Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Excavating thyroid carcinoma risk metabolic genes associated with biological clock by bioinformatics analysis: A study based on TCGA and GEO data

wendong wang
Hebei North University
Jiali Zhang
Hebei North University
Zhengjiu Liu
Hebei North University
Haoting Zhang
Hebei North University
Linrui Ren
Hebei North University
Xiaobo Zhu
Hebei North University
Xiaotong Chang (changxt1212@vip.sina.com)

Research article

Keywords: Risk gene, Biological rhythm, Chromosomal location, Enrichment analyses, KEGG pathways

Posted Date: February 5th, 2020

DOI: https://doi.org/10.21203/rs.2.22670/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Background

Thyroid carcinoma (TC) is malignant tumor of the endocrine gland with a significantly increased incidence in recent years, which related to metabolic diseases and biological clock disorder. The molecular mechanism and risk metabolic genes associated with biological clock of TC require conducting study.

Methods

We mined differentially expressed genes of TC (TC-DEGs) from TCGA and GEO database. Searching for biorhythm related genes (BRGs) based on core clock genes by STRING analysis. Next, we used a series of bioinformatics methods to analyze the TC-DEGs and BRGs, including chromosome location, GO and KEGG pathway annotation. Then, we extracted the expression of statistically significant TC metabolic genes by Perl program combined with a computational method of GSEA. Furthermore, screen the risk metabolic genes and reveal potential underlying KEGG pathways of the gene signatures.

Results

Our results identified 474 DEGs ($P < 0.05$; $|\log FC| > 1$) in patients with TC compared with the normal controls using the available numerical mRNA expression values. The chromosomal assignments of TC-DEGs and BRGs were inhomogeneous. These genes were mainly distributed on Chrom01, 02, 11, 12, 04, 03 and X, none of the associated genes were distributed on sex Y chromosome. ARNTL and BHLHE40 belong to both the TC-DEGs and BRGs. PER2, NCOR1, RXRG, CCND1, SERPINE1 and PLAU were mined which act as the hub of signaling pathways linking “Thyroid hormone signaling pathway”, “Circadian rhythm”, “Transcriptional misregulation in cancer” and so on. One hundred and forty-one metabolic genes were extracted to train the prognostic model, and 11 risk genes were found, among them, 10 risk gene products have indirect interaction with ARNTL and BHLHE40. Such important genes for TC were regulated by many common conserved microRNAs.

Conclusions

Our study identified 8 key genes and 11 risk metabolic genes of thyroid carcinoma associated with biological clock by bioinformatics analysis. And the coding products of these genes form a clear interaction network, which is conducive to further study the molecular mechanism, reveal the role of circadian clock genes and metabolic genes in thyroid cancer.

Background

Thyroid carcinoma (TC) is malignant tumor of the endocrine gland, and originates from follicular or parafollicular cells. Compared to other cancers, TC is relatively uncommon, which accounts for about 1% of systemic malignant tumors [1]. However, the incidence of TC has increased significantly in China in

recent years, which is more than three times that of 10 years ago. The incidence of thyroid carcinoma is related to have a history of exposure to high doses of radiation, family history of thyroid cancer, metabolic disorder, biological rhythm disorder and so on. However, for most patients, the cause of TC is unknown [2]. Thyroid carcinoma can occur at any age, but which incidence increases above 40 years old. Moreover, women (who are most often in their 40 s or 50 s when diagnosed) are at a higher risk of thyroid carcinoma than men (who are usually in their 60 s or 70 s) [3]. In general, TC has few obvious symptoms and is usually found by neck examination. Such cancer has a tendency of family aggregation, but its genetic basis is not totally clear [4]. At present, TC is mainly divided into four categories, namely papillary thyroid cancer (PTC), follicular thyroid cancer (FTC), medullary thyroid cancer (MTC) and anaplastic thyroid cancer (ATC). PTC is the most common type, which accounting for about 75% of thyroid carcinoma [5].

Metabolic reprogramming is emerging to be a novel symbol of cancer [6]. Thyroid carcinoma is associated with metabolic disorder, because the investigation shows that the incidence of this cancer is the same as that of metabolic syndrome. Glucose metabolism, insulin resistance, hypertension and high BMI significantly increase the risk of thyroid cancer [7]. Besides, biological rhythm disorder can affect metabolism, the biological clock genes are widely involved in the regulation of a variety of metabolic processes [8]. Therefore, Thyroid carcinoma is inextricably related to biological clock disorder and metabolic diseases.

In recent years, with the development of gene chip technologies and sequencing technologies, the biological database has been gradually improved, integrating biological clock genes, metabolic signature and prognosis-related gene signature analysis, which have shown a great advantage for revealing the molecular mechanism of thyroid cancer. Thyroid carcinoma is rapidly increasing in incidence worldwide, so which become a focus of medical research. However, the molecular mechanism of TC remains far from clear. In order to further study the molecular mechanism, reveal the role of circadian clock genes and metabolic genes in thyroid cancer, we conduct the study.

Materials And Methods

Data collection

The messenger RNA (mRNA) expression profiles and corresponding clinical information were obtained from The Cancer Genome Atlas (TCGA) official website (https://portal.gdc.cancer.gov/) for the Thyroid Carcinoma projects data set (TCGA-THCA) including 510 TC, 58 control thyroid gland samples and 507 clinical data. The GSE33630 [9] mRNA expression dataset was downloaded from the Gene Expression Omnibus (GEO) database (www.ncbi.nlm.nih.gov/geo). In the GSE33630 dataset, 60 TC and 45 normal thyroid gland controls were included and analyzed using the Affymetrix Human Genome U133 Plus 2.0 Array.

TC-DEGs And BRGs Mining

The pretreatment of original expression datasets, including background correction, normalization and log2 transformation were employed for the expression profiles using the R affy package (www.bioconductor.org/packages/release/bioc/html/affy.html). Furthermore, the TC-DEGs was identified by the Limma version 3. 6.1 R package (www.bioconductor.org/packages/release/bioc/html/limma.html) [10], using an adjusted P-value < 0.05 and a |log fold change (FC)| >1 as criteria. There are 12 core clock genes, CLOCK, CSNK1E, CRY1, CRY2, PER1, PER2, PER3, NPAS2, ARNTL, RORA, NR1D1 and NR1D2 in mammals [11]. Centre on these core clock genes, biorhythm related genes (BRGs) were mined by literature reviewing, NCBI database retrieval and Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; https://string-db.org/) analysis, which is a database of known and predicted protein-protein interactions [12].

Genes Chromosomal Location

The locations of TC-DEGs and BRGs on the chromosomes were identified by the UCSC Genome Browser Gateway (http://genome.ucsc.edu/cgi-bin/hgGateway), which provides a wide variety of genome mapping and sequencing data [13]. Then, the visualization of chromosome location of these genes was established by MapGene2Chrom web v2 online analytical tool (http://mg2c.iask.in/mg2c_v2.0) [14].

GO And KEGG Pathway Analysis

Enriched Gene Ontology[15] (GO, which is a commonly used approach for functional studies of large-scale transcriptomic data) terms of TC-DEGs and BRGs were acquired by the Database for Visualization, Annotation and Integrated Analysis [16] (DAVID, https://david.ncifcrf.gov). Furthermore, the enriched KEGG (Kyoto Encyclopedia of Genes and Genomes, https://www.kegg.jp/kegg/pathway.html [17]) pathways for these genes were identified by R " pathview" package, P-value < 0.05 and Count ≥ 10 genes as criteria.

Extraction Of The Expression Of Metabolic Genes

Download the dataset "KEGG gene sets, gene symbols" from Gene Set Enrichment Analysis (GSEA, http://software.broadinstitute.org/gsea/msigdb/index.jsp) [18–19], which as a computational method combined with Perl program to extract the expression of statistically significant metabolic genes between TC and control thyroid gland samples.

Identification the high-risk metabolic genes

Univariate Cox regression analysis [20] ($p < 0.01$ as criteria, were considered statistically significant) were performed to identify the high-risk metabolic genes. The R package “survival” [21] was used to explore the optimal cut-off of risk score. Then, The R package “limma” and “beeswarm” [22] were used for analysis the single gene differential expression of high-risk metabolic genes. The prognostic gene signature was shown as risk score = (CoefficientmRNA1 × expression of mRNA1) + (CoefficientmRNA2 × expression of mRNA2) +⋯+ (CoefficientmRNAn × expression mRNAn) [23]. The R package “survminer” was used to identify the optimal cut-off value to divide patients into high- risk and low-risk group.

Protein-protein Interactions And Conserved MicroRNAs Of Associated Genes

Protein-protein of genes of thyroid carcinoma associated with biological clock by were screened by STRING. Furthermore, cytoscape version 3.6.1 software [24] was utilized to establish the interaction network. The conserved microRNAs of the key genes and risk genes of TC-DEGs and BRGs were mined using Targetscan analysis online tool (http://www.targetscan.org/vert /) [25].

Gene Set Enrichment Analyses

To reveal potential underlying Kyoto Encyclopedia of and Genomes (KEGG) pathways of the gene signature, GSEA were used to find enriched terms in C2 in TCGA-THCA cohort. $p < 0.05$ and FDR (a false discovery rate) $q < 0.25$ were as criteria, which were considered statistically significant.

