroprotection in spinal cord injury. *Neurochem. Int.* 139, 104788. doi: 10.1016/j.neuint.2020.104788

Pranay, A., Shukla, S., Kannan, S., Malgundkar, S. A., Govekar, R. B., Patil, A., et al. (2013). Prognostic utility of autoantibodies to α-enolase and Hsp70 for cancer of the gingivo-buccal complex using immunoproteomics. *Proteomics Clin. Appl.* 7, 392–402. doi: 10.1002/prca.201200081

Principe, M., Borgoni, S., Cascione, M., Chattaragada, M. S., Ferri-Borgogno, S., Capello, M., et al. (2017). Alpha-enolase (ENO1) controls alpha v/beta 3 integrin expression and regulates pancreatic cancer adhesion, invasion, and metastasis. *J. Hematol. Oncol.* 10, 16. doi: 10.1186/s13045-016-0385-8

Principe, M., Ceruti, P., Shih, N. Y., Chattaragada, M. S., Rolla, S., Conti, L., et al. (2015). Targeting of surface alpha-enolase inhibits the invasiveness of pancreatic cancer cells. *Oncotarget* 6, 11098–11113. doi: 10.18632/oncotarget.3572

Qian, X., Xu, W., Xu, J., Shi, Q., Li, J., Weng, Y., et al. (2017). Enolase 1 stimulates glycolysis to promote chemoresistance in gastric cancer. *Oncotarget* 8, 47691–47708. doi: 10.18632/oncotarget.17868

Qiao, G., Xu, H., Li, C., Li, X., Farooqi, A. A., Zhao, Y., et al. (2018b). Granulin A synergizes with cisplatin to inhibit the growth of human hepatocellular carcinoma. *Int. J. Mol. Sci.* 19, 3060. doi: 10.3390/ijms19103060

Qiao, H., Wang, Y., Zhu, B., Jiang, L., Yuan, W., Zhou, Y., et al. (2019). Enolase1 overexpression regulates the growth of gastric cancer cells and predicts poor survival. *J. Cell. Biochem.* 120, 18714–18723. doi: 10.1002/jcb.29179

Qiao, H., Wang, Y. F., Yuan, W. Z., Zhu, B. D., Jiang, L., and Guan, Q. L. (2018a). Silencing of ENO1 by shRNA inhibits the proliferation of gastric cancer cells. *Technol. Cancer Res. Treat.* 17, 1533033818784411. doi: 10.1177/1533033818784411

Ray, A., Song, Y., Du, T., Chauhan, D., and Anderson, K. C. (2020). Preclinical validation of Alpha-Enolase (ENO1) as a novel immunometabolic target in multiple myeloma. *Oncogene* 39, 2786–2796. doi: 10.1038/s41388-020-1172-0

Redlitz, A., Fowler, B. J., Plow, E. F., and Miles, L. A. (1995). The role of an enolase-related molecule in plasminogen binding to cells. *Eur. J. Biochem.* 227, 407–415. doi: 10.1111/j.1432-1033.1995.tb20403.x

Ren, G., and Adamus, G. (2004). Cellular targets of anti-alpha-enolase autoantibodies of patients with autoimmune retinopathy. *J. Autoimmun.* 23, 161–167. doi: 10.1016/j.jaut.2004.06.003

Sanchez, T. W., Zhang, G., Li, J., Dai, L., Mirshahidi, S., Wall, N. R., et al. (2016). Immunoseroproteomic profiling in African american men with prostate cancer: evidence for an autoantibody response to glycolysis and plasminogen-associated proteins. *Mol. Cell. Proteomics.* 15, 3564–3580. doi: 10.1074/mcp.M116.060244

Santana-Rivera, Y., Rabelo-Fernández, R. J., Quiñones-Díaz, B. I., Grafals-Ruíz, N., Santiago-Sánchez, G., Lozada-Delgado, E. L., et al. (2020). Reduced expression of enolase-1 correlates with high intracellular glucose levels and increased senescence in cisplatin-resistant ovarian cancer cells. *Am. J. Transl. Res.* 12, 1275–1292.

