Unveiling Prognostics Biomarkers of Tyrosine Metabolism Reprogramming in Liver Cancer by Cross-platform Gene Expression Analyses

Tran Ngoc Nguyen[1*], Ha Quy Nguyen[2], Duc-Hau Le[1]

1. Department of Computational Biomedicine, Big Data Institute, Vingroup, Hanoi, Vietnam
2. Department of Computer Vision, Big Data Institute, Vingroup, Hanoi, Vietnam

* Corresponding author. Email: v.tranNN3@vintech.net.vn

Keywords: hepatocellular carcinoma, tyrosine, microRNA, GSTZ1, metabolic reprogramming, pan-cancer, cross-platform

Abstract

Tyrosine is mainly degraded in the liver by a series of enzymatic reactions. Abnormal expression of the tyrosine catabolic enzyme tyrosine aminotransferase (TAT) has been reported in patients with hepatocellular carcinoma (HCC). Despite this, aberration in tyrosine metabolism has not been investigated in cancer development. In this work, we conduct comprehensive cross-platform study to obtain foundation for discoveries of potential therapeutics and preventative biomarkers of HCC. We explore data from The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), Gene Expression Profiling Interactive Analysis (GEPIA), Oncomine and Kaplan Meier plotter (KM plotter) and performed integrated analyses to evaluate the clinical significance and prognostic values of the tyrosine catabolic genes in HCC. We find that five tyrosine catabolic enzymes are downregulated in HCC compared to normal liver at mRNA and protein level. Moreover, low expression of these enzymes correlates with poorer survival in patients with HCC. Notably, we identify pathways and upstream regulators that might involve in tyrosine catabolic reprogramming and further drive HCC development. In total, our results underscore tyrosine metabolism alteration in HCC and lay foundation for incorporating these pathway components in therapeutics and preventative strategies.

Introduction

Hepatocellular carcinoma (HCC) remains the most common cancer in the word, especially in Asia and Africa, and the third leading cause of cancer-related death worldwide[1]. It is believed that the pathogenesis of HCC is a long-term process that involves constant metabolic reprogramming. Previous efforts to investigate metabolic programming of HCC have largely focused on aerobic

glycolysis, commonly referred to as the Warburg effect, which supports tumor growth in part by accumulating glycolytic intermediates for anabolic biosynthesis[2,3]. For instance, HCC tumors express high levels of the hexokinase isoform 2 (HK2), which converts glucose to glucose-6-phosphate, and its expression is associated with the pathological stage of the tumor[4,5]. HK2 silencing acted synergistically with sorafenib to inhibit HCC tumor growth in mice[5]. Besides glucose, HCC has been reported to alter its lipid and lipoprotein catabolic and anabolic pathways and increased HCC risks have been observed in patients with obesity[6], diabetes[7], and hepatic steatosis[8]. Recent studies defined a functional association among lipogenesis, multifunctional enzyme fatty acid synthase (FASN), sterol regulatory element-binding protein-1 (SREBP-1), a transcription factor regulating FASN expression, and HCC[9,10].

Recently there are increasing evidences suggesting that cancer cells have increased levels of oxidative stress and ROS production compared to normal cells[11]. Thus, redox homeostasis is finely tuned in cancer cells with a role in the control of cell signaling and metabolism[12]. For instance, ROS-mediated inhibition of PKM2 allows cancer cells to sustain antioxidant responses by diverting glucose flux into the pentose phosphate pathway and increasing the production of reducing equivalents for ROS detoxification[12]. Oxidative damage is considered as a key pathway in HCC progression and increases patient vulnerability for HCC recurrence[13]. As previously reported, accumulation of a *m*-tyrosine may disrupt cellular homeostasis and contribute to disease pathogenesis and the elimination of this isomer can be an effective defense against oxidative stress[14].

Tyrosine, like other amino acids, is the building block for proteins as well as an alternative energy source for cellular functions. Liver is the major organ where tyrosine degradation takes place to produce intermediates or precursors for gluconeogenesis and ketogenesis. The degradation of

tyrosine is catalyzed through a series of five enzymatic reactions. Disturbed tyrosine metabolism has been implicated in several types of disease such as Huntington's disease[15] and esophageal cancer[16,17]. Previously reported. patients with hereditary tyrosinemia are more likely to develop HCC[18,19]. In patients with HCC, an upregulation of serum tyrosine has been recorded[20,21], suggesting a deregulated tyrosine metabolism in HCC. However, to date, there is a lack of systematic study to profile the state of tyrosine catabolic enzymes and molecular impacts of alteration in tyrosine catabolism in HCC development.

As previously reported, the frequent deletion of 16q22 and aberrant methylation led to the downregulation of the first tyrosine catabolic enzyme TAT (tyrosine aminotransferase)[22]. Functional analyses showed that TAT harbored proapoptotic effect and that TAT suppression could promote liver tumorigenesis[22]. Glutathione S-transferases (GSTs) are a family of phase II isoenzymes that detoxify toxicant to lower toxic[23] and its dysfunction has been found to be closely related with response to chemotherapy[24-26]. *GSTZ1* belongs to the zeta class of GSTs and is the fourth enzyme in tyrosine metabolism. Patients carrying GSTZ1 variants had an increased risk of bladder cancer when exposed to trihalomethanes[27]. Furthermore, a computational-based investigation suggested *GSTZ1* might act as a protective factor in ovarian cancer[28].

In this study, we aim to systematically investigate the expression and prognostic value of tyrosine catabolism enzymes (TAT, HPD, HGD, GSTZ1 and FAH) in HCC by integrating large-scale datasets. We further detect enriched pathways associated with overexpression of a tyrosine catabolic enzyme in HCC cells. Our comprehensive, gene-centric analysis shed light on the genomic changes, clinical relevance, upstream regulators and possible impact of tyrosine catabolic genes on HCC development.

Results

A cross-platform, pan-cancer analysis of tyrosine catabolic enzyme expression

We first set out to investigate the expression profiles of tyrosine catabolic genes in cancer transcriptomes (Figure 1A). Here, we used the Oncomine online database[29] to perform pan-cancer transcriptome analysis on its available data sets. The top mRNA differences between cancer samples and normal samples were analyzed by default selective criteria. Figure 1B showed that there was a total of 390, 428, 431, 457 and 448 Oncomine data sets involving the genes, *TAT*, *HPD*, *HGD*, *GSTZ1* and *FAH*, respectively. Remarkably, in most data sets, a large proportion of patients demonstrated downregulation of these genes in the tumorous parts compared to those of normal samples. Specifically, in HCC, all of the gene sets show downregulation of the investigated tyrosine catabolic enzyme-encoding genes (Figure 1B, highlighted in red box). Furthermore, the Gene expression heat map from GEPIA pan-cancer transcriptome analysis[30] showed markedly downregulation of TAT, HPD and GSTZ1 in HCC (Figure 1C). Additionally, in cervical squamous cell carcinoma, all of the tyrosine catabolic genes were visibly downregulated in tumors compared to normal tissue adjacent to the tumor. Through this initial observation, we found evidences to support that tyrosine catabolic genes expression were downregulated in many cancers, including HCC.

Tyrosine catabolic genes are downregulated in HCC

Next, to further investigate the role of tyrosine catabolic enzymes, we performed analysis of a publicly available dataset (The Cancer Genome Atlas[31] [TCGA], Liver Cancer [LIHC]) including gene expression in 369 HCC tissues vs 160 normal liver tissues. Here, the data demonstrated that TAT, HPD and GTSZ1 were decreased in HCC tissues compared to normal liver (Figure 2A).

However, the gene expression of HGD and FAH were virtually unchanged in HCC samples compare to normal liver samples.

To gain supporting evidence on the downregulation of tyrosine catabolic genes in HCC, the GSE89377 (Data Citation 1) dataset was employed to assess the expression of these genes in normal liver samples, early HCC and HCC from stage 1 to 3. Interestingly, we found that in early HCC, the expression of tyrosine catabolic genes was insignificantly changed compared to normal liver. However, the transcripts of *TAT*, *HPD*, *HGD*, *GSTZ1* and *FAH* significantly reduced in the HCC stage 2 and stage 3 compared to normal liver (Figure 2B).

Overall, our combined analysis on TCGA data and an independent GSE dataset showed that tyrosine catabolic genes were downregulated in late stage HCC compared to normal liver.