Results

A total of 474 TC-DEGs and 64 BRGs are identified

Four hundred and seventy-four DEGs were identified ($P < 0.05$; $|\log FC| > 1$) in patients with thyroid carcinoma compared with the normal controls using the available numerical mRNA expression values, including 295 up-regulated genes and 179 down regulated genes (Fig. 1). The 64 BRGs ((including 12 core clock genes) were mined by literature reviewing, NCBI database retrieval and STRING, which are presented in Table 1.

Table 1

The 64 biorhythm related genes

Gene. symbol	Entrez ID	Gene. symbol	Entrez ID	Gene. symbol	Entrez ID
ARNTL●	406	CHEK11	1111	DEC1	50514
ARNTL2	56938	CIPC	85457	DLAT	1737
ATR	545	CLOCK●	9575	DLD	1738
BATF2	116071	CLSPN	63967	EP300	2033
BHLHE40	8553	CRY1●	1407	EPAS1	2034
BHLHE41	79365	CRY2●	1408	GRP	2922
BLM	641	CSNK1D	1453	GSK3B	2932
BTRC	8945	CSNK1E●	1454	HDAC3	8841
CDC45	8318	DBP	1628	HIF1A	3091
CEBPA	1050	DDIT3	1649	KPTAP23-1	3123
SERPINE1	5054	SIRT1	23411	TIMELESS	8914
NRIP1	8204	PGC	5225	PDPR	55066
PDHA1	5160	PHLPP1	23239	PDXK	8566
PDHB	5162	POLA1	5422	PER1●	5187
PDHX	8050	PPM1A	5494	PER2●	8864
PDP1	54704	PPTC7	160760	PER3●	8863
MCM2	4171	NFIL3	4783	TIPIN	54962
MCM3	4172	NONO	4841	RIPK4	54101
MCM7	4176	NPAS2●	4862	ROR	100885779
MED1	5469	NR1D1●	9572	RORA●	6095
NCOR1	9611	NR1D2●	9975	RORB	6096
RORC	6097				

Entrez ID is the number of genes in Entrez database of NCBI. The core genes of biological clock are marked with black dot superscript.

TC-DEGs And BRGs Chromosomal Mapping

The chromosomal locations of 474 TC-DEGs and 64 BRGs were identified using the UCSC Genome Browser Gateway resources, and MapGene2Chrom web v2 online analytical tool was used to establish the chromosomal location map (Fig. 2). The chromosomal assignments of these genes were inhomogeneous. These genes were mainly distributed on Chrom01, 02, 11, 12, 04, 03 and X, none of the associated genes were distributed on sex Y chromosomes.

GO and KEGG pathway analysis

Using DAVID, Eighteen most significantly enriched GO process terms (FDR ≤ 0.05 and Count ≥ 5 genes as criteria) were obtained for TC-DEGs and BRGs, including “circadian rhythm”, “extracellular space”, “extracellular region”, “circadian regulation of gene expression”, “extracellular exosome”, “regulation of circadian rhythm”, “oxidation-reduction process”, “proteinaceous extracellular matrix”, “retinol metabolic process”, “arachidonic acid secretion”, “regulation of acetyl-CoA biosynthetic process from pyruvate”, “nuclear hormone receptor binding”, “pyruvate metabolic process”, “pyruvate dehydrogenase complex”, “negative regulation of glucocorticoid receptor signaling pathway”, “positive regulation of gene expression”, “response to drug” and “phospholipase A2 activity”, which are presented in Fig. 3A. KEGG pathway analysis showed that 18 significantly enriched pathways (P-value ≥ 0.05 and Count ≥ 10)for TC-DEGs and BRGs, including “Purine metabolism”, “Retinol metabolism”, “Circadian rhythm”, “Metabolism of xenobiotics by cytochrome P450”, “Transcriptional misregulation in cancer”, “Glycerophospholipid metabolism”, “Arachidonic acid metabolism”, “Drug metabolism - cytochrome P450”, “Chemical carcinogenesis”, “Pancreatic secretion”, “Tyrosine metabolism”, “Ether lipid metabolism”, “Glycolysis/Gluconeogenesis”, “Vascular smooth muscle contraction”, “Thyroid hormone signaling pathway”, “Complement and coagulation cascades”, “Tryptophan metabolism” and “p53 signaling pathway”. ARNTL and BHLHE40 belong to both the TC-DEGs and BRGs, and ARNTL gene is one of the core clock genes. PER2 is one of the core clock genes, and it also exists in the signal pathway of “Transcriptional misregulation in cancer”. NCOR1 and RXRG are the key links between “Transcriptional misregulation in cancer” and “Thyroid hormone signaling pathway”. CCND1 links “Thyroid hormone signaling pathway” with “p53 signaling pathway”. SERPINE1 links “p53 signaling pathway” with “Complement and coagulation cascades”. And, PLAU is the key link between “Transcriptional misregulation in cancer” and “Complement and coagulation cascades” (Fig. 3B). Our study identified 8 key genes of thyroid carcinoma associated with biological clock.

Construction The Prognostic Metabolic Gene Signature In TCGA

The 141 metabolic genes (77 up-regulated and 64 down-regulated) were included in TCGA-THCA and GSE33630 to train the prognostic model (Table 2). Univariate Cox regression model found 11 risk genes

($p < 0.01$), which are ITPKA, FBP2, AOX1, PLA2G7, GPAT3, HAGHL, PFKFB1, GGCT, PGK1, PAPSS2 and SGPP2 (Fig. 4A). GGCT (Fig. 4B), HAGHL (Fig. 4C) and ITPKA (Fig. 4D) are highly expressed in tumor cells ($p < 0.01$). PLA2G7 (Fig. 4E), GPAT3 (Fig. 4F), PAPSS2 (Fig. 4G), AOX1 (Fig. 4H), FBP2 (Fig. 4I), SGPP2 (Fig. 4J), PGK1 (Fig. 4K) and PFKFB1 (Fig. 4L) are lower expressed in tumor cells compared with the normal controls ($p < 0.01$).

Table 2

The 141 metabolic genes (77 up-regulated and 64 down-regulated).

Differential expression	Gene. symbol				
77 up-regulated genes	ADCY8	AK1	ALOX5	UPP1	PC
	PLA2G2E	PDE6A	CYP26A1	PNP	ISYNA1
	PLA2G12B	MTHFD1L	RDH5	PLA2G10	HPGDS
	ACOT12	GUCY1B1	PNMT	DPYS	PLA2G5
	ALDH3B2	SPHK1	GSTA1	CDA	CYP26C1
	NT5C1A	ACER2	LPL	P4HA2	ETNK2
	AWAT2	GPT	ALDH1A3	PDE1C	NT5E
	CYP2S1	CHST13	CA7	HDC	HAGHL
	PLCD3	ME3	CA12	ALOX12B	TDO2
	AOC1	TH	ALDH3B1	PLPP2	CA2
	ENTPD2	CYP4A22	ENTPD1	ITPKA	PCYT1B
	GGT6	AMY2A	ENTPD8	DHRS3	PLA2G2F
	GALE	MBOAT2	PDE5A	SULT2B1	B4GALT6
	ALOX15B	RENBP	PLA2G2C	PDE9A	CHIT1
	CYP1B1	CBR3	GGCT	CPT1C	GLS2
	DHDH	CHPT1			
64 down-regulated genes	GPAT3	TPH2	SGPP2	ACACB	TPO
	IDO1	DDC	LIPC	FMO1	IDO2
	CP	FBP2	CA4	RDH12	FMO2
	GBA3	GSTM5	PFKFB2	PFKFB1	ADH1A
	ACSM5	ACSM2A	AK7	ENPP6	HGD
	HAL	PAPSS2	CYP1A1	GYS2	AGPAT4
	GSTM3	DGKI	UGT8	PLA2G4E	PDE11A
	GATM	OTC	CD38	ALDH1A2	BCO1
	ADH6	ADH1C	NOS1	GLDC	ACSL6
	IMPA2	AOX1	PLA2G7	PLA2G2D	OGDHL

Differential expression	Gene. symbol				
	ENPP1	ADH1B	LDHAL6A	LIPF	CA5A
	SORD	PIK3C2G	NMNAT3	CTH	MIOX
	PDE10A	UGT2B11	ALDH1A1	ALAS2	

Protein-protein Interactions Of Associated Genes

The protein-protein interaction network of key genes (ARNTL, BHLHE40, PER2, NCOR1, RXRG, CCND1, SERPINE1 and PLAU) and risk metabolic genes (ITPKA, FBP2, AOX1, PLA2G7, GPAT3, HAGHL, PFKFB1, GGCT, PGK1, PAPSS2 and SGPP2) of thyroid carcinoma associated form a closely related network diagram.

There are indirect interaction among ARNTL, BHLHE40 and 10 risk genes, which including ITPKA, FBP2, AOX1, PLA2G7, GPAT3(AGPAT9), PFKFB1, GGCT, PGK1, PAPSS2 and SGPP2 (Fig. 5).

