Identification of potential biomarkers in hepatocellular carcinoma: a network-based approach

Mehrdad Ameri ¹, Haniye Salimi ², Sedigheh Eskandari ³, Navid Nezafat ^{3,4*}

- ^{1.} Diagnostic Laboratory Sciences and Technology Research Center, School of Paramedical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran.
- ^{2.} Organ transplant research center, Shiraz University of Medical Sciences, Shiraz, Iran.
- ^{3.} Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran.
- ^{4.} Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, P.O. Box 71345-1583, Shiraz, Iran.

Abstract

Background: Hepatocellular carcinoma (HCC) is one of the leading causes of death worldwide. Identification of potential therapeutic and diagnostic biomarkers can be helpful to screen cancer progress. This study implemented with the aim of discovering potential biomarkers for HCC within a network-based approach integrated with microarray data.

Methods: Through downloading a gene expression profile GSE62232 differentially expressed genes (DEGs) were identified. Gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis for DEGs were performed utilizing enrichr server. Following reconstruction of protein-protein interaction network of DEGs with STRING, network visualization, analyses, and clustering into structural modules carried out using Cytoscape. Considering degree centrality, 15 hub genes were selected as early biomarker candidates for final validation. In order to validate hub genes, GEPIA server was used to perform overall survival (OS) and disease-free survival (DFS).

Results: In our approach 1996 DEGs were identified including 995 up-regulated genes and 1001 down-regulated genes. KEGG pathway enrichment analysis shown that DEGs are associated with Chemical carcinogenesis, and Cell cycle. GO term enrichment analysis indicated the relation of DEGs with epoxigenase P450 pathway, arachidonic acid monooxygenase activity, and secretory granule lumen. Following analysis of protein-protein interaction network of DEGs top three structural modules and 15 early hub genes were selected. Validation of hub genes performed using GEPIA. Consequently, CDK1, CCNB1, CCNA2, CDC20, AURKA, MAD2L1, TOP2A, KIF11, BUB1B, TYMS, EZH2, and BUB1 were considered as our final proposed biomarkers.

Conclusion: using an integrated network-based approach with microarray data our results revealed 12 final candidates with potential to considered as biomarkers in hepatocellular carcinoma.

Keywords

Network-based, biomarkers, hepatocellular carcinoma (HCC), hub genes

Introduction

Hepatocellular carcinoma (HCC), a predominant primary liver cancer, is the sixth leading causes of death by cancer (1). A multi-step mechanism consists of an accumulation of gene alterations that is resulted in various molecular and cellular modifications is the main cause of HCC pathogenesis (2). Recent studies, have indicated various genes such as epidermal growth factor receptor (EGFR) (3), transforming growth factor-beta 1 (TGF- β 1) (4), c-myc (5), to name but few involve in tumorigenesis and progress of HCC. However, there is still a lack of a specific biomarker for precise diagnostic of HCC (6). Therefore, finding an accurate biomarker for screening HCC at different stages with high specificity and sensitivity is an urgent need.

Data obtained from microarray technology in combination with bioinformatics approaches provides a prominent strategy at molecular level for an extensive studying of dysregulation in gene expression between cancer and normal samples from patients. A large-scale microarray data has been published in different databases through recent years and utilizing these data to perform integrated analysis is valuable strategy to follow up the cancer progression in the patients. Despite all efforts in finding different gene alterations in HCC, the molecular mechanisms of HCC progression remain unclear. Consequently, analyzing microarray data to obtain differentially expressed genes (DEGs) between tumor and normal samples is a promising method to discover hub genes and key pathways, which results in HCC tumorigenesis.

In this study, expression profile of hepatocellular carcinoma was obtained from the gene expression omnibus (GEO) database for identification of hub genes in HCC. Following the DEGs identification, they were utilized for gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis, and protein-protein interaction (PPI) network reconstruction studies. Ultimately, survival analyses were performed using GEPIA to evaluate the prognostic value of each hub genes. CDK1, CCNB1, CCNA2, CDC20, AURKA, MAD2L1, TOP2A, KIF11, BUB1B, TYMS, EZH2, and BUB1 were considered as our suggested biomarkers.