Prognostic value of tyrosine catabolic genes in patients with HCC

Subsequently, we sought to determine the clinical relevance of *TAT*, *HPD*, *HGD*, *GSTZ1* and *FAH* expression in term of prognosis in HCC patients since these genes were highly enriched in liver tissues (Supplementary Figure S1). Kaplan–Meier analysis was employed to compare between the subgroups with high and low gene expression (using the median, 25% or 75% quartile values of gene expression as cut-off points) in TCGA-LIHC cohort of 364 liver cancer patients. The overall survival was significantly associated with *TAT*, *HGD* and *GSTZ1* expression in HCC samples ($p = 0.0067$, $p = 0.0039$ and $p = 0.036$, respectively) (Figure 3). Similarly, lower expression of *TAT*, *HGD* and *GSTZ1* could also translate to a worse disease-free survival in HCC patients ($p = 0.011$, $p = 0.0038$ and $p = 0.036$, respectively) (Supplementary Figure S2).

To further validate the potential application of tyrosine catabolic genes in the clinic, we extracted the characterized IHC images from the Human Protein Atlas. HCC tumor tissue staining of tyrosine catabolic enzymes showed significant decrease in positive staining compared with normal liver tissue. Specifically, HPD staining decreased by 2.26-fold ± 2.10 ($p = 0.0388$), HGD decreased by 1.67-fold ±0.87 ($p = 0.0423$) and GSTZ1 decreased by 2.27-fold ±1.09 ($p = 0.0007$) in HCC tumor compared to normal liver tissue (Figure 4).

These findings highlighted that the expression of tyrosine catabolic enzyme-encoding genes correlated with worse overall survival and disease-free survival in HCC and that TAT, HGD and GSTZ1 had potential prognostic value in patients with HCC.

Gene expression profiling of GSTZ1 expressing HCC cell line

Following the previous analyses, we noted that the fourth rate-limiting enzyme, GSTZ1 (Glutathione S-transferase Zeta 1) had significant downregulation and prognosis. We therefore sought to study the molecular pathway alterations associated with this gene. Here, we explored the publicly available data set GSE117822 (Data Citation 2) where GSTZ1 is overexpressed in Huh7 HCC cell line by adenoviral transfection. R software[32] with the DESeq2[33] package was applied to screen DEGs from the gene expression dataset GSE between control vectors and overexpressed GSTZ1. A total of 3163 DEGs ($p < 0.05$) were identified from this dataset, 1742 upregulated genes and 1421 downregulated genes.

To investigate changes in molecular pathways associated with GSTZ1 overexpression, we use GSEA to rank the DEGs against the C2 canonical pathway gene set[34]. We were able to profile positively and negatively enriched pathways in GSTZ1 overexpressed Huh7 (Supplementary Figure S3). For better visualization of related gene sets and identification of important pathway

families, we presented the pathways using Enrichment Map[35] in Cytoscape[36] (Figure 5). As expected, we observed a positive enrichment for multiple metabolism related pathways including Metabolism of Lipids, Metabolism of Proteins and Metabolism of Amino Acids. Noticeably, increased GSTZ1 expression led to heightened Oxidative Phosphorylation and Respiratory Electron Transport. On the other hand, genes involved in glycolysis, such as *HK2*, *PDK2* were downregulated (1.88-fold and 2.05-fold, respectively) in cell expressing GSTZ1 compared with vector control. Most importantly, overexpression of GSTZ1 in HCC cell led to the downregulation in several pathways in cancer gene sets (Kegg Small Cell Lung Cancer and Kegg Chronic Myeloid Leukemia). Together, these data highlighted the changes in molecular pathways that correspond to GSTZ1 expression and critically, provided insights on how overexpression of GSTZ1 might negate HCC development.

Mutation profiles of tyrosine catabolic genes in HCC

We extended our studies to investigate on underlying mechanism of how tyrosine catabolic genes were downregulated in HCC. First, we explored mutation profiles of tyrosine catabolic genes in 353 HCC patients by exploring TCGA data using cBioPortal[37]. We found that each individual gene was mutated in less than 1.1% of patients with HCC (Supplementary Figure S4). In all genes, there were 8/21 missense mutations that harbor deleterious effect (Table 1). However, when incorporating mutation type with mRNA expression profile, we did not observe a correlation where amplification led to increased expression or vice versa (Supplementary Figure S5). Second, we explored copy number status of tyrosine catabolic genes using data from GISTIC analysis[38] (TCGA Copy Number Portal). We found that even though the genes were located near the peak region of deletion, none of them were in focal (Table 2). Except for HPD (Q value = 0.019), the

rest of the genes were less likely to suffer copy number alterations (Table 2). Taken together, we found that several base substitution mutation scenarios can lead to the deletion of tyrosine catabolic genes but these is not a strong association between mutation status and mRNA expression.

MicroRNAs regulate the expression of tyrosine catabolic genes

Next, we sought out to explore microRNAs as possible negative regulators of *TAT*, *HPD*, *HGD*, *GSTZ1* and *FAH*. Using Target Scan database[39], we found there were two microRNAs that targeted *TAT*, *HPD*, *GSTZ1* and *FAH* (Figure 6A), which were miR-539 and miR-661. There were no common microRNAs that target all tyrosine catabolic genes. First, investigation of 370 HCC samples and 50 normal samples (TCGA-LIHC) showed that miR-539 increased by 2.84-fold in HCC samples compared to normal liver ($p = 0.05$). Second, pan-cancer co-expression analysis for miRNA-target interaction in HCC using starBase[40] showed that miR-539 level negatively correlated with TAT, HPD, GSTZ1 and FAH expression ($r = -0.221$, $r = -0.193$, $r = -0.123$, $r = -0.166$) (Supplementary Figure S6). More importantly, our Kaplan-Meier analysis by KM-plotter[41] of TCGA-LIHC data set showed that high miR-539 expression led to worse overall survival in in HCC patients (Figure 6B).

Additionally, Kaplan-Meier analysis on CapitalBio miRNA Array liver dataset[42] also showed that miR-661 expression positively correlated with worse overall survival (Figure 6C). Overall, these findings suggested that in HCC, the downregulation of tyrosine catabolic genes can be due to microRNA regulation. We found that miR-539 and miR-661 can potentially suppress TAT, HPD, GSTZ1 and FAH expression and that expression of miR-539 and miR-661 can provide prognostic insights for patients with HCC.

Discussion

We explored publicly available gene expression data sets and database to investigate the roles of genes in the tyrosine degradation pathway in the development of HCC. Our results indicated that all tyrosine catabolic genes decreased in HCC compared to normal liver tissues. Furthermore, we found that these genes gradually decreased from normal liver through early HCC to late HCC. We demonstrated that the fourth rate-limiting enzyme, GSTZ1 expression significantly reduced, either in protein or mRNA level, in HCC (Figure 2A, Figure 4). Even though the tyrosine catabolic gene expression remained unchanged at early stage HCC, they were significantly down-regulated in late stage HCC (Figure 2B). We also found that TAT, HGD and GSTZ1 expression levels positively correlated with overall survival and disease-free survival of HCC (Figure 3, Supplementary Figure S2). Previously shown, TAT, the first rate-limiting enzymes in the pathways, was downregulated in HCC, possibly due to the frequent deletion of 16q[22,45]. Another study found that aberrant DNA hypermethylation on chromosome 16 has been described as an early event in HCC tumorigenesis[46]. Functional *in vitro* validations showed that TAT induced apoptosis and that TAT possessed tumor-suppressive functions[22].

GSTZ1, which is expressed in both hepatic cytosol and mitochondria, has shown to be oxidative stress-related[28]. High levels of GSTZ1 expression conferred resistance to the effect of anti-cancer therapy of dichloroacetate in hepatocellular carcinoma cell lines by an independent mechanism to tyrosine metabolism[47,48]. We decided to further study the roles of GSTZ1 in HCC development by exploring a public dataset where GSTZ1 were overexpressed in HCC cell line Huh7. We found that with the expression of GSTZ1, there was positive enrichment of Biological Oxidations (Figure 5, Supplementary Figure S2). Generally, cell generates ATP through oxidative phosphorylation and produces ROS as a byproduct. Cancer cells have a higher tolerance for ROS and it is known

that low doses of ROS can stimulate growth in various types of cancer[11]. However, unbalanced increase in ROS level can induce cancer cell cycle arrest, senescence and apoptosis[11]. Here, it is appealing to assume that HCC can reprogram its tyrosine metabolism to maintain ROS balance as a growth strategy. Additionally, we detected an overall enrichment in metabolism of proteins and lipids pathways and decrease in glycolysis genes following GSTZ1 expression (Figure 5). Liver is a dynamic organ which constantly undergoes metabolic shift. Cancer cells, including HCC, usually switch to aerobic glycolysis to maximize energy usage and further fuel growth[2]. Since overexpression of GSTZ1 led to downregulation of several glycolysis genes, we consider it possible that the suppression of tyrosine catabolism can be a mechanism by which HCC switch to aerobic glycolysis during cancer progression.