Conserved MicroRNAs Interactions Of Associated Genes

The thyroid carcinoma risk metabolic genes associated with biological clock were mined, which are ARNTL, BHLHE40, PER2, NCOR1, RXRG, CCND1, SERPINE1, PLAU, ITPKA, FBP2, AOX1, PLA2G7, GPAT3, HAGHL, PFKFB1, GGCT, PGK1, PAPSS2 and SGPP2. The conserved microRNAs of these genes were identified using Targetscan. Contrast research found that ARNTL, BHLHE40, PAPSS2, CCND1, PLAU, SERPINE1, PER2 and NCOR1 are regulated by many common conserved microRNAs, such as hsa-let-7a-5p, hsa-miR-106, hsa-miR-181, hsa-miR-449, hsa-miR-34, hsa-miR-30, hsa-miR-27, hsa-miR-23 and so on (Fig. 6).

Gene Set Enrichment Analyses

Five hundred and seven patients (with a follow-up period longer than 30 days) were divided into a high-risk and low-risk group with an optimal cut-off of 0.5 for risk score (Fig. 7A). Seventy-two significantly enriched KEGG pathways in TCGA-THCA were mined by GSEA. Vast majority of the enriched KEGG pathways were related to metabolism or metabolic diseases, such as the metabolism of tyrosine, sphingolipid, pyrimidine, glutathione, glycerolipid, fatty acid, insulin signaling pathway, autoimmune thyroid disease, the maturity-onset diabetes of the young and type II diabetes mellitus. Besides, the other enriched KEGG pathways were not related to metabolism but frequently pathways in carcinoma, such as the systemic lupus erythematosus, colorectal cancer, bladder cancer, pancreatic cancer and thyroid cancer. Besides, most of the metabolism-related pathways were enriched in the low risk group, whereas most of the non-metabolism related pathways were enriched in the high-risk group (Fig. 7B).

Discussion

The coming of the era of big data has a huge impact on the biomedical field and is the latest driving force in such field. A variety of high-throughput technologies have emerged accumulating a large number of experimental datum and results, which need corresponding database to storage and bioinformatics method to analyze urgently. In recent years, TCGA database provides a large amount of genomic and clinical data for cancer basic medicine and translational medicine study, thus providing a large amount of data basis for discovering biological mechanisms affecting tumor initiation, development, differentiation and metastasis [26]. GEO database is the largest and most comprehensive public gene expression data resource, which contains a variety of disease types except tumor [27]. At present, the combination of TCGA and GEO database for data mining is an important method of cancer analysis, and would facilitate further study and practical application in future. Recent years, the incidence of thyroid carcinoma in China is on the rise, the incidence is almost 3.3-times as high in women versus men [28]. It is very important to screen molecular targets and study the pathogenesis of thyroid cancer because of its disease characteristics, such as insidious incidence, low rate of preoperative diagnosis and poor prognosis of advanced patients.

The occurrence of cancer is closely related to metabolic disorder, which is usually caused by the break of biological rhythm. Biological rhythm events are controlled by biological clocks, which influenced by environment, such as temperature, light, diet, physical exercise and so on[29]. More than 20% of genes in mammal are controlled by biological clock genes [30], which as internal timekeeping mechanism to regulate metabolic and physiological processes rhythmically. Therefore, biological clock disorders can lead to a variety of diseases, such as depression, obesity, cancer and so on [31–33]. Consequently, excavating thyroid carcinoma risk metabolic genes associated with biological clock is imminent.

In the study, four hundred and seventy-four DEGs were identified in patients with thyroid carcinoma compared with the normal controls, including 295 up-regulated genes and 179 down regulated genes, and 64 biorhythm related genes ((including 12 core clock genes) were mined, which is helpful for us to study the relationship between thyroid cancer and biological clock genes. Furthermore, BHLHE40 and ARNTL of 141 TC-DEGs and 64 BRGs were identified to overlap. BHLHE40 is transcriptional repressor regulates the activity of the clock genes and clock-controlled genes, involving in the control of circadian rhythm and cell differentiation [34–35]. ARNTL is core clock gene, that encode a basic helix-loop-helix protein that forms a heterodimer (CLOCK/ARNTL complexes) with CLOCK. This heterodimer can activate transcription of PER1, PER2, PER3, CRY1 and CRY2 genes [36]. Defects in ARNTL have been linked to infertility, problems with gluconeogenesis and lipogenesis, and altered sleep patterns [37]. The BHLHE40 protein can interact with ARNTL or compete for E-box binding sites in the promoter of PER1 and repress CLOCK/ARNTL's transactivation of PER1[38]. Among ARNTL related pathways are regulation of lipid metabolism by Peroxisome proliferator-activated receptor alpha and Circadian entrainment [39].

The chromosomal assignments of TC-DEGs and BRGs were inhomogeneous, which mainly distributed on Chrom01, 02, 11, 12, 04, 03 and X. These genes have the characteristics of linkage distribution, suggesting that these genes may have related functions, which help in understanding of their regulation and role between biorhythm disorder and thyroid carcinoma. None of the associated genes were distributed on sex Y chromosomes, it may be related to the higher incidence of female than male.

Eighteen most significantly enriched GO process terms (FDR $\leq$ 0.05 and Count $\geq$ 5 genes as criteria) and 18 significantly enriched pathways (P-value $\geq$ 0.05 and Count $\geq$ 10)for TC-DEGs and BRG were obtained. Abnormal expression of PER2, NCOR1, RXRG, CCND1, SERPINE1 and PLAU genes in thyroid cancer, which were mined acting as the hub of signaling pathways linking “Thyroid hormone signaling pathway”, “Circadian rhythm”, “Transcriptional misregulation in cancer” and so on. The specific functions of PER2 are mainly reflected in the following three aspects: regulating the physiological rhythm of the body, regulating cell cycle and promoting DNA damage repair [40]. PER2 is one of the core clock genes of physiological rhythm regulation, and it also exists in the signal pathway of “Transcriptional misregulation in cancer” in our study, which means PER2 can affect the development of tumor. NCOR1 and RXRG are the key links between “Transcriptional misregulation in cancer” and “Thyroid hormone signaling pathway”. NCOR1 mediates ligand-independent transcription repression of thyroid-hormone and retinoic-acid receptors by promoting chromatin condensation and preventing access of the transcription machinery [41]. Diseases associated with NCOR1 include Rett Syndrome and Endometrial Hyperplasia [42–43]. RXRG regulate gene expression in various biological processes. RXRG forms heterodimer with retinoic acid, thyroid hormone and vitamin D receptor, which increases the DNA binding and transcription function of their respective reaction elements [44]. Amplification and overexpression of CCND1gene, which alters cell cycle progression, are observed frequently in a variety of tumors and may contribute to tumorigenesis [45]. Defects in SERPINE1 gene are the cause of plasminogen activator inhibitor-1 deficiency, and high concentrations of the gene product are associated with thrombophilia [46]. PLAU has been implicated in many pathophysiological processes, as a potent marker of invasion and metastasis in a variety of human cancers associated with breast, stomach, colon, bladder, ovary, brain and endometrium [47]. In the paper, one hundred and forty-one metabolic genes (77 up-regulated and 64 down-regulated) were included in TCGA-THCA and GSE33630 to train the prognostic model. Eleven risk genes including ITPKA, FBP2, AOX1, PLA2G7, GPAT3, HAGHL, PFKFB1, GGCT, PGK1, PAPSS2 and SGPP2 were found by univariate Cox regression model ($p < 0.01$). Gene GGCT, HAGHL and ITPKA are highly expressed in tumor cells, and PLA2G7, GPAT3, PAPSS2, AOX1, FBP2, SGPP2, PGK1 and PFKFB1 are lower expressed in tumor cells compared with the normal controls. The levels of ITPKA that are important in cellular signaling [48]. FBP2 as gluconeogenesis regulatory enzyme which catalyzes the hydrolysis of fructose 1,6-bisphosphate to fructose 6-phosphate and inorganic phosphate [49]. Compared with normal cells, tumor cells have significant changes in the pattern of glucose metabolism. Under the condition of sufficient oxygen, normal cells metabolize glucose by oxidative phosphorylation, while tumor cells metabolize glucose mainly by glycolysis [50]. The expression level of FBP2 was verified in thyroid carcinoma. AOX1 gene encodes aldehyde oxidase to catalyze the formation of superoxide [51]. Abnormal expression of AOX1 can

lead to metabolic disorder and cancer. Defects in PLA2G7 are a cause of platelet-activating factor acetyl hydrolase deficiency [52]. GPAT3 protein catalyzes the conversion of glycerol-3-phosphate to lysophosphatidic acid in the synthesis of triacylglycerol, this gene is regulated by insulin-stimulated phosphorylation and play distinct roles in adipogenesis [53–54]. Diseases associated with HAGHL include myopathy, centronuclear [55]. The abnormal expression of PFKFB1 in thyroid tissue, which play an important role in the proliferation, migration, invasion and metastasis of thyroid cancer [56]. PFKFB1 can be used as a potential molecular target for thyroid biotherapy. The GGCT play a role in cell proliferation, and the expression of this gene is a potential marker for cancer [57].