Methods

Data selection and preliminary analysis

A gene expression profile of hepatocellular carcinoma (GSE62232) was obtained from gene expression omnibus (GEO) which is a repository of transcriptional and expression data (7). Our selected dataset contained samples from patients with age from 21 to 82 years old and different etiological factors. 14 tumor samples were from female patients while 67 belonged to male;

moreover, 10 normal samples were used as control. Normalization and expression calculation of downloaded CELL files were implemented with transcriptome analysis console (TAC) through summarization using RMA method additionally eBayes was applied as ANOVA method. To determine differentially expressed genes (DEGs) a threshold value of fold change > 2 or $-2 <$ and p-value < 0.05 were chosen.

Gene ontology (GO) term and KEGG pathway enrichment analysis of DEGs

Following DEGs identification, gene ontology (GO) term and kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis were performed utilizing Enrichr (8, 9) which is available at <https://amp.pharm.mssm.edu/Enrichr>. GO (10, 11) is capable of determine details about three functional features of a genes set, including biological process (BP), molecular function (MF), and cellular component (CC). KEGG (12) is a part of Japanese GenomeNet service and is available at <https://www.genome.jp/kegg/>. It can reveal the relationships between biological pathways related to disease and drugs. $P < 0.05$ is considered cutoff criteria to obtain significant result for GO and KEGG enrichment analysis.

PPI-network of DEGs and module analysis

To discover the connection between differentially expressed genes, a protein-protein interaction (PPI) network was reconstructed utilizing STRING v11.0 (13) at <https://string-db.org/>. Cytoscape 3.6.0 (14) was used for network visualization and analysis. Clustering network into structural modules was performed utilizing MCODE app (15) based on degree cutoff = 2, node score cutoff = 0.2, K-core = 2, and max-depth = 100. GO term and KEGG pathway enrichment analysis of top three modules were carried out with Enrichr software.

Hub genes PPI network and enrichment analysis

Based on highest degree centrality, top 15 nodes from PPI-network of DEGs were considered as hub genes. Selected hub genes were inputted into STRING to analysis their relationships. Furthermore, GO term and KEGG pathway enrichment analysis was also performed for hub genes.

Survival analysis of hub genes

In order to validate the selected hub genes, gene expression profiling interactive analysis (GEPIA) which is available at <http://gepia.cancerpku.cn> were employed to investigate their potential to choose as biomarkers. Therefore, in patients with liver hepatocellular carcinoma (LIHC) overall survival (OS) and disease-free survival (DFS) related to hub genes were examined.

Results

Differentially expressed genes in hepatocellular carcinoma

After downloading a microarray dataset of HCC (array type: HG-U133_Plus_2) containing 81 carcinoma and 10 normal samples, the sample normalization and comparison analysis was performed between two groups (shown in figure 1), resulted in identifying 1996 DEGs. Among identified DEGs, 995, and 1001 genes have been up-regulated and down-regulated, respectively; moreover, volcano plot and scatter plot were applied for visualization of DEGs between carcinoma and normal groups (shown in figure 2). Each of top 20 up- and down-regulated DEGS are listed in table 1, additionally expression pattern of DEGs were revealed by performing hierarchical clustering analysis (shown in figure 3).

Figure 1. Principal components analysis (PCA) of samples. Blue circles are cancer samples and red circles are normal samples.

Figure 2. A) Volcano plot and b) scatter plot for DEGs. In both pictures, green and red circles represent down-regulated, and up-regulated DEGs, respectively.

(A)

(B)

Figure 3. Hierarchical clustering of DEGs. The x axis is belonging to samples and the y axis represent the genes. Blue and red color indicates cancer and normal condition, respectively.

Table 1. Top 20 up-regulated and down-regulated DEGs.