The downregulation of other genes in the tyrosine catabolic pathways have not been linked to changes in DNA. Thus, we reason that the downregulation of *HPD, HGD, GSTZ1* and *FAH* might be dependent or independent of the downregulation of TAT. We found that four out of five genes were predicted to be regulated by miR-539, miR-661. Noticeably, investigation of TCGA-LIHC dataset showed that miR-539 significantly increased in HCC compared to normal skin and that the miR-539 level inversely correlated with expression of *TAT, GSTZ1, HPD, FAH* (Supplementary Figure S4). Here, our analyses showed that expression of two of these microRNAs positively correlated with overall survival in HCC patients (Figure 6B). As previously reported, miR-539 was usually downregulated and acted as tumor suppressors in various tumor types[49,50]. In HCC, miR-539 was also demonstrated to suppress HCC development *in vitro* by targeting FSCN1 and suppressing apoptosis[51,52]. Here, our findings suggested that on miR-539 might be a tumor promoter in contrast to previous experimental studies. On the other hand, prior studies showed that miR-661 was a tumor promoter in non-small cell lung cancer, colon cancer and ovarian cancer [53-

[55]. However, the roles of miR-661 in HCC development has not been investigated. Taken together, we speculate that miR-539 and miR-661 can be potential regulators of tyrosine catabolic genes and whether these regulation lead to HCC development need to be validated by functional studies.

Tyrosine metabolism is an important process that is often dysregulated in various diseases including cancers and chronic disorders[56]. Tyrosinemia type I patients have a higher risk of developing HCC[56]. The reasons for the high incidence of HCC are unknown but it has been suggested that it may be caused by accumulated metabolites such as fumarylacetoacetate (FAA) and maleylacetoacetate (MAA)[56]. A metabolomics study on esophageal cancer (EC) showed that tyrosine decreased in serum of patients with EC compared with healthy control[57,58]. There has been little evidence on how tyrosine metabolism might contribute to cancer development even though changes in expression of some tyrosine metabolic genes have been reported in HCC patients[22,59].

To summarize, our findings from the integrative databases and comprehensive analysis of this study demonstrated the downregulation of tyrosine catabolic genes and their prognostic value in HCC. We provided evidence on how suppressing these genes can benefit HCC development and that tyrosine catabolism is a novel pathway through which HCC reprogram its metabolism. Finally, we provided evidence suggesting that microRNAs can regulate the expression of tyrosine catabolic genes and might be a potential prognostic biomarker for HCC. However, more investigations need to be applied to fully reveal the role of tyrosine catabolism in HCC for further translational study.

We further discovered that the expression of GSTZ1, the fourth rate limiting enzyme in tyrosine catabolism, regulates glycolytic gene expression. We provided evidence to support that tyrosine catabolic genes can be regulated by microRNAs. Thus, we present here a novel function for tyrosine catabolic genes in tumorigenesis and provide a previously unappreciated event by which cancer cells reprogram tyrosine metabolism during cancer progression.

Methods

Oncomine analysis

The Oncomine online databases[29] were accessed for the visualization of gene expression. Oncomine is an online cancer microarray database used to facilitate and promote discoveries from genome-wide expression analyses. The pan-cancer studies in Oncomine were selected to compare the expression levels in tumor vs normal tissue adjacent to the tumor. The selection criteria for the Oncomine studies were $p < 0.05$ as a threshold, 2-fold change and gene rank in the top 10%. The p value, fold changes, and cancer subtypes were extracted.

Gene expression analysis

The TCGA data was analysed by GEPIA[30] (http://gepia.cancer-pku.cn/). For the differential expression analysis, the genes were log2(TPM + 1) transformed. One-way ANNOVA was used to compute p value. Those with $\log2(TPM+1) > 1$ and $p < 0.01$ were then considered differentially expressed genes. Normal tissues are matched TCGA adjacent tissue and GTEx normal tissue.

Survival analysis

The overall survival curves of *TAT*, *HPD*, *HGD*, *GSTZ1* and *FAH* were investigated using the Kaplan-Meier method with the log-rank test. We set the high and low gene expression level groups by the median value. The overall survival plot was obtained with the hazard ratios (HR) and the 95% confidence interval information. The whole process was implemented using the web-based tool GEPIA[30].

The prognostic values of hsa-miR-539 and has-miR-661 in HCC were analyzed using Kaplan Meier plotter (KM plotter) database[44]. In brief, the miRNAs were entered into the database, after which survival plots were generated and hazard ratio, 95% confidence intervals, log rank *P*-value were displayed on the webpage. The log-rank p value was calculated with <0.05 considered statistically significant.

Gene expression omnibus data mining

We retrieved transcriptome profiles of HCC tissues from GEO which is a public genomics database, allowing users to investigate gene expression profiles of interest[60]. The GSE89377 is a microarray dataset of multi-stage HCC in a GPL570 Affymetrix Human Genome U133 Plus 2.0 Array Platform. The GSE89377 dataset contains 108 samples in total, including 13 healthy people, 5 with early HCC, 9 with Stage 1 HCC, 12 with Stage 2 HCC and 14 with Stage 3 HCC. Processed gene expression dataset was downloaded using *GEOquery*[61]. *limma*[62] R packages was used to determine the DEGs between normal and HCC tissues. p<0.01 was considered as the cutoff value.

Differentially expressed genes identification and Gene set enrichment analysis (GSEA)

The GSE89377 dataset published in 2017 was processed by Bioconductor package DESeq2[33] to identify DEGs and analyzed by GSEA with the Molecular Signatures Database "Canonical Pathways" gene set collection[34]. The default GSEA basic parameters were used; to find gene sets that correlate with *GSTZ1* expression profile (continuous phenotype label), Pearson metric was used for ranking genes.

Quantification of immunohistochemistry images from Human Protein Atlas

Immunohistochemistry (IHC) images were downloaded from the publicly available The Human Protein Atlas[63] (HPA; http://www.proteinatlas.org) version 8.0. The analyses in present study were performed using HPA images of liver sections that were labeled with antibodies for HPD (antibody HPA038321), HGD (antibody HPA047374), GSTZ1 (antibody HPA004701) and FAH (antibody HPA041370). A custom script written in MATLAB programming language was used to detect positive staining based on brown pixel-counting. The absolute amount of antibody-specific chromogen per pixel was determined and normalized against total tissue area. Code is available at http://github.com/nguyenquyha/IHC-method.

Identify miRNA candidates by Targetscan

Targetscan database (http://www.targetscan.org) were accessed for identifying miRNA candidates. In brief, gene name was entered to retrieve a list of microRNAs that was predicted to target the input gene. Default parameters were used. After that, the miRNA lists were merged to find common miRNAs that target *TAT*, *HPD*, *GSTZ1* and *FAH*.

Copy Number Analysis

Copy number alteration data from Gene-Centric GISTIC analyses was retrieved from TCGA Copy Number Portal (http://portals.broadinstitute.org/tcga/home). Liver hepatocarcinoma tumor type was selected for this analysis using the stddata__2015_04_02 TCGA/GDAC tumor sample sets from FireHose.

Statistical Analysis

Statistical analyses were performed using GraphPad Prism 8.0.2. Independent Student's t test was used to compare the mean value of two groups. Bars and error represent mean ± standard deviations (SD) of replicate measurements. Statistical significance was defined as $p \leq 0.05$. $^{*}p < 0.05$, $^{**}p < 0.01$ and $^{***}p < 0.001$.

Resource table

Software and Algorithms	Version	Source
GraphPad PRISM	8.0.2	https://www.graphpad.com
R	3.5.3	https://www.r-project.org/
limma R package	3.8	https://bioconductor.org/packages/release/bioc/html/limma.html
Cytoscape	3.7.1	https://cytoscape.org/
EnrichmentMAP	3.2.0	http://apps.cytoscape.org/apps/enrichmentmap
GEPIA	1	http://gepia.cancer-pku.cn
Oncomine	NA	https://www.oncomine.org
KMPlotter	NA	https://kmplot.com
GSEA software	2-2.2.3	http://software.broadinstitute.org/gsea/index.jsp

Conflict of Interest

The authors declare that this study received funding from Vingroup. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Acknowledgements

We also thank Dr. Ban Xuan Dong at Institute for Molecular Engineering, University of Chicago for proof-reading the manuscript.

Author Contributions

T.N.N conceived the study, performed research, analyzed, interpreted the data and wrote the manuscript. H.Q.N contributed to the data analysis. D.H.L provided critical scientific input. The final manuscript was reviewed and approved by all listed authors.