PGK1 protein as a glycolytic enzyme which is secreted by tumor cells leads to the release of the tumor blood vessel inhibitor angiostatin [58]. Defects in PAPSS2 cause the Pakistani type of spondyloepimetaphyseal dysplasia [59]. SGPP2 is a transmembrane protein that degrades the bioactive signaling molecule sphingosine 1-phosphate and induce during inflammatory responses [60]. The protein-protein interaction network of ARNTL, BHLHE40, PER2, NCOR1, RXRG, CCND1, SERPINE1, PLAU, ITPKA, FBP2, AOX1, PLA2G7, GPAT3, HAGHL, PFKFB1, GGCT, PGK1, PAPSS2 and SGPP2 form a closely related network diagram.

Vast majority of the enriched KEGG pathways were related to metabolism, metabolic diseases, biologic rhythm and frequently pathways in carcinoma. Most of the metabolism-related pathways were enriched in the low risk group, whereas most of the non-metabolism related pathways were enriched in the high-risk group. It can be seen that the occurrence of thyroid carcinoma is closely related to metabolic disorder and biological clock. Therefore, future biological experiments should pursue the function of important genes and associated miRNAs in TC.

Conclusions

Our study excavated a novel nineteen risk metabolic genes associated with biological clock by bioinformatics analysis base on TCGA and GEO data, which provided potential biomarkers for metabolic therapy and treatment response prediction of thyroid carcinoma. This way is useful to establish the information mining method that integrates biological rhythm disorder and related diseases, and has important theoretical guiding significance and broad application prospects for exploring new treatment approaches, which helpful to improve health and prevent diseases. However, the experimental validations of the risk metabolic genes are warranted.

Abbreviations

TC: Thyroid carcinoma; TC-DEGs: differentially expressed genes of Thyroid carcinoma; BRGs: biorhythm related genes; PTC: papillary thyroid cancer; FTC: follicular thyroid cancer; MTC: medullary thyroid cancer; ATC: anaplastic thyroid cancer; mRNA: messenger RNA; TCGA: The Cancer Genome Atlas; GEO: Gene

Expression Omnibus; GO: Gene Ontology; DAVID: Database for Visualization, Annotation and Integrated Analysis; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSEA: Gene Set Enrichment Analysis;

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All the data supporting the findings of this study are contained in the

manuscript and the corresponding supplementary files.

Competing interests

The authors declare that they have no competing interests.

Funding

The present study was supported by a grant from the Key Programs of the Educational Commission of Hebei Province of China (grant no. ZD2018076); Youth Science and Technology Project of Health and Municipal Commission of Population and Family of Hebei Province of China (grant no. 20180811); Science and technology research project of Hebei higher education institution (grant no. Z018032);

Authors' contributions

WW and XC designed the study, WW, JZ, ZL, HZ and XZ performed the data analysis, and WW wrote the manuscript. XC supervised this work. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Author details

Department of Biochemistry College of Laboratory Medicine, Hebei North University, Zhangjiakou 075000, Hebei, China.

References

1. James BC, Aschebrook-Kilfoy B, Cipriani N, Kaplan EL, Angelos P, Grogan RH. The Incidence and Survival of Rare Cancers of the Thyroid, Parathyroid, Adrenal, and Pancreas. Ann Surg Oncol.2016; doi:1245/s10434-015-4901-9
2. Pishkari S, Paryan M, Hashemi M, Baldini E, Mohammadi-Yeganeh S. The role of microRNAs in different types of thyroid carcinoma: a comprehensive analysis to find new miRNA supplementary therapies. J Endocrinol Invest.2018; doi:10.1007/s40618-017-0735-6.
3. Qin WY, Guan CN, Xu ZM, Li JW. Analysis of clinical characteristics of 854 patients with thyroid cancer. Lin Chung Er Bi Yan Hou Tou Jing Wai Ke Za Zhi.2019; doi:10.13201/j.issn.1001-1781.2019.08.009.
4. Diquigiovanni C, Bergamini C, Evangelisti C, Isidori F, Vettori A, Tiso N, et al. Mutant MYO1F alters the mitochondrial network and induces tumor proliferation in thyroid cancer. Int J Cancer.2018; doi:10.1002/ijc.31548.
5. Qiu J, Zhang W, Xia Q, Liu F, Zhao S, Zhang K, et al. Investigating the mechanisms of papillary thyroid carcinoma using transcriptome analysis. Mol Med Rep.2017; doi:10.3892/mmr.2017.7346.
6. Hanahan, D, Weinberg, R. A. Hallmarks of cancer: The next generation. Cell. 2011; doi:10.1016/j. cell.2011.02.013.
7. Yin DT, He H, Yu K, Xie J, Lei M, Ma R, et al. The association between thyroid cancer and insulin resistance, metabolic syndrome and its components: A systematic review and meta-analysis. Int J Surg.2018; doi:10.1016/j. ijsu.2018.07.013.
8. Feng D, Lazar MA. Clocks, metabolism, and the epigenome. Mol Cell. 2012; doi:10.1016/j.molcel.2012.06.026
9. Tomás G, Tarabichi M, Gacquer D, Hébrant A, Dom G, Dumont JE, et al. A general method to derive robust organ-specific gene expression-based differentiation indices: application to thyroid carcinoma diagnostic. Oncogene. 2012; doi:10.1038/onc.2011.626.
10. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res.2015; doi:10.1093/nar/gkv007.
11. Lin E, Kuo PH, Liu YL, Yang AC, Kao CF, Tsai SJ. Effects of circadian clock genes and health-related behavior on metabolic syndrome in a Taiwanese population: Evidence from association and interaction analysis. PLoS One.2017; doi:10.1371/journal. pone. 0173861.
12. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017; doi:10.1093/nar/gkw937.
13. Miga KH, Newton Y, Jain M, Altemose N, Willard HF, Kent WJ. Centromere reference models for human chromosomes X and Y satellite arrays. Genome Res. 2014; doi:10.1101/gr.159624.113.

14. Jiangtao C, Yingzhen K, Qian W, Yuhe S, Daping G, Jing L, et al. MapGene2Chrom, a tool to draw gene physical map based on Perl and SVG languages. Yi Chuan. 2015; doi:10.16288/j.yczz.2015.01.013.
15. Hulsegge I, Kommadath A, Smits MA. Two different approaches for Gene Ontology analysis. BMC Proc 2009;3 Suppl 4: S10.
16. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. NatProtoc. 2009; doi:10.1038/nprot.2008.211.
17. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; doi:1093/nar/28.1.27.
18. Powers RK, Goodspeed A, Pielke-Lombardo H, Tan AC, Costello JC. GSEA-InContext: identifying novel and common patterns in expression experiments. 2018; doi:10.1093/bioinformatics/bty271.
19. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst.2015; doi:1016/j.cels.2015.12.004.
20. Yang R, Xiong J, Peng Y, Peng X, Zeng X. A SAS marco program for batch processing of univariate Cox regression analysis for great database. Zhong Nan Da Xue Xue Bao Yi Xue Ban.2015; doi:10.11817/j. issn.1672-7347.2015.02.013.
21. Rizvi AA, Karaesmen E, Morgan M, Preus L, Wang J, Sovic M, et al. gwasurvivr: an R package for genome-wide survival analysis. 2019; doi:10.1093/bioinformatics/bty920.
22. Anders Eklund, Thomas E. Nichols, Hans Knutsson. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. PNAS. 2016; doi:1073/pnas.1602413113
23. Liu GM, Xie WX, Zhang CY, Xu JW. Identification of a four-gene metabolic signature predicting overall survival for hepatocellularcarcinoma. J Cell Physiol.2020; doi:10.1002/jcp.29081.
24. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; doi:1101/gr.1239303.
25. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. 2015; doi: 10.7554/eLife.05005.
26. Jia D, Li S, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. Aging (Albany NY).2018; doi:10.18632/aging.101415.
27. Clough E, Barrett T. The Gene Expression Omnibus Database. Methods Mol Biol.2016; doi:10.1007/978-1-4939-3578-9_5.
28. Yang L, Zheng RS, Wang N, Zeng HM, Yuan YN, Zhang SW, et al. Analysis of Incidence and Mortality of Thyroid Cancer in China, 2013. Zhonghua Zhong Liu Za Zhi.2017; doi:10.3760/cma.j. issn.0253-3766.2017.11.010.
29. Tognini P, Murakami M, Liu Y, Eckel-Mahan KL, Newman JC, Verdin E, et al. Distinct Circadian Signatures in Liver and Gut Clocks Revealed by Ketogenic Diet. Cell Metab.2017; doi:10.1016/j.

cmet.2017.08.015.