Satani, N., Lin, Y. H., Hammoudi, N., Raghavan, S., Georgiou, D. K., and Muller, F. L. (2016). ENOblock does not inhibit the activity of the glycolytic enzyme enolase. *PLoS One* 11:e0168739. doi: 10.1371/journal.pone.0168739

Sedoris, K. C., Thomas, S. D., and Miller, D. M. (2007). c-myc promoter binding protein regulates the cellular response to an altered glucose concentration. *Biochemistry* 46, 8659–8668. doi: 10.1021/bi7003558

Sedoris, K. C., Thomas, S. D., and Miller, D. M. (2010). Hypoxia induces differential translation of enolase/MBP-1. *BMC Cancer* 10:157. doi: 10.1186/1471-2407-10-157

Shih, N. Y., Lai, H. L., Chang, G. C., Lin, H. C., Wu, Y. C., Liu, J. M., et al. (2010). Anti-alpha-enolase autoantibodies are down-regulated in advanced cancer patients. *Jpn. J. Clin. Oncol.* 40, 663–669. doi: 10.1093/jjco/hyq028

Song, Y., Luo, Q., Long, H., Hu, Z., Que, T., Zhang, X., et al. (2014). Alpha-enolase as a potential cancer prognostic marker promotes cell growth, migration, and invasion in glioma. *Mol. Cancer* 13, 65. doi: 10.1186/1476-4598-13-65

Subramanian, A., and Miller, D. M. (2000). Structural analysis of alpha-enolase. Mapping the functional domains involved in down-regulation of the c-myc protooncogene. *J. Biol. Chem.* 275, 5958–5965. doi: 10.1074/jbc.275.8.5958

Sun, L., Guo, C., Cao, J., Burnett, J., Yang, Z., Ran, Y., et al. (2017). Over-expression of alpha-Enolase as a prognostic biomarker in patients with pancreatic cancer. *Int. J. Med. Sci.* 14, 655–661. doi: 10.7150/ijms.18736

Sun, L., Lu, T., Tian, K., Zhou, D., Yuan, J., Wang, X., et al. (2019). Alpha-enolase promotes gastric cancer cell proliferation and metastasis via regulating AKT signaling pathway. *Eur. J Pharmacol.* 845, 8–15. doi: 10.1016/j.ejphar.2018.12.035

Takashima, M., Kuramitsu, Y., Yokoyama, Y., Iizuka, N., Fujimoto, M., Nishisaka, T., et al. (2005). Overexpression of alpha enolase in hepatitis C virus-related hepatocellular carcinoma: association with tumor progression as determined by proteomic analysis. *Proteomics* 5, 1686–1692. doi: 10.1002/pmic.200401022

Tomaino, B., Cappello, P., Capello, M., Fredolini, C., Sperduti, I., Migliorini, P., et al. (2011). Circulating autoantibodies to phosphorylated alpha-enolase are a hallmark of pancreatic cancer. *J. Proteome Res.* 10, 105–112. doi: 10.1021/pr100213b

Tsai, S. T., Chien, I. H., Shen, W. H., Kuo, Y. Z., Jin, Y. T., Wong, T. Y., et al. (2010). ENO1, a potential prognostic head and neck cancer marker, promotes transformation partly via chemokine CCL20 induction. *Eur. J. Cancer* 46, 1712–1723. doi: 10.1016/j.ejca.2010.03.018

Tu, S. H., Chang, C. C., Chen, C. S., Tam, K. W., Wang, Y. J., Lee, C. H., et al. (2010). Increased expression of enolase alpha in human breast cancer confers tamoxifen resistance in human breast cancer cells. *Breast Cancer Res. Treat.* 121, 539–553. doi: 10.1007/s10549-009-0492-0

Wang, L., Bi, R., Yin, H., Liu, H., and Li, L. (2019). ENO1 silencing impairs hypoxia-induced gemcitabine chemoresistance associated with redox modulation in pancreatic cancer cells. *Am. J. Transl. Res.* 11, 4470–4480.