Figure Legends

Figure 1: Downregulation of the tyrosine catabolic genes in several types of cancer, including HCC**.** (A) Graphics of tyrosine catabolism process. (B) The mRNA expression levels of the tyrosine catabolic genes according to Oncomine database. The mRNA expression of the genes (cancer versus normal tissue) in pan-cancers analyzed with the Oncomine database. The graphic demonstrates the numbers of datasets that meet our threshold in each cancer type. Cell color was defined as the gene rank percentile in the study. (C) The heat map indicates the expression after normalization by TPM+1 for comparison between tumor (T) and normal (N) across cancer types.

Normal tissues are matched TCGA adjacent tissue and GTEx data. The cancer abbreviation names are shown according to TCGA study abbreviations (Supplementary Table S1). TPM, transcript per million.

Figure 2: Gene expression profile of the tyrosine catabolic genes in HCC. (A) Gene expression analysis of tyrosine catabolic genes using GEPIA based on the TCGA and GTEx database. Box plots represent the gene expression level in terms of $\log_2$(TPM+1) in the tumor (red, n=369) and normal (grey, n=160) samples, respectively. Normal tissues are matched TCGA adjacent tissue and GTEx data. The method for differential analysis is one-way ANOVA. (B) Gene expression analysis across stages of the tyrosine catabolic genes in GSE89377 dataset. Violin plots represent $\log_2$(TPM+1) of genes in normal (grey, n= 13), early HCC (red, n=5), stage 1 HCC (red, n=9), stage 2 HCC (red, n=12) and stage 3 HCC (red, n=14). A t-test was used to compare the expression difference between tumor and normal tissue; $p < 0.05$ was considered statistically significant. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ based on the Student's t test. Values are mean ± SEM. TPM, transcript per million.

Figure 3: Overall survival outcomes of 364 HCC patients were analyzed using log-rank tests based on gene expression in HCC tissues from the TCGA cohort. Kaplan-Meier curves are plotted using GEPIA for TAT, HPD, HGD, GSTZ1 and FAH, and HRs and 95% confidence intervals are shown. Abbreviation: HCC, hepatocellular carcinoma, HRs, hazard ratios; TCGA, the Cancer Genome Atlas.

Figure 4: The protein expression profile of the tyrosine catabolic genes in the pan-cancer analysis (A) Quantification of HPD, HGD, GSTZ1 and FAH expression in IHC images obtained from HPA. A t-test was used to compare the expression difference between tumor and normal tissue adjacent to the tumor; $p < 0.05$ was considered statistically significant. *$p < 0.05$, ***$p < 0.001$ based on the Student's t test. Values are mean ± SEM. (B) Representative images of normal liver tissue and HCC tissue stained with antibody against GSTZ1.

Figure 5: Enrichment Map of GSTZ1 overexpressed huh7 and non-targeted control: GSEA was used to obtain canonical pathway gene sets that were visualized using the Enrichment Map plug-in for Cytoscape. Each node represents a gene set with similar nodes clustered together and connected by edges with the number of known interactors between the nodes being represented by the thickness of edges. The size of each node denotes the gene set size for each specific.

Figure 6: Prognostic value of microRNAs that target the tyrosine catabolic genes (A) The Venn diagram demonstrated the number of predicted miRNAs that target TAT, HPD, FAH and GSTZ1 from TargetScan database. (B) Survival analysis with miR-539 and miR-661 (KM Plotter dataset). The TCGA-LIHC dataset[43] from Kaplan-Meier Plotter[44] was used to test for survival prediction capacity of miR-539 in liver cancer. The CapitalBio miRNA Array liver dataset[42] was used to test for survival prediction capacity of miR-661 in liver cancer. Cox regression model was used for each gene to predict relapse-free survival. Samples are divided into Low (black) and High (red) expression groups for each gene. Hazard ratio (HR) and p value for each association are shown within each plot.

Table 1. Summary of mutations of tyrosine catabolic genes in patients with HCC

Gene	DNA change	Type	Consequences	SIFT Impact
TAT	chr16:g.71568080 G>C	Substitution	3 Prime UTR	N/A
	chr16:g.71570753_71570754insG	Insertion	Frameshift	N/A
	chr16:g.71576063G>T	Substitution	Intron	N/A
	chr16:g.71568109A>T	Substitution	3 Prime UTR	N/A
	chr16:g.71570812T>C	Substitution	Missense	Deleterious
	chr16:g.71572596A>G	Substitution	Synonymous	N/A
	chr16:g.71568283C>A	Substitution	Missense	Deleterious
HPD	chr12:g.121847089T>C	Substitution	Missense	Deleterious
	chr12:g.121849748T>C	Substitution	Missense	Deleterious
	chr12:g.121858824G>A	Substitution	5 Prime UTR	N/A
HGD	chr3:g.120646351T>A	Substitution	Missense	Deleterious
	chr3:g.120650834A>T	Substitution	Missense	Deleterious
	chr3:g.120670454C>A	Substitution	Missense	Deleterious
	chr3:g.120682178delTTCT	Deletion	5 Prime UTR	N/A
GSTZ1	chr14:g.77330329G>T	Substitution	Missense	N/A
FAH	chr15:g.80172237A>G	Substitution	Missense	Deleterious
	chr15:g.80160464G>T	Substitution	Splice Region	N/A
	chr15:g.80173063G>A	Substitution	Synonymous	N/A
	chr15:g.80186294G>T	Substitution	3 Prime UTR	N/A
	chr15:g.80186299G>A	Substitution	3 Prime UTR	N/A

	chr15:g.80162464C>T	Substitution	Intron	N/A

Table 2. Summary of CNAs of tyrosine catabolic genes in patients with HCC

Gene name	Location	Nearest peak	In peak ?	Q-value	Frequency of detection		
					Overall	Focal	High value
TAT	chr16:71600753-71610998	chr16:78129906-79627535	No	1	0.4108	0.0135	0
HPD	chr12:122277432-122326517	chr12:123453469-133155338	No	0.0191	0.1432	0.0486	0
HGD	chr3:120347014-120401418	chr3:114042610-115341566	No	1	0.1135	0.0081	0
GSTZ1	chr14:77787229-77797940	chr14:66969095-67653632	No	0.856	0.3405	0.0324	0.0054
FAH	chr15:80445232-80478924	chr15:88785838-101883952	No	1	0.1892	0.0189	0

Data Citations

1. Gene Expression Omnibus GSE89377 (2017)
2. Gene Expression Omnibus GSE117822 (2019)

References

1 Loeb, L. A. & Harris, C. C. Advances in chemical carcinogenesis: a historical review and prospective. *Cancer research* **68**, 6863-6872, doi:10.1158/0008-5472.Can-08-2852 (2008).

2 Warburg, O. On the origin of cancer cells. *Science (New York, N.Y.)* **123**, 309-314 (1956).

3 Iansante, V. *et al.* PARP14 promotes the Warburg effect in hepatocellular carcinoma by inhibiting JNK1-dependent PKM2 phosphorylation and activation. *Nature communications* **6**, 7882, doi:10.1038/ncomms8882 (2015).

4 Gong, L. *et al.* Reduced survival of patients with hepatocellular carcinoma expressing hexokinase II. *Medical oncology (Northwood, London, England)* **29**, 909-914, doi:10.1007/s12032-011-9841-z (2012).

5 DeWaal, D. *et al.* Hexokinase-2 depletion inhibits glycolysis and induces oxidative phosphorylation in hepatocellular carcinoma and sensitizes to metformin. *Nature communications* **9**, 446, doi:10.1038/s41467-017-02733-4 (2018).

6 Gan, L., Liu, Z. & Sun, C. Obesity linking to hepatocellular carcinoma: A global view. *Biochimica et biophysica acta. Reviews on cancer* **1869**, 97-102, doi:10.1016/j.bbcan.2017.12.006 (2018).

7 Li, X., Wang, X. & Gao, P. Diabetes Mellitus and Risk of Hepatocellular Carcinoma. *BioMed research international* **2017**, 5202684, doi:10.1155/2017/5202684 (2017).

8 Borrelli, A. *et al.* Role of gut microbiota and oxidative stress in the progression of non-alcoholic fatty liver disease to hepatocarcinoma: Current and innovative therapeutic approaches. *Redox biology* **15**, 467-479, doi:10.1016/j.redox.2018.01.009 (2018).

9 Calvisi, D. F. *et al.* Increased lipogenesis, induced by AKT-mTORC1-RPS6 signaling, promotes development of human hepatocellular carcinoma. *Gastroenterology* **140**, 1071-1083, doi:10.1053/j.gastro.2010.12.006 (2011).

10 Na, T. Y. *et al.* Liver X receptor mediates hepatitis B virus X protein-induced lipogenesis in hepatitis B virus-associated hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* **49**, 1122-1131, doi:10.1002/hep.22740 (2009).