30. Benna C, Helfrich-Förster C, Rajendran S, Monticelli H, Pilati P, Nitti D. Genetic variation of clock genes and cancer risk: a field synopsis and meta-analysis. Oncota get. 2017; doi:10.18632/oncotarget.15074.
31. Deng TS. Biological clocks, some clock-related diseases, and medicinal plants. Psych J.2018; doi:10.1002/pchj.263.
32. Fuhr L, El-Athman R, Scrima R, Cela O, Carbone A, Knoop H, et al. The Circadian Clock Regulates Metabolic Phenotype Rewiring Via HKDC1 and Modulates Tumor Progression and Drug Response in Colorectal Cancer. 2018; doi:10.1016/j. ebiom.2018.07.002.
33. Weger M, Diotel N, Dorsemans AC, Dickmeis T, Weger BD. Stem cells and the circadian clock. Dev Biol.2017; doi:10.1016/j. ydbio.2017.09.012.
34. Nakashima A, Kawamoto T, Noshiro M, Ueno T, Doi S, Honda K, et al. Dec1 and CLOCK Regulate Na+/K+-ATPase β1 Subunit Expression and Blood Pressure. 2018; doi:10.1161/HYPERTENSIONAHA.118.11075.
35. Wang C, Liu W, Liu Z, Chen L, Liu X, Kuang S. Hypoxia Inhibits Myogenic Differentiation through p53 Protein-dependent Induction of BHLHE40 Protein. J Biol Chem.2015; doi:10.1074/jbc.M115.688671.
36. Terajima H, Yoshitane H, Ozaki H, Suzuki Y, Shimba S, Kuroda S, et al. ADARB1 catalyzes circadian A-to-I editing and regulates RNA rhythm. Nat Genet.2017; doi:10.1038/ng.3731.
37. Sato F, Kohsaka A, Bhawal UK, Muragaki Y. Potential Roles of Dec and Bmal1 Genes in Interconnecting Circadian Clock and Energy Metabolism. Int J Mol Sci.2018; doi:10.3390/ijms19030781.
38. Zheng Q, Wang C, Wang L, Zhang D, Liu N, Ming X, et al. Interaction with SP1, but not binding to the E-box motifs, is responsible for BHLHE40/DEC1-induced transcriptional suppression of CLDN1 and cell invasion in MCF-7 cells. Mol Carcinog.2018; doi:10.1002/mc.22829.
39. Godinho-Silva C, Domingues RG, Rendas M, Raposo B, Ribeiro H, da Silva JA, et al. Light-entrained and brain-tuned circadian circuits regulate ILC3s and gut homeostasis. 2019; doi:10.1038/s41586-019-1579-3.
40. Corrà S, Salvadori R, Bee L, Barbieri V, Mognato M. Analysis of DNA-damage response to ionizing radiation in serum-shock synchronized human fibroblasts. Cell Biol Toxicol.2017; doi:10.1007/s10565-017-9394-9.
41. Shimizu H, Lu Y, Vella KR, Damilano F, Astapova I, Amano I, et al. Nuclear corepressor SMRT is a strong regulator of body weight independently of its ability to regulate thyroid hormone action. PLoS One.2019; doi:10.1371/journal. pone. 0220717.
42. Kruusvee V, Lyst MJ, Taylor C, Tarnauskaitė Ž, Bird AP, Cook AG. Structure of the MeCP2-TBLR1 complex reveals a molecular basis for Rett syndrome and related disorders. Proc Natl Acad Sci U S A.2017; doi:10.1073/pnas.1700731114.
43. Kashima H, Horiuchi A, Uchikawa J, Miyamoto T, Suzuki A, Ashida T, et al. Up-regulation of nuclear receptor corepressor (NCoR) in progestin-induced growth suppression of endometrial

hyperplasia and carcinoma. Anticancer Res.2009; 29(4):1023-9.

44. Gely-Pernot A, Raverdeau M, Teletin M, Vernet N, Féret B, Klopfenstein M, et al. Retinoic Acid Receptors Control Spermatogonia Cell-Fate and Induce Expression of the SALL4A Transcription Factor. PLoS Genet.2015; doi:10.1371/journal. pgen.1005501.
45. Sathyan KM, Nalinakumari KR, Abraham T, Kannan S. CCND1 polymorphisms (A870G and C1722G) modulate its protein expression and survival in oral carcinoma. Oral Oncol.2008; doi:1016/j.oraloncology.2007.09.003
46. Sirgo G, Morales P, Rello J. PAI-1 gene: pharmacogenetic association of 4G/4G genotype with bleeding after cardiac surgery--pilot study. Eur J Anaesthesiol.2009; doi:10.1097/EJA.0b013e3283240412.
47. Huang Q, Wei H, Wu Z, Li L, Yao L, Sun Z, et al. Preferentially Expressed Antigen of Melanoma Prevents Lung Cancer Metastasis. PLoS One.2016; doi:10.1371/journal. pone. 0149640.
48. Windhorst S, Song K, Gazdar AF. Inositol-1,4,5-trisphosphate 3-kinase-A (ITPKA) is frequently over-expressed and functions as an oncogene in several tumor types. Biochem Pharmacol.2017; doi:10.1016/j. bcp.2017.03.023.
49. Bakshi I, Suryana E, Small L, Quek LE, Brandon AE, Turner N, et al. Fructose bisphosphatase 2 overexpression increases glucose uptake in skeletal muscle. J Endocrinol.2018; doi:10.1530/JOE-17-0555.
50. Li Z, Zhang H. Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. Cell Mol Life Sci.2016; doi:10.1007/s00018-015-2070-4.
51. Foti A, Dorendorf F, Leimkühler S. A single nucleotide polymorphism causes enhanced radical oxygen species production by human aldehyde oxidase. PLoS One.2017; doi:10.1371/journal. pone. 0182061.
52. Stafforini DM. Biology of platelet-activating factor acetylhydrolase (PAF-AH, lipoprotein associated phospholipase A2). Cardiovasc Drugs Ther.2009; doi:10.1007/s10557-008-6133-8.
53. Yu J, Loh K, Song ZY, Yang HQ, Zhang Y, Lin S. Update on glycerol-3-phosphate acyltransferases: the roles in the development of insulin resistance. Nutr Diabetes.2018; doi:10.1038/s41387-018-0045-x.
54. Shan D, Li JL, Wu L, Li D, Hurov J, Tobin JF, et al. GPAT3 and GPAT4 are regulated by insulin-stimulated phosphorylation and play distinct roles in adipogenesis. J Lipid Res.2010; doi:10.1194/jlr.M006304.
55. Beltran J, Jbara M, Maimon R. Shoulder: labrum and bicipital tendon. Top Magn Reson Imaging.2003; doi:1097/00002142-200302000-00004.
56. Peng F, Li Q, Sun JY, Luo Y, Chen M, Bao Y. PFKFB3 is involved in breast cancer proliferation, migration, invasion and angiogenesis. Int J Oncol.2018; doi:10.3892/ijo.2018.4257.
57. Kageyama S, Ii H, Taniguchi K, Kubota S, Yoshida T, Isono T, et al. Mechanisms of Tumor Growth Inhibition by Depletion of γ-Glutamylcyclotransferase (GGCT): A Novel Molecular Target for Anticancer Therapy. Int J Mol Sci.2018; doi:10.3390/ijms19072054.

58. Zhang Y, Yu G, Chu H, Wang X, Xiong L, Cai G, et al. Macrophage-Associated PGK1 Phosphorylation Promotes Aerobic Glycolysis and Tumorigenesis. Mol Cell.2018; doi:10.1016/j. molcel.2018.06.023.
59. Tüysüz B, Yılmaz S, Gül E, Kolb L, Bilguvar K, Evliyaoğlu O, et al. Spondyloepimetaphyseal dysplasia Pakistani type: expansion of the phenotype. Am J Med Genet A.2013; doi:10.1002/ajmg.a.35906.
60. Taguchi Y, Allende ML, Mizukami H, Cook EK, Gavrilova O, Tuymetova G, et al. Sphingosine-1-phosphate Phosphatase 2 Regulates Pancreatic Islet β-Cell Endoplasmic Reticulum Stress and Proliferation. J Biol Chem.2016; doi:10.1074/jbc.M116.728170.