Rank	Up-regulated DEGs	Description	Down-regulated DEGs	Description
1	AKR1B10	Aldo-keto reductase family 1, member B10 (aldose reductase)	ASCL1	Achaete-scute family bHLH transcription factor 1
2	SPINK1	Serine peptidase inhibitor, Kazal type 1	CNDP1	Carnosine dipeptidase 1 (metallopeptidase M20 family)
3	RPS4Y1	Ribosomal protein S4, Y-linked 1	HAMP	Hepcidin antimicrobial peptide
4	DDX3Y	DEAD (Asp-Glu-Ala-Asp) box helicase 3, Y-linked	CYP26A1	Cytochrome P450, family 26, subfamily A, polypeptide 1
5	GABBR1	Gamma-aminobutyric acid (GABA) B receptor, 1; ubiquitin D	MT1M	Metallothionein 1M
6	GPC3	Glypican 3	ASCL1	Achaete-scute family bHLH transcription factor 1
7	EIF1AY	Eukaryotic translation initiation factor 1A, Y-linked	CXCL14	Chemokine (C-X-C motif) ligand 14
8	COL15A1	Collagen, type XV, alpha 1	KCNN2	Potassium channel, calcium activated intermediate/small conductance subfamily N alpha, member 2
9	ACSL4	Acyl-CoA synthetase long-chain family member 4	FCN2	Ficolin (collagen/fibrinogen domain containing lectin) 2
10	CCL20	Chemokine (C-C motif) ligand 20	LINC01093	Long intergenic non-protein coding RNA 1093
11	ASPM	Abnormal spindle microtubule assembly	SLCO1B3	Solute carrier organic anion transporter family, member 1B3
12	IL32	Interleukin 32	FCN3	Ficolin (collagen/fibrinogen domain containing) 3
13	RRM2	Ribonucleotide reductase M2	CYP2C19	Cytochrome P450, family 2, subfamily C, polypeptide 19
14	TMEM45B	Transmembrane protein 45B	CYP1A2	Cytochrome P450, family 1, subfamily A, polypeptide 2
15	CRNDE	Colorectal neoplasia differentially expressed (non-protein coding)	CYP1A2	Cytochrome P450, family 1, subfamily A, polypeptide 2
16	TOP2A	Topoisomerase (DNA) II alpha	MT1F	Metallothionein 1F
17	RRM2	Ribonucleotide reductase M2	OIT3	Oncoprotein induced transcript 3
18	LYZ	Lysozyme	CLEC1B	C-type lectin domain family 1, member B
19	CAP2	CAP, adenylate cyclase-associated protein, 2 (yeast)	AVPR1A	Arginine vasopressin receptor 1A
20	SPP1	Secreted phosphoprotein 1	CRHBP	Corticotropin releasing hormone binding protein

Gene ontology (GO) term and KEGG pathway enrichment analysis of DEGs

The enriched GO term and KEGG pathway enrichment analysis were performed for identified DEGs. In regard to GO biological process results, DEGs are related to 'epoxygenase P450 pathway', 'alpha-amino acid catabolic process', 'arachidonic acid metabolic process', 'steroid metabolic process', 'monocarboxylic acid metabolic process', 'organic cyclic compound catabolic process', 'exogenous drug catabolic process', 'cellular amino acid biosynthetic process', 'monocarboxylic acid biosynthetic process', and 'drug catabolic process'.

GO molecular function showed DEGs are associated with 'arachidonic acid monooxygenase activity', 'arachidonic acid epoxygenase activity', 'steroid hydroxylase activity', 'oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor', 'protein homodimerization activity', 'oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen', 'heme binding', 'transaminase activity', 'oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen', and 'oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor'.

GO cellular component in regard to enrichment analysis were related to 'secretory granule lumen', 'membrane attack complex', 'spindle', 'condensed nuclear chromosome kinetochore', 'condensed nuclear chromosome, centromeric region', 'chromosome, centromeric region', 'cytoplasmic vesicle lumen', 'endoplasmic reticulum lumen', 'intercalated disc', 'mitotic spindle'. (Table 2, Figure 4a, 4b, 4c).

KEGG pathway enrichment analysis of DEGs are associated with ‘Chemical carcinogenesis’, ‘Cell cycle’, ‘Drug metabolism’, ‘Retinol metabolism’, ‘Glycine, serine and threonine metabolism’, ‘Metabolism of xenobiotics by cytochrome P450’, ‘p53 signaling pathway’, ‘Fatty acid degradation’, ‘PPAR signaling pathway’, ‘Tryptophan metabolism’. (Table 3, figure 4d)

Table 2. GO term enrichment analysis of DEGs.