11 Liou, G. Y. & Storz, P. Reactive oxygen species in cancer. *Free radical research* **44**, 479-496, doi:10.3109/10715761003667554 (2010).

12 Anastasiou, D. *et al.* Inhibition of pyruvate kinase M2 by reactive oxygen species contributes to cellular antioxidant responses. *Science (New York, N.Y.)* **334**, 1278-1283, doi:10.1126/science.1211485 (2011).

13 Fitian, A. I. & Cabrera, R. Disease monitoring of hepatocellular carcinoma through metabolomics. *World journal of hepatology* **9**, 1-17, doi:10.4254/wjh.v9.i1.1 (2017).

14 Ipson, B. R. & Fisher, A. L. Roles of the tyrosine isomers meta-tyrosine and ortho-tyrosine in oxidative stress. *Ageing research reviews* **27**, 93-107, doi:10.1016/j.arr.2016.03.005 (2016).

15 Herman, S. *et al.* Alterations in the tyrosine and phenylalanine pathways revealed by biochemical profiling in cerebrospinal fluid of Huntington's disease subjects. *Scientific reports* **9**, 4129, doi:10.1038/s41598-019-40186-5 (2019).

16 Lai, H. S., Lee, J. C., Lee, P. H., Wang, S. T. & Chen, W. J. Plasma free amino acid profile in cancer patients. *Seminars in cancer biology* **15**, 267-276, doi:10.1016/j.semcancer.2005.04.003 (2005).

17 Wiggins, T., Kumar, S., Markar, S. R., Antonowicz, S. & Hanna, G. B. Tyrosine, phenylalanine, and tryptophan in gastroesophageal malignancy: a systematic review. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **24**, 32-38, doi:10.1158/1055-9965.Epi-14-0980 (2015).

18 Kim, S. Z. *et al.* Hepatocellular carcinoma despite long-term survival in chronic tyrosinaemia I. *Journal of inherited metabolic disease* **23**, 791-804 (2000).

19 Russo, P. & O'Regan, S. Visceral pathology of hereditary tyrosinemia type I. *American journal of human genetics* **47**, 317-324 (1990).

20 Watanabe, A., Higashi, T., Sakata, T. & Nagashima, H. Serum amino acid levels in patients with hepatocellular carcinoma. *Cancer* **54**, 1875-1882, doi:10.1002/1097-0142(19841101)54:9<1875::aid-cncr2820540918>3.0.co;2-o (1984).

21 Hirayama, C., Suyama, K., Horie, Y., Tanimoto, K. & Kato, S. Plasma amino acid patterns in hepatocellular carcinoma. *Biochem Med Metab Biol* **38**, 127-133, doi:10.1016/0885-4505(87)90071-5 (1987).

22 Fu, L. *et al.* Down-regulation of tyrosine aminotransferase at a frequently deleted region 16q22 contributes to the pathogenesis of hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* **51**, 1624-1634, doi:10.1002/hep.23540 (2010).

23 Chasseaud, L. F. The role of glutathione and glutathione S-transferases in the metabolism of chemical carcinogens and other electrophilic agents. *Advances in cancer research* **29**, 175-274 (1979).

24 Lee, W. H. *et al.* Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 11733-11737 (1994).

25 Niu, D., Zhang, J., Ren, Y., Feng, H. & Chen, W. N. HBx genotype D represses GSTP1 expression and increases the oxidative level and apoptosis in HepG2 cells. *Molecular oncology* **3**, 67-76, doi:10.1016/j.molonc.2008.10.002 (2009).

26 Zhang, Y. J. *et al.* Silencing of glutathione S-transferase P1 by promoter hypermethylation and its relationship to environmental chemical carcinogens in hepatocellular carcinoma. *Cancer letters* **221**, 135-143, doi:10.1016/j.canlet.2004.08.028 (2005).

27 Cantor, K. P. *et al.* Polymorphisms in GSTT1, GSTZ1, and CYP2E1, disinfection by-products, and risk of bladder cancer in Spain. *Environmental health perspectives* **118**, 1545-1550, doi:10.1289/ehp.1002206 (2010).

28 Zhang, J. *et al.* A panel of three oxidative stress-related genes predicts overall survival in ovarian cancer patients received platinum-based chemotherapy. *Aging* **10**, 1366-1379, doi:10.18632/aging.101473 (2018).

29 Rhodes, D. R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia (New York, N.Y.)* **6**, 1-6 (2004).

30 Tang, Z. *et al.* GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic acids research* **45**, W98-w102, doi:10.1093/nar/gkx247 (2017).

31 Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics* **45**, 1113-1120, doi:10.1038/ng.2764 (2013).

32 R: A Language and Environment for Statistical Computing v. R 3.5.3 (2019).

33 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

34 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

35 Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PloS one* **5**, e13984, doi:10.1371/journal.pone.0013984 (2010).

36 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

37 Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* **2**, 401-404, doi:10.1158/2159-8290.Cd-12-0095 (2012).

38 Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology* **12**, R41, doi:10.1186/gb-2011-12-4-r41 (2011).

39 Garcia, D. M. *et al.* Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nature structural & molecular biology* **18**, 1139-1146, doi:10.1038/nsmb.2115 (2011).

40 Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic acids research* **42**, D92-97, doi:10.1093/nar/gkt1248 (2014).

41 Lanczky, A. *et al.* miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. *Breast cancer research and treatment* **160**, 439-446, doi:10.1007/s10549-016-4013-7 (2016).

42 Li, W. *et al.* Diagnostic and prognostic implications of microRNAs in human hepatocellular carcinoma. *International journal of cancer* **123**, 1616-1622, doi:10.1002/ijc.23693 (2008).

43 Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **169**, 1327-1341.e1323, doi:10.1016/j.cell.2017.05.046 (2017).

44 Nagy, A., Lanczky, A., Menyhart, O. & Gyorffy, B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Scientific reports* **8**, 9227, doi:10.1038/s41598-018-27521-y (2018).

45 Piao, Z., Park, C., Kim, J. J. & Kim, H. Deletion mapping of chromosome 16q in hepatocellular carcinoma. *British journal of cancer* **80**, 850-854, doi:10.1038/sj.bjc.6690431 (1999).

46 Kanai, Y. *et al.* Aberrant DNA methylation on chromosome 16 is an early event in hepatocarcinogenesis. *Japanese journal of cancer research : Gann* **87**, 1210-1217 (1996).

47 Jahn, S. C. *et al.* GSTZ1 expression and chloride concentrations modulate sensitivity of cancer cells to dichloroacetate. *Biochimica et biophysica acta* **1860**, 1202-1210, doi:10.1016/j.bbagen.2016.01.024 (2016).

48 James, M. O. *et al.* Therapeutic applications of dichloroacetate and the role of glutathione transferase zeta-1. *Pharmacology & therapeutics* **170**, 166-180, doi:10.1016/j.pharmthera.2016.10.018 (2017).

49 Zhao, J., Xu, J. & Zhang, R. MicroRNA-539 inhibits colorectal cancer progression by directly targeting SOX4. *Oncology letters* **16**, 2693-2700, doi:10.3892/ol.2018.8892 (2018).

50 Zhang, H. *et al.* miR-539 inhibits prostate cancer progression by directly targeting SPAG5. *Journal of experimental & clinical cancer research : CR* **35**, 60, doi:10.1186/s13046-016-0337-8 (2016).

51 Liu, Y. *et al.* miR-539 inhibits FSCN1 expression and suppresses hepatocellular carcinoma migration and invasion. *Oncology reports* **37**, 2593-2602, doi:10.3892/or.2017.5549 (2017).

52 Zhu, C., Zhou, R., Zhou, Q., Chang, Y. & Jiang, M. microRNA-539 suppresses tumor growth and tumorigenesis and overcomes arsenic trioxide resistance in hepatocellular carcinoma. *Life sciences* **166**, 34-40, doi:10.1016/j.lfs.2016.10.002 (2016).

53 Gomez de Cedron, M. *et al.* MicroRNA-661 modulates redox and metabolic homeostasis in colon cancer. *Molecular oncology* **11**, 1768-1787, doi:10.1002/1878-0261.12142 (2017).

54 Liu, F. *et al.* MiR-661 promotes tumor invasion and metastasis by directly inhibiting RB1 in non small cell lung cancer. *Molecular cancer* **16**, 122, doi:10.1186/s12943-017-0698-4 (2017).

55 Zhu, T. *et al.* MiR-661 contributed to cell proliferation of human ovarian cancer cells by repressing INPP5J expression. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **75**, 123-128, doi:10.1016/j.biopha.2015.07.023 (2015).