Figures

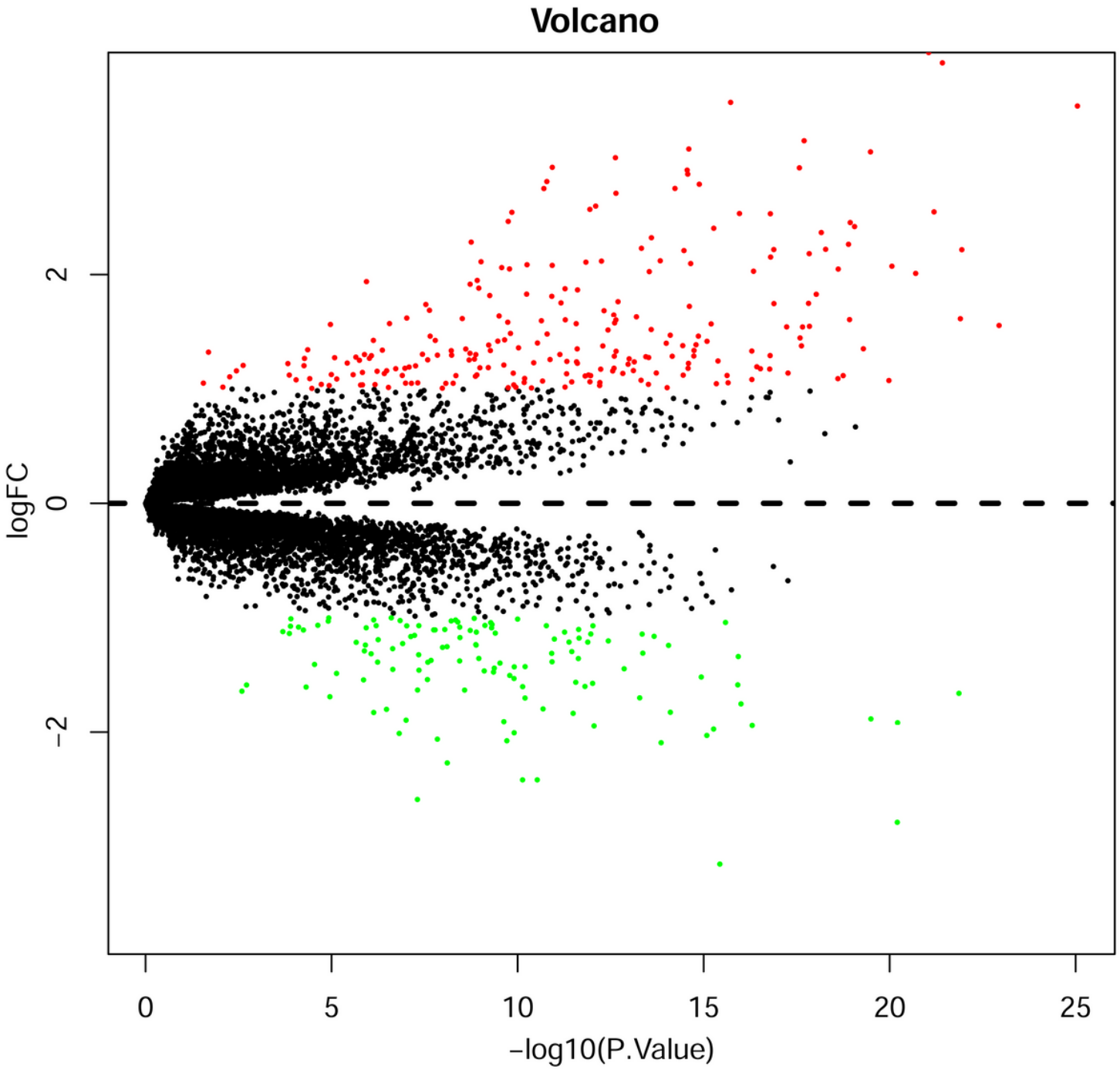

Figure 1

Volcano map of 474 differentially expressed genes (P<0.05; |log FC| >1). The genes with log FC > 1 are up-regulated genes, and the genes with log FC < -1 are down-regulated genes. The volcano map was accomplished by “limma” data package of R software.

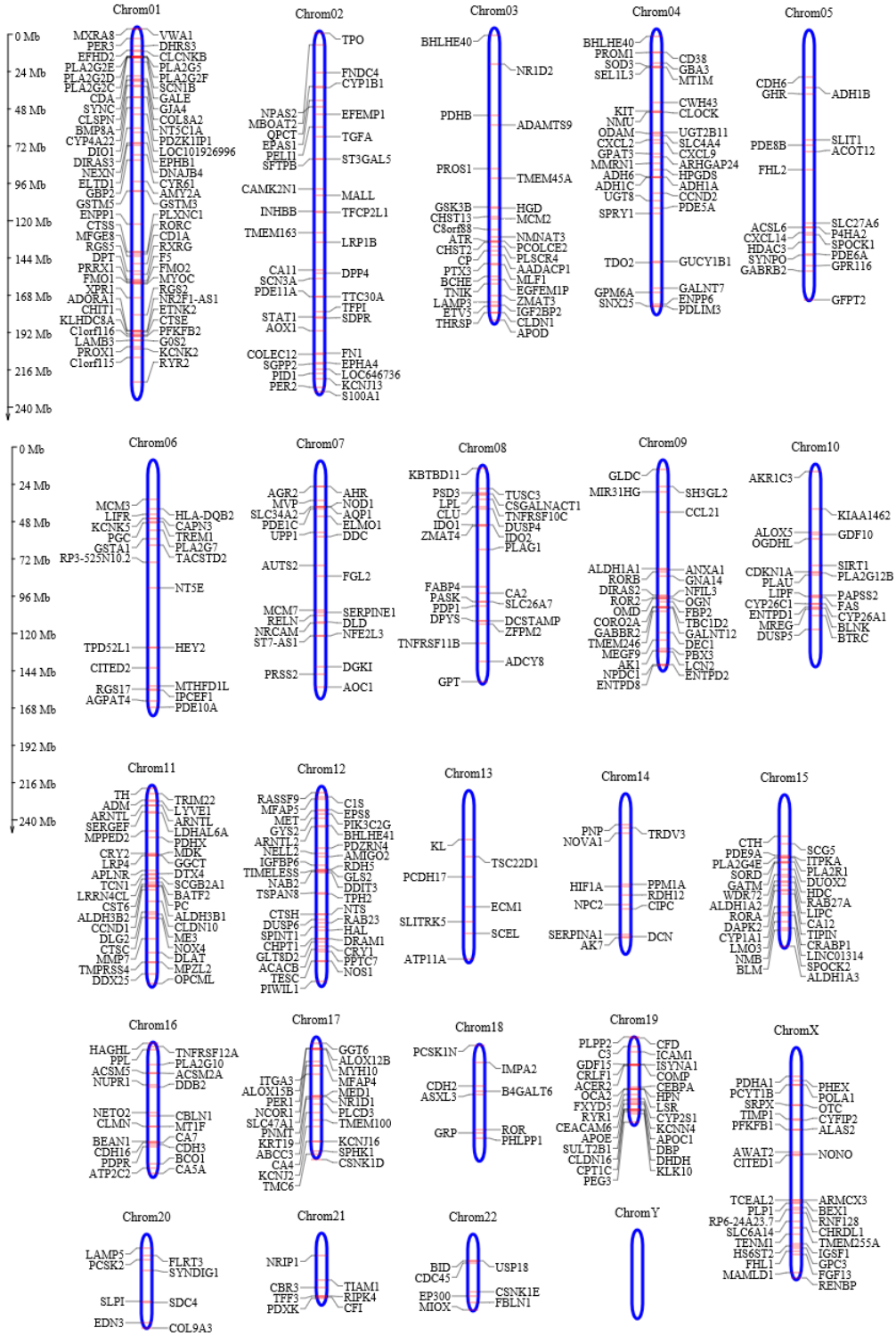

Figure 2

The chromosomal mapping of 474 TC-DEGs and 64 BRGs by UCSC and MapGene2Chrom web v2 online analytical tool.