Term	P-value	Adjusted P-value	Combined Score	Genes
Biological process				
Epoxygenase P450 pathway (GO:0019373)	2.18E-10	1.11E-06	186.2269	EPHX2; CYP4A11; CYP2C19; CYP4F12; CYP2C9; CYP2A7; CYP2A6; CYP2C8; CYP2B6; CYP2A13; CYP1A2; CYP1A1; CYP2E1
Alpha-amino acid catabolic process (GO:1901606)	5.44E-09	1.39E-05	105.573	AADAT; TAT; HAAO; HOGA1; KMO; FTCD; ACMSD; ALDH6A1; HAL; TDO2; UROC1; CBS; AMDHD1; KYNU; SARDH; IDO2
Arachidonic acid metabolic process (GO:0019369)	1.04E-08	1.78E-05	87.70913	PTGIS; EPHX2; CYP4A11; PLA2G4C; AKR1C3; CYP2C19; PTGS2; CYP4F12; CYP2A7; CYP2C9; CYP2A6; CYP2C8; CYP2B6; CYP2A13; CYP1A2; CYP1A1; CYP2E1; MGLL
Steroid metabolic process (GO:0008202)	3.77E-08	4.81E-05	58.37231	AKR1D1; NR1I2; HSD17B14; CYP2C19; CYP3A4; HSD11B1; CYP2B6; SULT1E1; CYP4V2; ACAA1; SRD5A2; SRD5A1; UGT2B15; AKR1C3; UGT2B17; CYP8B1; DHRS2; CYP39A1; CYP2C9; CYP26A1; CYP2C8; CYP2A6; CYP1A2; CYP1A1; CYP2E1
Monocarboxylic acid metabolic process (GO:0032787)	5.81E-08	5.93E-05	54.23803	GLYAT; GHR; MTHFD1L; KYNU; BSG; ME1; BCO2; ACAA1; PCK1; PCK2; GGT5; AGXT2; CYP4A11; ACOT12; HOGA1; IGF1; PRODH2; GNPAT; CYP2C9; GRHPR; CYP26A1; ALDH1A3; VNN1; CYP1A2; AGXT; ALDH8A1
Organic cyclic compound catabolic process (GO:1901361)	1.18E-07	1.01E-04	156.857	HSD11B1; AOC1; EPHX2; CYP1A2; HSD17B14; CYP3A4; ACAA1; PRODH2
Exogenous drug catabolic process (GO:0042738)	2.05E-07	1.50E-04	99.59451	CYP2C9; CYP2A7; CYP2A6; CYP2C8; CYP2B6; CYP2A13; NR1I2; CYP1A2; CYP2E1; CYP3A4; CYP2C19
Cellular amino acid biosynthetic process (GO:0008652)	2.55E-07	1.63E-04	87.98716	OAT; FOLH1; GLS2; FOLH1B; PAH; ASPG; ASNS; ALDH18A1; BCAT1; CDO1; ASPA; GLS
Monocarboxylic acid biosynthetic process (GO:0072330)	2.91E-07	1.65E-04	56.86805	SDS; OSBPL6; ACSM3; OSBPL3; AKR1D1; BGN; LPL; ASNS; ACSM5; HOGA1; CYP8B1; DCN; ACMSD; CYP39A1; ACLY; ALDH1A3; DSEL; DSE; SLC27A5
Drug catabolic process (GO:0042737)	3.53E-07	1.80E-04	92.07408	CYP2C9; CYP2A7; CYP2A6; CYP2C8; CYP2B6; CYP2A13; NR1I2; CYP1A2; CYP2E1; CYP3A4; CYP2C19
Molecular function				
Arachidonic acid monooxygenase activity (GO:0008391)	5.70E-08	6.56E-05	132.6786	CYP2C9; CYP2A7; CYP2A6; CYP2C8; CYP2B6; CYP2A13; CYP4A11; CYP2E1; CYP2C19; CYP4F12