56 Tanguay, R. M., Jorquera, R., Poudrier, J. & St-Louis, M. Tyrosine and its catabolites: from disease to cancer. *Acta biochimica Polonica* **43**, 209-216 (1996).

57 Cheng, J., Zheng, G., Jin, H. & Gao, X. Towards Tyrosine Metabolism in Esophageal Squamous Cell Carcinoma. *Combinatorial chemistry & high throughput screening* **20**, 133-139, doi:10.2174/1386207319666161220115409 (2017).

58 Zhang, J. *et al.* Esophageal cancer metabolite biomarkers detected by LC-MS and NMR methods. *PloS one* **7**, e30181, doi:10.1371/journal.pone.0030181 (2012).

59 Cheng, Z. *et al.* Conversion of hepatoma cells to hepatocyte-like cells by defined hepatocyte nuclear factors. *Cell research* **29**, 124-135, doi:10.1038/s41422-018-0111-x (2019).

60 Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic acids research* **41**, D991-D995, doi:10.1093/nar/gks1193 (2013).

61 Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846-1847, doi:10.1093/bioinformatics/btm254 (2007).

62 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47-e47, doi:10.1093/nar/gkv007 (2015).

63 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science (New York, N.Y.)* **347**, 1260419, doi:10.1126/science.1260419 (2015).

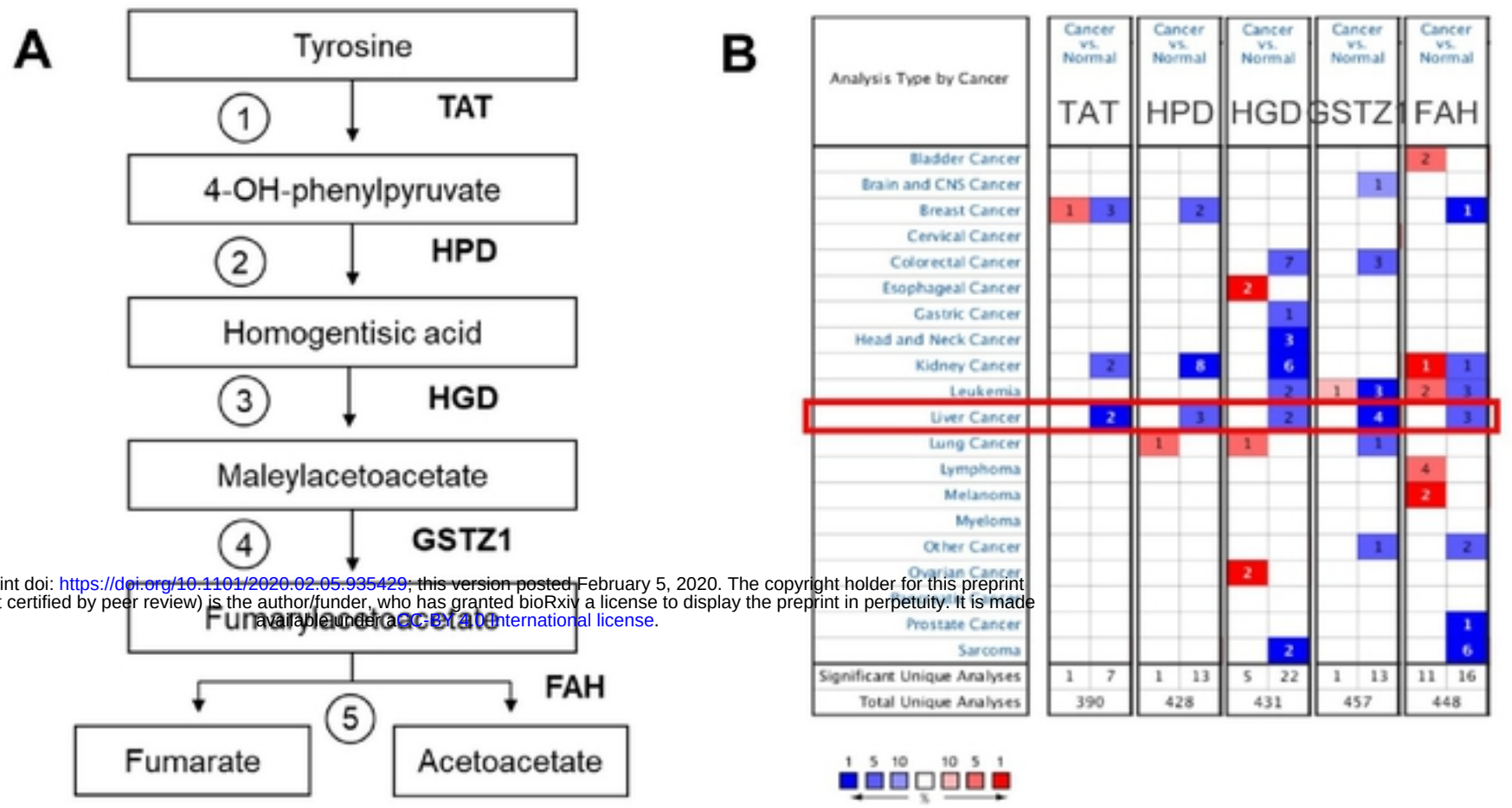

Figure 1: Downregulation of the tyrosine catabolic genes in several types of cancer, including HCC. (A) Graphics of tyrosine catabolism process. (B) The mRNA expression levels of the tyrosine catabolic genes according to Oncomine database. The mRNA expression of the genes (cancer versus normal tissue) in pan-cancers analyzed with the Oncomine database. The graphic demonstrates the numbers of datasets that meet our threshold in each cancer type. Cell color was defined as the gene rank percentile in the study. (C) The heat map indicates the expression after normalization by TPM+1 for comparison between tumor (T) and normal (N) across cancer types. Normal tissues are matched TCGA adjacent tissue and GTEx data. The cancer abbreviation names are shown according to TCGA study abbreviations (Supplementary Table S1). TPM, transcript per million.