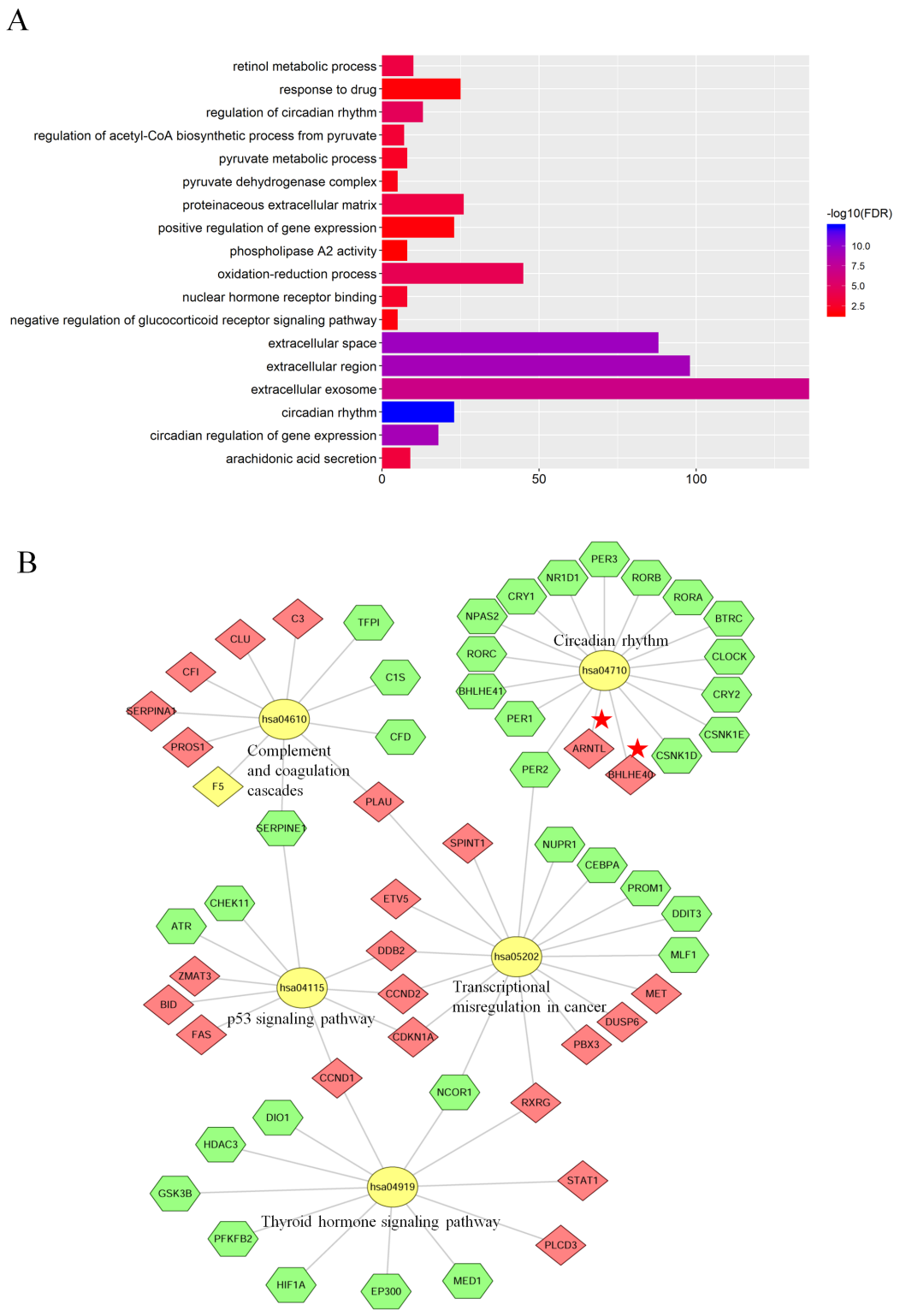

Figure 3

Enriched GO process terms and KEGG pathways of TC-DEGs and BRGs (A) Enriched GO process terms of TC-DEGs and BRGs obtained from the DAVID, visualization of the picture is accomplished by “ggplot2” data package of R software. GO, gene ontology. (B) KEGG pathways of TC-DEGs and BRGs obtained from the Kyoto Encyclopedia of Genes and Genomes database resource, visualization of the picture is accomplished by cytoscape version 3.6.1 software.

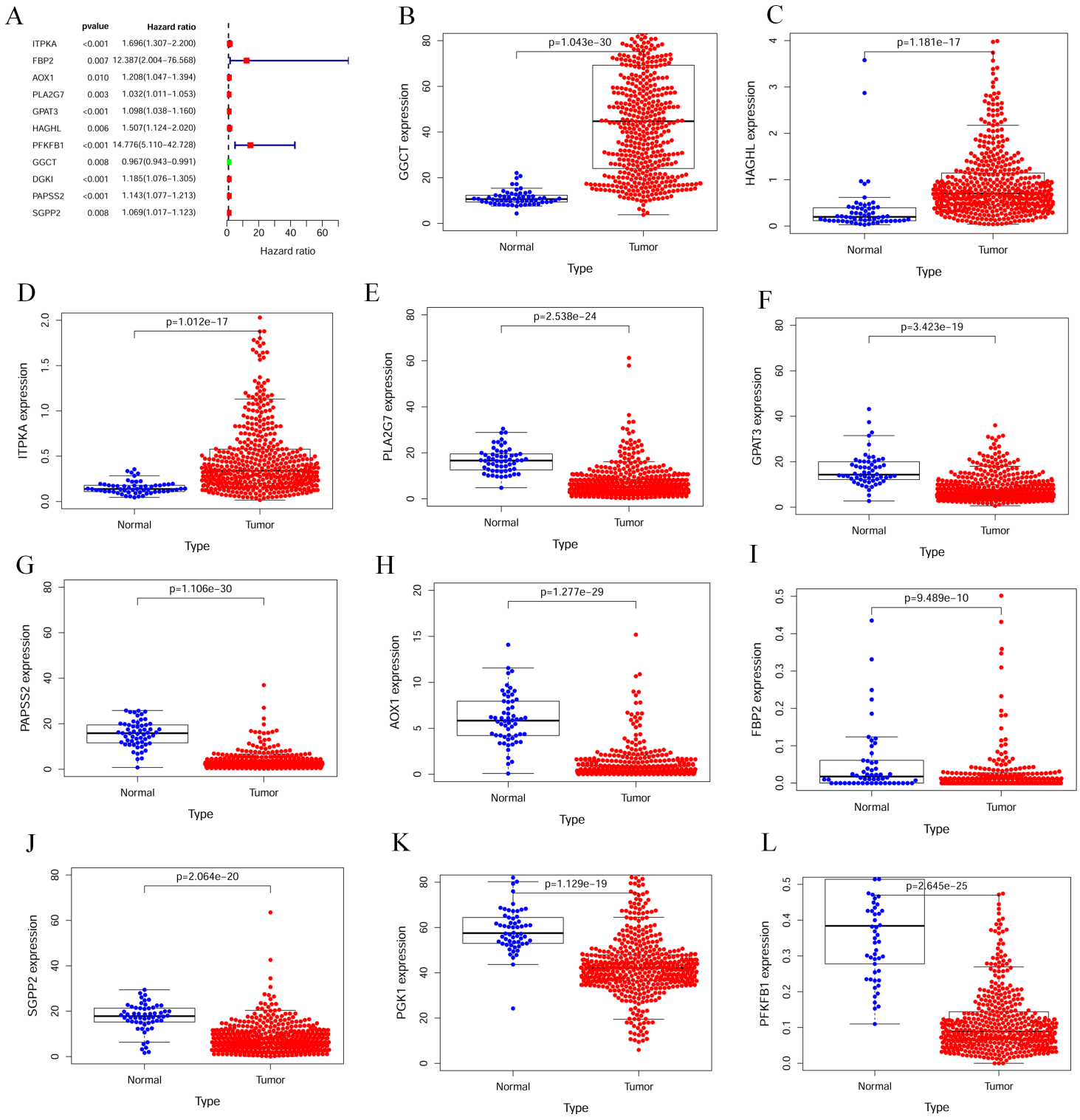

Figure 4

The forest map of 11 high-risk metabolic genes (A) and single gene differential expression of 11 high-risk metabolic genes (B - L). The forest map was accomplished by “survival” data package of R software, single gene differential expression was accomplished by “limma” and “beeswarm” data package of R software. $p < 0.01$ as criteria, which were considered statistically significant.

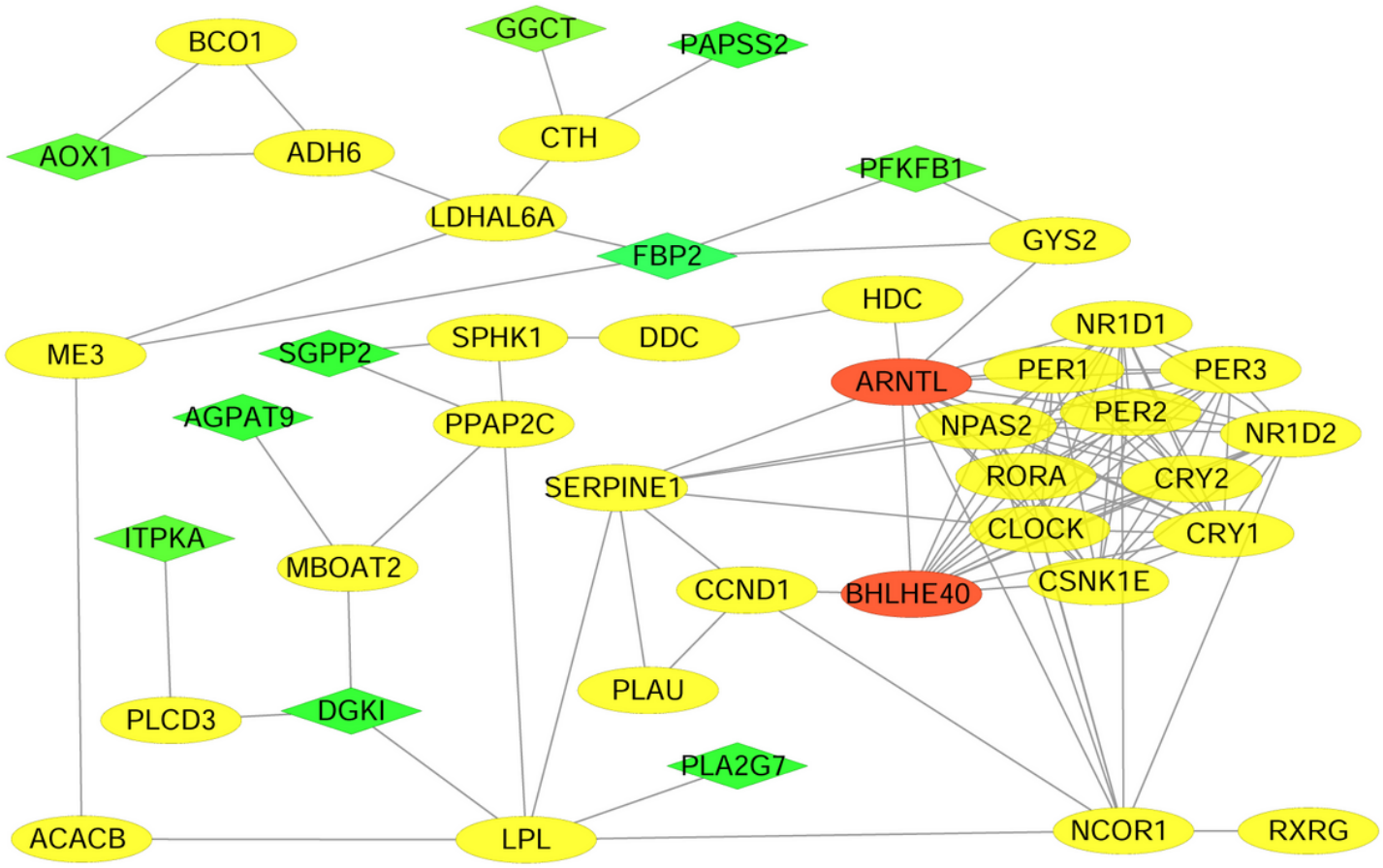

Figure 5

Protein-protein interaction network of key genes and risk metabolic genes. The figure was plotted using cytoscape version 3.6.1 software. Diamond represents risk gene.