Arachidonic acid epoxygenase activity (GO:0008392)	5.70E-08	3.28E-05	132.6786	CYP2C9; CYP2A7; CYP2A6; CYP2C8; CYP2B6; CYP2A13; CYP4A11; CYP2E1; CYP2C19; CYP4F12
Steroid hydroxylase activity (GO:0008395)	4.30E-06	0.001649	57.29352	CYP39A1; CYP2C9; CYP2A7; CYP2A6; CYP2C8; CYP2B6; CYP2A13; CYP1A1; CYP2E1; CYP3A4; CYP2C19; CYP8B1
Oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor (GO:0016616)	4.98E-06	0.001432	37.52801	ABCC4; ADH1C; HPGD; CYP4A22; AKR1D1; AKR1C3; DHRS2; ADH6; ADH4; GRHPR; AKR7A3; BDH2; FAM213B; AKR1B10; GPD1; ME1; RDH16; RDH5; PHGDH; GLYR1
Protein homodimerization activity (GO:0042803)	5.49E-06	0.001264	19.98055	TOP2A; CDA; TENM1; HSP90AB1; HIP1; UXS1; HEXB; HSPB1; PTH1R; EPRS; JCHAIN; GYS2; GHR; KYNU; MAP3K9; HIF1AN; PSPH; PDGFRA; SDS; TPM4; VWF; TPM1; TAP1; ADAM10; APOA4; HOGA1; MUT; DCK; FOXP2; TOX3; ALDH1A3; MSH2; CDH11; CAT; PECAM1; MASP1; PLCB1; AGXT; ATF3; MGLL; ABCG2; CDH19; CAMK2B; PON3; HPGD; SHMT1; PON1; PDGFA; ADRB2; TMEM27; INPP5F; MTHFD1L; IRAK1; ERBB3; CBS; TPR; KCNN2; ECT2; XDH; S100A10; UGT1A6; BARD1; AOC1; GCH1; STAT1; AADAT; EPHX2; ASNS; GSTZ1; NR4A2; GRHPR; CENPF; CD4; TP53I3; NR4A3; GSTA4; APOC2; BAX; GLA; CD200; MAD2L1
Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen (GO:0016712)	8.38E-06	0.001608	67.74414	CYP2C9; CYP2A7; CYP2A6; CYP2C8; CYP2B6; CYP2A13; CYP1A2; CYP2E1; CYP2C19
Heme binding (GO:0020037)	1.47E-05	0.002416	35.53057	PTGIS; CBSL; CYP2C19; PTGS2; CYP8B1; CYP2A7; CYP26A1; CYP2C9; CYP2A6; CYP2C8; CYP2B6; TDO2; CYP2A13; CBS; CAT; CYP1A2; CYP2E1
Transaminase activity (GO:0008483)	2.57E-05	0.003703	71.45018	OAT; AADAT; AGXT2; PSAT1; BCAT1; AGXT; ETNPPL
Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen (GO:0016709)	1.39E-04	0.017727	32.23068	SQLE; CYP26A1; FAXDC2; AKR1D1; CYP4A22; CYP4A11; AKR1C3; CYP2E1; KMO; CYP8B1; CYP4F12
Oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor (GO:0016620)	3.04E-04	0.034934	35.05117	ADH4; ALDH1A3; ALDH6A1; AKR1B10; ALDH1B1; AKR1C3; ALDH18A1; ALDH8A1
Cellular component				
Secretory granule lumen (GO:0034774)	1.74E-08	7.75E-06	40.3977	CDA; ECM1; HSP90AB1; HEXB; SERPINE1; DBH; CLU; SERPINA4; PNP; GM2A; PGM2; ACAA1; CTSC; SERPINB1; ANXA2; VWF; HGF; NME2; GTPBP2; ILF2; PSMA5; ACLY; PKM; CRISPLD2; CAT; CHI3L1; S100A8; CFD; CSTB; PSMD14; MVP; PDGFA; PLG; CFP; GNS; THBS1; PLAC8; NEU1; AGA; CTSA; DYNC1H1; AOC1; GSN; AGL; IGF2; GGH; IGF1; ARPC5; LYZ; FABP5; ALB; LCN2; GLA