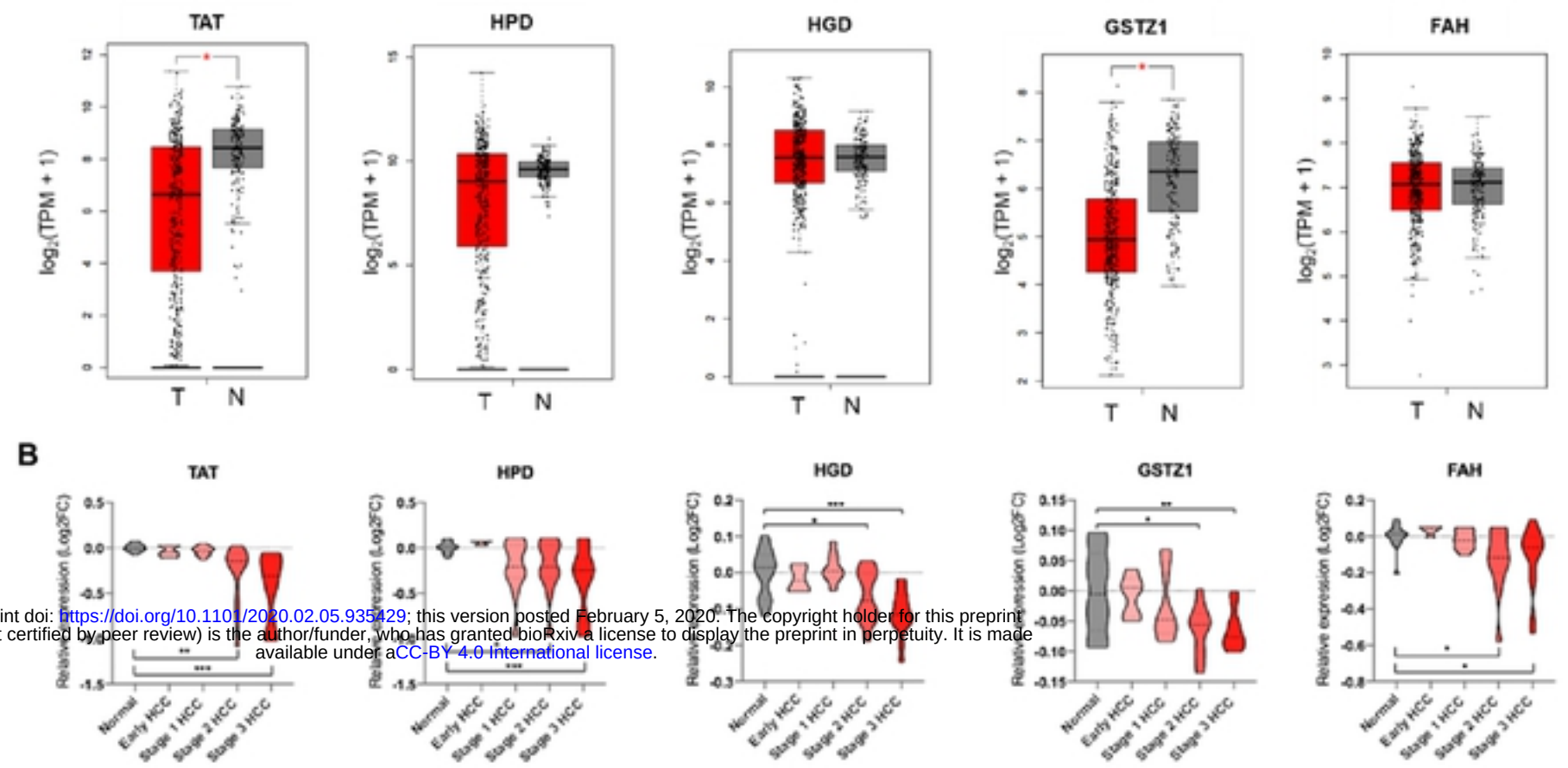

Figure 2: Gene expression profile of the tyrosine catabolic genes in HCC. (A) Gene expression analysis of tyrosine catabolic genes using GEPIA based on the TCGA and GTEx database. Box plots represent the gene expression level in terms of $\log_2$(TPM+1) in the tumor (red, n=369) and normal (grey, n=160) samples, respectively. Normal tissues are matched TCGA adjacent tissue and GTEx data. The method for differential analysis is one-way ANOVA. (B) Gene expression analysis across stages of the tyrosine catabolic genes in GSE89377 dataset. Violin plots represent $\log_2$(TPM+1) of genes in normal (grey, n= 13), early HCC (red, n=5), stage 1 HCC (red, n=9), stage 2 HCC (red, n=12) and stage 3 HCC (red, n=14). A t-test was used to compare the expression difference between tumor and normal tissue; $p < 0.05$ was considered statistically significant. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ based on the Student's t test. Values are mean ± SEM. TPM, transcript per million.

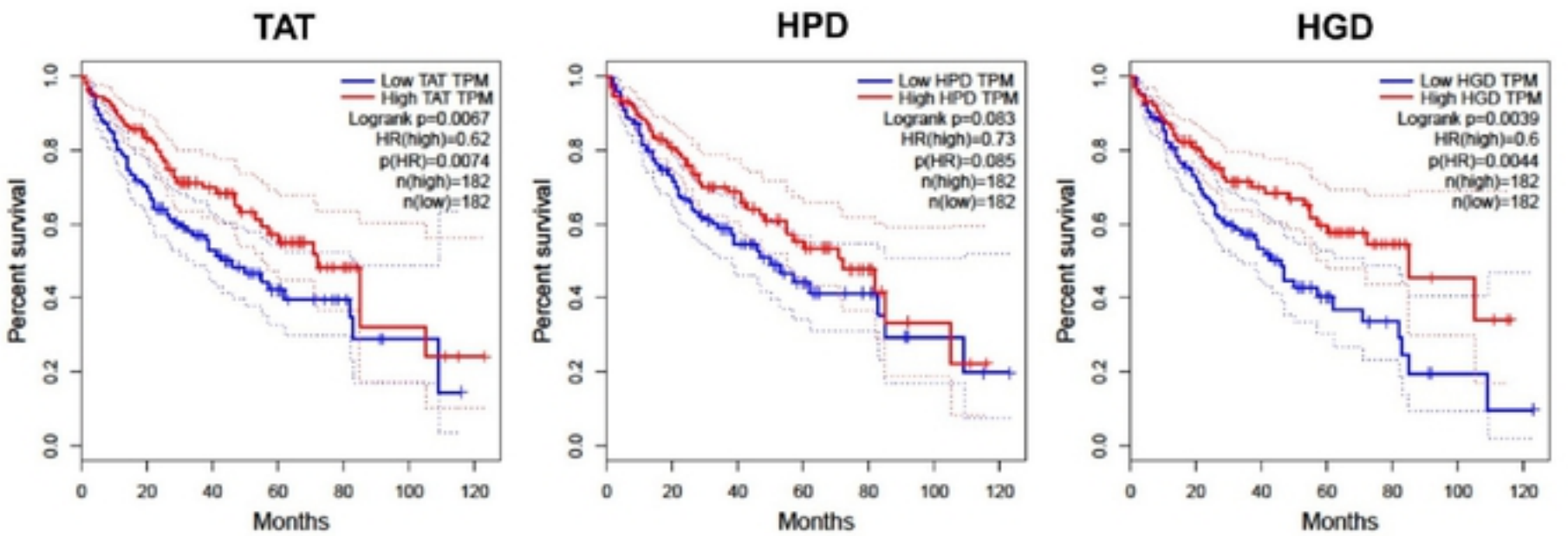

Figure 3: Overall survival outcomes of 364 HCC patients were analyzed using log-rank tests based on gene expression in HCC tissues from the TCGA cohort. Kaplan-Meier curves are plotted using GEPIA for TAT, HPD, HGD, GSTZ1 and FAH, and HRs and 95% confidence intervals are shown. Abbreviation: HCC, hepatocellular carcinoma, HRs, hazard ratios; TCGA, the Cancer Genome Atlas.

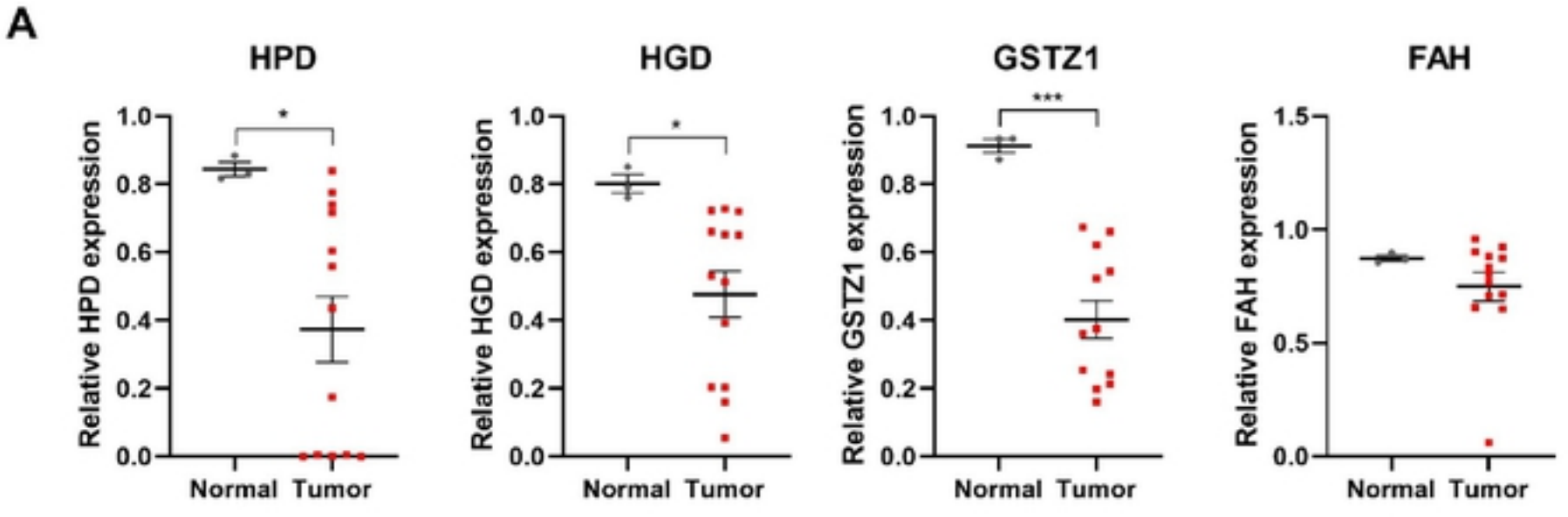

Figure 4: The protein expression profile of the tyrosine catabolic genes in the pan-cancer analysis (A) Quantification of HPD, HGD, GSTZ1 and FAH expression in IHC images obtained from HPA. A t-test was used to compare the expression difference between tumor and normal tissue adjacent to the tumor; $p < 0.05$ was considered statistically significant. $*p < 0.05$, $***p < 0.001$ based on the Student's t test. Values are mean ± SEM. (B) Representative images of normal liver tissue and HCC tissue stained with antibody against GSTZ1.