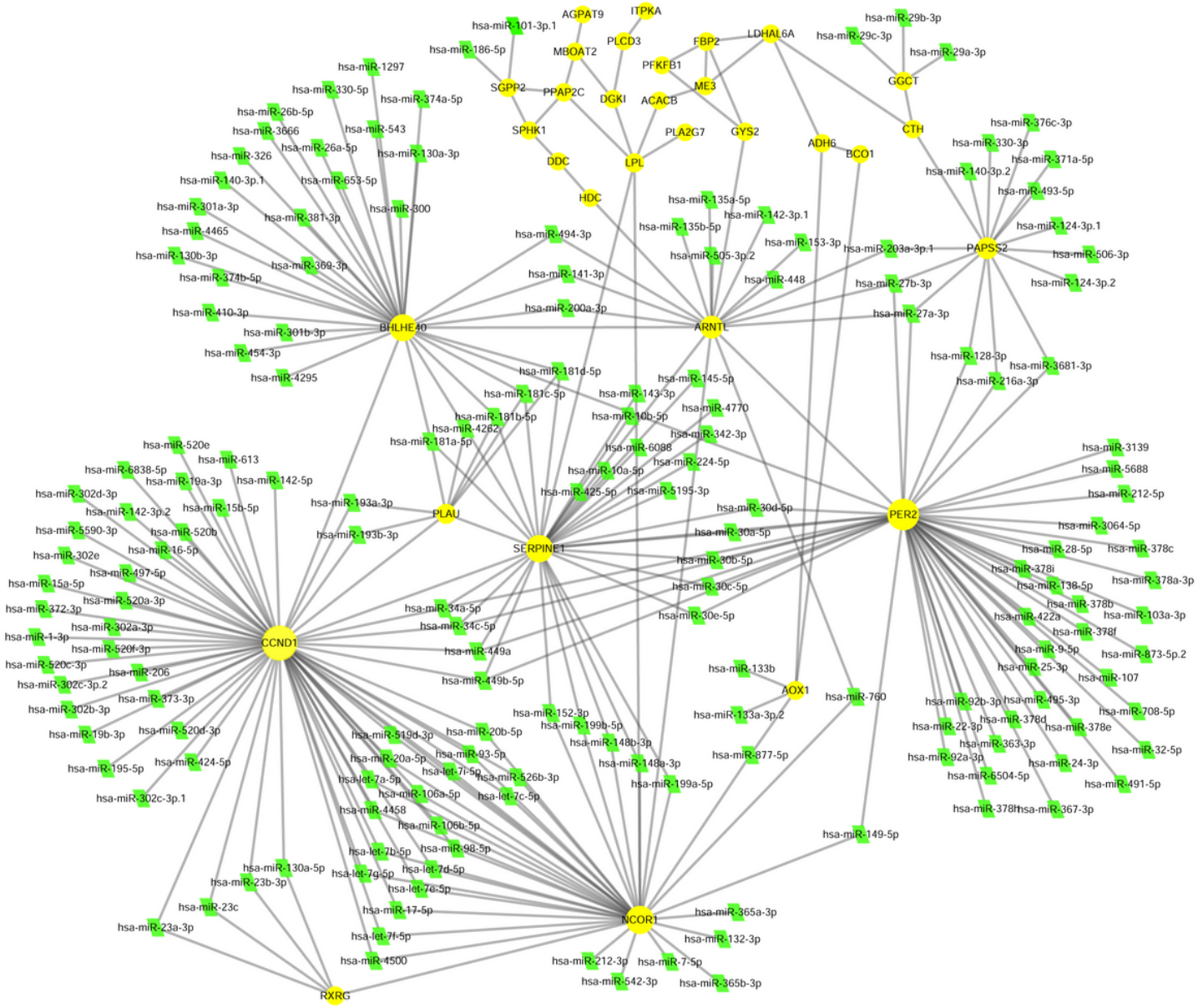

Figure 6

The conserved microRNAs of TC risk metabolic genes. The conserved microRNAs were identified using Targetscan and visualized by cytoscape version 3.6.1 software. Diamond represents conserved microRNA.

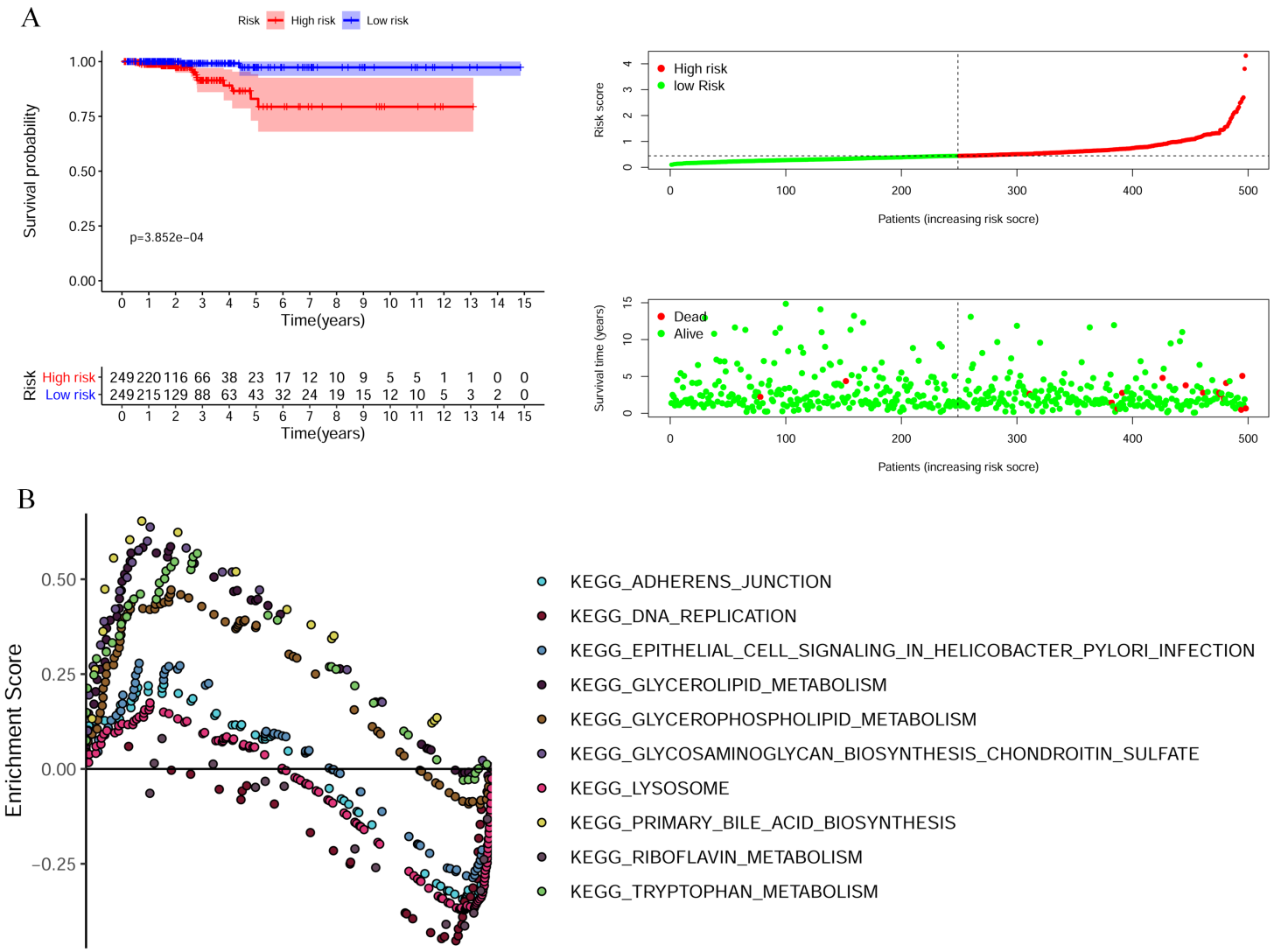

Figure 7

Risk score analysis（A）and The significantly enriched KEGG pathways in TCGA-THCA and GSE33630 cohort by GSEA (B). GSEA, Gene Set Enrichment Analyses; TCGA, The Cancer Genome Atlas [Color figure can be viewed at wileyonlinelibrary.com]. There shows Five representative KEGG pathways in high-risk patients and five representative KEGG pathways in low-risk patients.