Membrane attack complex (GO:0005579)	1.24E-05	0.00276	127.336	C6; C7; C9; C8B; C8A
Spindle (GO:0005819)	4.23E-05	0.006282	21.96729	KIF14; BUB1B; CAPG; TTK; KIF11; AURKA; CDC20; CCNB1; RACGAP1; TPR; NUSAP1; SPDL1; ECT2; CDK5RAP2; RMDN2; CDC7; CDC6; TUBG1; CKAP5; MZT1; TPX2; CENPF; NCOR1; KIF2A; PRC1; KIF4A; CDK1; KIF20A; MAD1L1; MAD2L1
Condensed nuclear chromosome kinetochore (GO:0000778)	5.41E-05	0.006028	72.47135	CCNB1; NUF2; BUB1B; CENPA; BUB1; NDC80
Condensed nuclear chromosome, centromeric region (GO:0000780)	1.01E-04	0.009041	62.18329	NUF2; BUB1B; CENPA; BUB1; NDC80; AURKA
Chromosome, centromeric region (GO:0000775)	1.80E-04	0.013408	28.54812	HELLS; CENPF; CENPW; MIS18A; H2AFY; RAD21; NUF2; BIRC5; NCAPD2; SMC3; CENPA; NDC80
Cytoplasmic vesicle lumen (GO:0060205)	1.88E-04	0.011995	19.78179	CFD; CDA; CSTB; SERPINB1; HSP90AB1; GSN; PSMD14; MVP; AGL; NME2; HBB; ARPC5; DBH; HBA1; PSMA5; PNP; PKM; CRISPLD2; CAT; PGM2; HYOU1; S100A8
Endoplasmic reticulum lumen (GO:0005788)	2.04E-04	0.011395	15.74304	COL15A1; PROZ; PDGFA; LAMC1; CFP; PTGS2; THBS1; ADAMTS13; GPC3; SPP1; ARSD; APOL1; ARSE; CTSC; ERAP2; GOLM1; IGFBP3; P3H2; ADAM10; APOA4; PDIA6; APOA5; CKAP4; COL1A1; COL1A2; F9; COL4A2; COL4A1; P4HA2; CANX; DNAJB11; DNAJC10; ALB; CALU; SPARCL1; HYOU1; PPIB
Intercalated disc (GO:0014704)	2.92E-04	0.014493	31.94614	GJC1; GJA1; TMEM65; FXYP1; CTNNA3; ANK3; FGF13; VAMP5; SLC8A1
Mitotic spindle (GO:0072686)	3.86E-04	0.017203	20.24638	CAPG; CDC7; RMDN2; KIF11; SMC3; AURKA; ASPM; TPX2; RACGAP1; TPR; NUSAP1; CDK1; ECT2; FAM83D; MAD1L1; MAD2L1

Table 3. The top 10 KEGG pathway enrichment analysis of DEGs.

Term	P-value	Adjusted P-value	Combined Score	Genes
Chemical carcinogenesis	9.12E-11	2.81E-08	99.12267	ADH1C; ADH1B; GSTT2; CYP3A4; CYP2C19; PTGS2; ADH6; HSD11B1; ADH4; CYP2A13; GSTT2B; UGT1A6; CYP3A7-CYP3A51P; CYP3A43; UGT2B15; UGT2B17; ALDH1A3; CYP2C9; CYP2A6; CYP2C8; GSTA4; GSTA3; NAT2; CYP1A2; CYP1A1; CYP2E1
Cell cycle	2.67E-10	4.11E-08	76.92983	PCNA; PRKDC; BUB1B; TTK; SMC3; CDC20; CCNB2; CCNB1; PTTG1; RAD21; SFN; E2F3; BUB1; SMAD2; CDKN2B; CDKN2C; GADD45B; CDKN2A; CDC7; CDC6; GADD45G; CCNA2; DBF4; CCNE2; CDK1; MCM3; MCM4; MCM5; MCM6; MAD1L1; MCM2; MAD2L1
Drug metabolism	3.08E-09	3.16E-07	68.70754	CDA; ADH1C; ADH1B; GSTT2; CYP2C19; CYP3A4; ADH6; NME1-NME2; ADH4; CYP2B6; UPP2; GSTT2B; XDH; UGT1A6; RRM2; UGT2B15; NME2; UGT2B17; NME1; CYP2C9; ALDH1A3; CYP2C8; CYP2A6; GSTA4; NAT2; GSTA3; CYP1A2; CYP2E1
Retinol metabolism	7.33E-09	5.64E-07	79.39321	CYP3A7-CYP3A51P; ADH1C; ADH1B; CYP4A22; CYP4A11; UGT2B15; UGT2B17; CYP3A4; ADH6; ADH4; CYP26A1; CYP2C9; CYP2A6; CYP2C8; CYP2B6; CYP1A2; RDH16; CYP1A1; RDH5; PNPLA4; UGT1A6