Wang, L., Qu, M., Huang, S., Fu, Y., Yang, L., He, S., et al. (2018). A novel α-enolase-targeted drug delivery system for high efficacy prostate cancer therapy. *Nanoscale* 10, 13673–13683. doi: 10.1039/c8nr03297a

Wygrecka, M., Marsh, L. M., Morty, R. E., Henneke, I., Guenther, A., Lohmeyer, J., et al. (2009). Enolase-1 promotes plasminogen-mediated recruitment of monocytes to the acutely inflamed lung. *Blood* 113, 5588–5598. doi: 10.1182/blood-2008-08-170837

Xu, X., Chen, B., Zhu, S., Zhang, J., He, X., Cao, G., et al. (2019). Hyperglycemia promotes Snail-induced epithelial-mesenchymal transition of gastric cancer via activating ENO1 expression. *Cancer Cell Int.* 19, 344. doi: 10.1186/s12935-019-1075-8

Yin, H., Wang, L., and Liu, H. L. (2018). ENO1 overexpression in pancreatic cancer patients and its clinical and diagnostic significance. *Gastroenterol. Res. Pract.* 2018, 3842198. doi: 10.1155/2018/3842198

Yu, S., Li, N., Huang, Z., Chen, R., Yi, P., Kang, R., et al. (2018). A novel lncRNA, TCONS_00006195, represses hepatocellular carcinoma progression by inhibiting enzymatic activity of ENO1. *Cell Death Dis.* 9, 1184. doi: 10.1038/s41419-018-1231-4

Zakrzewicz, D., Didiasova, M., Krüger, M., Giaimo, B. D., Borggrefe, T., Mieth, M., et al. (2018). Protein arginine methyltransferase 5 mediates enolase-1 cell surface trafficking in human lung adenocarcinoma cells. *Biochim. Biophys. Acta Mol. Basis Dis.* 1864, 1816–1827. doi: 10.1016/j.bbadis.2018.02.021

Zakrzewicz, D., Didiasova, M., Zakrzewicz, A., Hocke, A. C., Uhle, F., Markart, P., et al. (2014). The interaction of enolase-1 with caveolae-associated proteins

regulates its subcellular localization. *Biochem. J.* 460, 295–307. doi: 10.1042/BJ20130945

Zang, H. Y., Gong, L. G., Li, S. Y., and Hao, J. G. (2020). Inhibition of α-enolase affects the biological activity of breast cancer cells by attenuating PI3K/Akt signaling pathway. *Eur. Rev. Med. Pharmacol. Sci.* 24, 249–257. doi: 10.26355/eurrev_202001_19917

Zang, R., Li, Y., Jin, R., Wang, X., Lei, Y., Che, Y., et al. (2019). Enhancement of diagnostic performance in lung cancers by combining CEA and CA125 with autoantibodies detection. *Oncoimmunology* 8, e1625689. doi: 10.1080/2162402X.2019.1625689

Zhan, P., Zhao, S., Yan, H., Yin, C., Xiao, Y., Wang, Y., et al. (2017). α-enolase promotes tumorigenesis and metastasis via regulating AMPK/mTOR pathway in colorectal cancer. *Mol. Carcinog.* 56, 1427–1437. doi: 10.1002/mc.22603

Zhang, L., Lu, T., Yang, Y., and Hu, L. (2020). α-enolase is highly expressed in liver cancer and promotes cancer cell invasion and metastasis. *Oncol. Lett.* 20, 152. doi: 10.3892/ol.2020.12003

Zhang, L., Wang, H., and Dong, X. (2018). Diagnostic value of α-enolase expression and serum α-enolase autoantibody levels in lung cancer. *J. Bras. Pneumol.* 44, 18–23. doi: 10.1590/S1806-37562016000000241

Zhou, W., Capello, M., Fredolini, C., Piemonti, L., Liotta, L. A., Novelli, F., et al. (2010). Mass spectrometry analysis of the post-translational modifications of alpha-enolase from pancreatic ductal adenocarcinoma cells. *J. Proteome Res.* 9, 2929–2936. doi: 10.1021/pr901109w

Zhu, X., Miao, X., Wu, Y., Li, C., Guo, Y., Liu, Y., et al. (2015). ENO1 promotes tumor proliferation and cell adhesion mediated drug resistance (CAM-DR) in Non-Hodgkin's Lymphomas. *Exp. Cell Res.* 335, 216–223. doi: 10.1016/j.yexcr.2015.05.020

Zuo, J., Wang, B., Long, M., Gao, Z., Zhang, Z., Wang, H., et al. (2018). The type 1 transmembrane glycoprotein B7-H3 interacts with the glycolytic enzyme ENO1 to promote malignancy and glycolysis in HeLa cells. *FEBS Lett.* 592, 2476–2488. doi: 10.1002/1873-3468.13164

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.

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