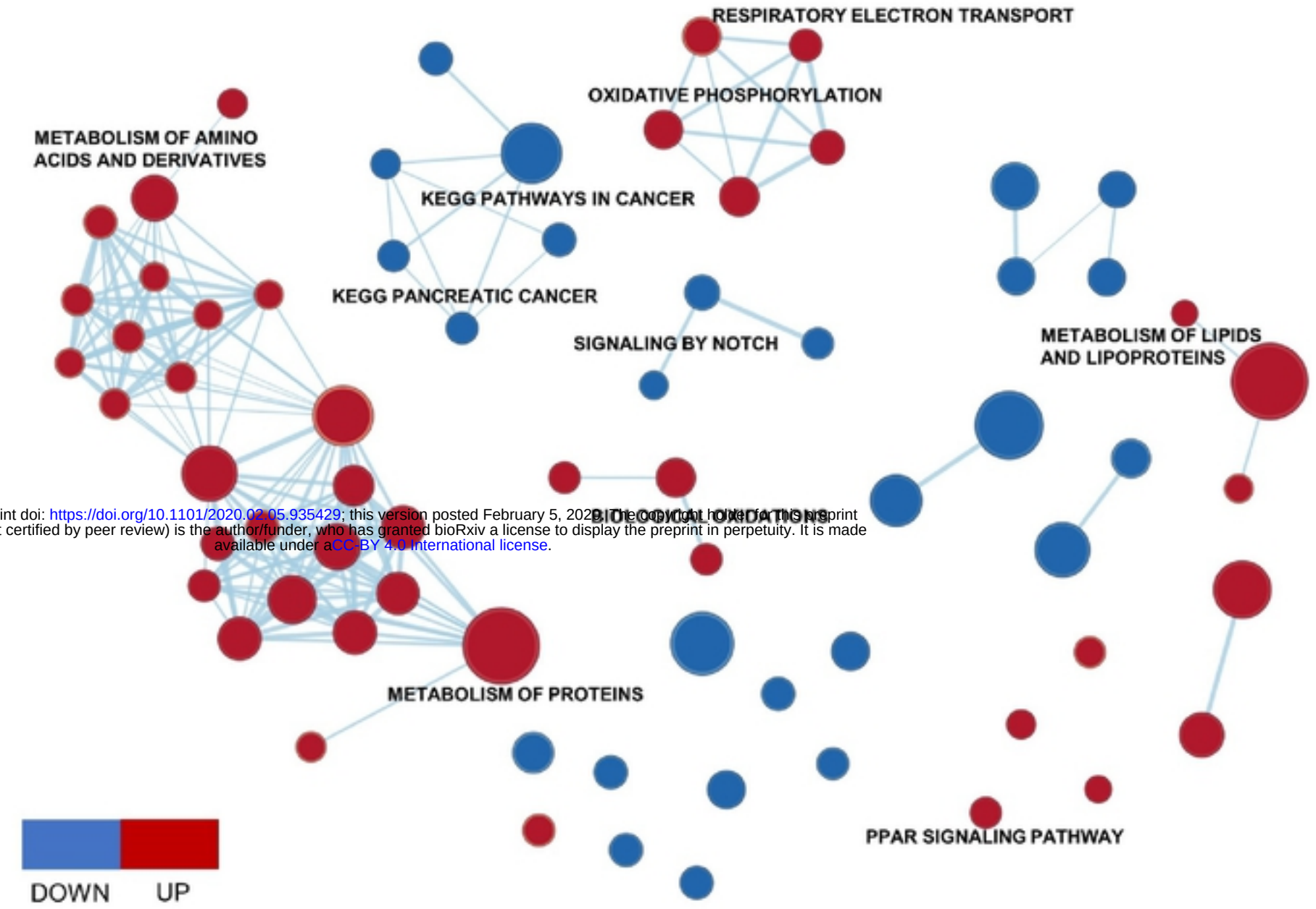

Figure 5: Enrichment Map of GSTZ1 overexpressed huh7 and non-targeted control: GSEA was used to obtain canonical pathway gene sets that were visualized using the Enrichment Map plug-in for Cytoscape. Each node represents a gene set with similar nodes clustered together and connected by edges with the number of known interactors between the nodes being represented by the thickness of edges. The size of each node denotes the gene set size for each specific.

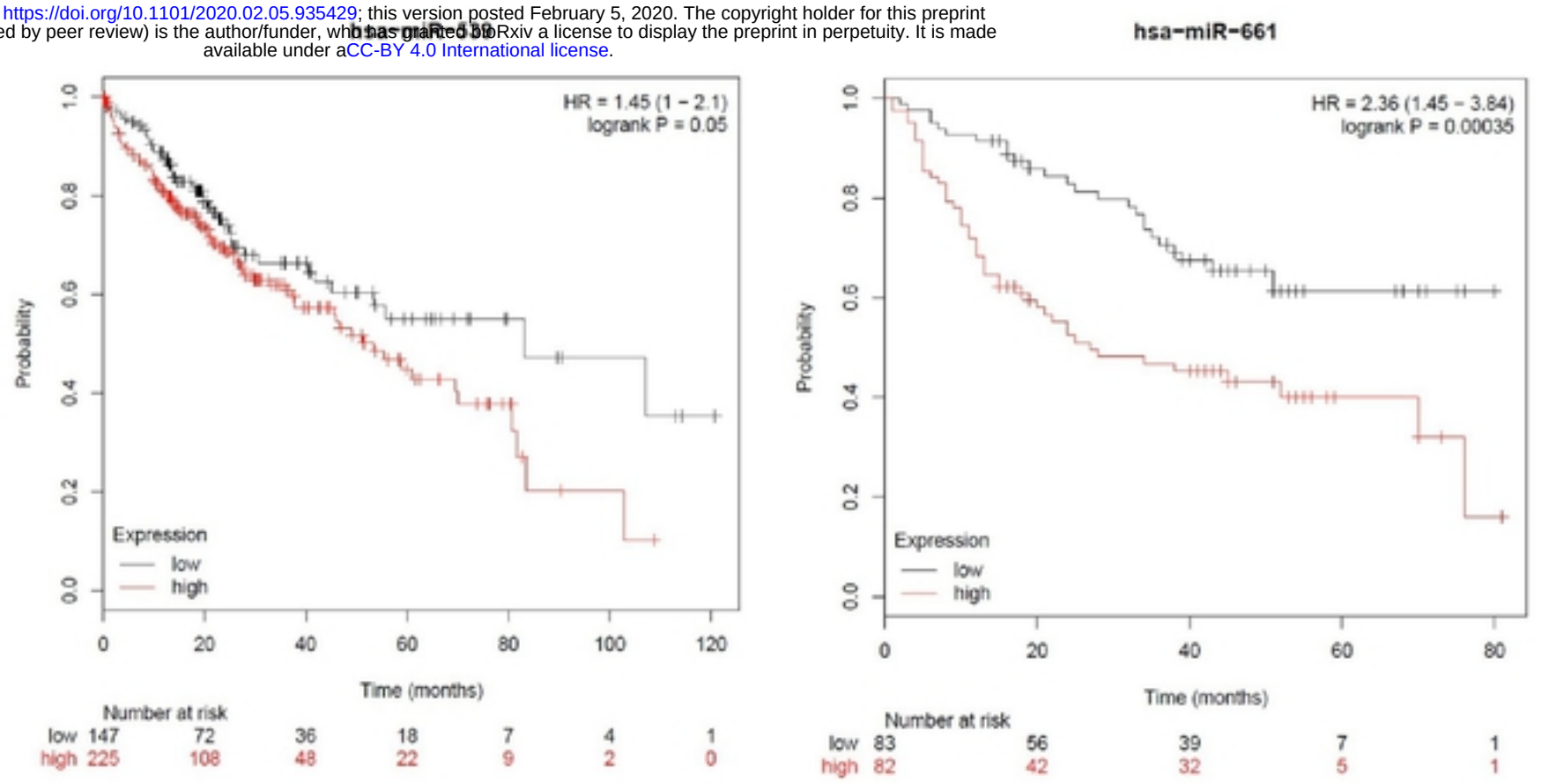

Figure 6: Prognostic value of microRNAs that target the tyrosine catabolic genes (A) The Venn diagram demonstrated the number of predicted miRNAs that target TAT, HPD, FAH and GSTZ1 from TargetScan database. (B) Survival analysis with miR-539 and miR-661 (KM Plotter dataset). The TCGA-LIHC dataset[43] from Kaplan-Meier Plotter[44] was used to test for survival prediction capacity of miR-539 in liver cancer. The CapitalBio miRNA Array liver dataset[42] was used to test for survival prediction capacity of miR-661 in liver cancer. Cox regression model was used for each gene to predict relapse-free survival. Samples are divided into Low (black) and High (red) expression groups for each gene. Hazard ratio (HR) and p value for each association are shown within each plot.