Glycine, serine and threonine metabolism	8.44E-09	5.20E-07	100.5555	DMGDH; ALAS1; SDS; AGXT2; SHMT1; CBSL; GNMT; GRHPR; BHMT; CBS; PSAT1; CTH; SARDH; PHGDH; AGXT; PSPH
Metabolism of xenobiotics by cytochrome P450	9.71E-09	4.98E-07	74.17367	ADH1C; ADH1B; UGT2B15; UGT2B17; GSTT2; CYP3A4; ADH6; HSD11B1; ALDH1A3; ADH4; CYP2C9; CYP2A6; AKR7A3; CYP2B6; CYP2A13; GSTA4; GSTA3; CYP1A2; CYP1A1; CYP2E1; GSTT2B; UGT1A6
p53 signaling pathway	3.05E-08	1.34E-06	68.25618	RRM2; GADD45B; CDKN2A; GORAB; IGFBP3; SERPINE1; SIAH1; IGF1; THBS1; GADD45G; CCNB2; CCNB1; TP53I3; CCNE2; ZMAT3; CCNG2; CDK1; BAX; SFN; MDM4; GTSE1
Fatty acid degradation	4.21E-08	1.62E-06	83.50768	GCDH; ACAA2; ADH1C; ADH1B; ACSL1; CYP4A22; CYP4A11; ACSL5; ACSL4; ACADSB; ADH6; ADH4; ACADL; ALDH1B1; ACAA1; ACADS
PPAR signaling pathway	5.18E-08	1.77E-06	64.37895	ACSL1; CYP4A22; CYP4A11; ACSL5; LPL; ACSL4; APOA5; CYP8B1; FABP4; ACADL; FABP5; SCD; ME1; PLIN4; PLIN2; HMGCS2; ANGPTL4; ACAA1; PCK1; SLC27A5; PCK2
Tryptophan metabolism	1.01E-06	3.12E-05	62.21888	GCDH; AOC1; AADAT; HAAO; KMO; ACMSD; TDO2; ALDH1B1; KYNU; CAT; CYP1A2; CYP1A1; ALDH8A1; IDO2

PPI-network reconstruction of DEGs and network clustering

PPI network of DEGs were reconstructed utilizing STRING server. Cytoscape version 3.6.0 was used to visualize, analysis and clustering network into structural modules. Total nodes were 1236 and total edges were 11624 in the constructed PPI network. 594 nodes were belonging to up-regulated DEGs and 562 nodes were related to down-regulated DEGs. Top 3 structural modules are shown in Table 4. GO term (Table 5) and KEGG pathway enrichment analysis (Table 6) were performed for each module. PPI network of DEGs and top 3 modules are represented in Figure 5. Top 15 hub genes based on their associated degree are listed in Table 7.

Table 4. Top 3 structural modules.

modules	Score	Nodes	Edges	Genes
1	70.263	77	2670	DTL, CEP55, BUB1B, CKAP2, ANLN, MCM4, MCM6, ECT2, RACGAP1, KIF20A, FOXM1, BUB1, CASC5, MAD2L1, NEK2, MCM2, PTTG1, CDC7, NUSAP1, NDC80, NUF2, CCNB2, PBK, CDC20, PRC1, TOP2A, EZH2, HMMR, GTSE1, UBE2C, KIAA0101, FANCI, UBE2T, CDC6, CDCA3, RRM2, ATAD2, ASPM, NCAPD2, MELK, TTK, NCAPG, CENPA, MCM5, CKS2, TRIP13, KIF14, GMNN, MCM3, UHRF1, BIRC5, FEN1, GINS1, CDKN3, RFC4, CENPF, POLE2, KPNA2, AURKA, TPX2, HELLS, CENPU, KIF11,

				CCNE2, SMC2, SMC4, MKI67, ZWINT, CCNA2, TYMS, KIF4A, RAD51AP1, CCNB1, DLGAP5, CDK1, PCNA, FAM83D
2	13	13	78	FABP5, ACLY, LYZ, CTSC, ANXA2, AGA, PLAC8, GLA, CTSA, GNS, GGH, HEXB, GM2A
3	9.714	29	136	SPP2, APOA5, LAMC1, C6, ELN, PRKDC, FTCD, PDIA6, APOL1, GOLM1, GPC3, F11, IGFBP3, FETUB, THBS1, PSMD4, CDKN2A, CKAP4, ADAM10, PAH, C8A, SPARCL1, KLKB1, SERPINE1, APOA4, CALU, HIST1H2BK, HIST1H2AC, HGFAC

Figure 4. Top 10 GO term and KEGG pathways enrichment analysis,

a) GO enrichment analysis – Biological process

b) GO enrichment analysis – Molecular function

c) GO enrichment analysis – Cellular component

d) KEGG enrichment analysis for DEGs

Figure 5. PPI network of DEGs and top 3 modules. A) PPI network of DEGs. B) module 1. C) module 2. D) module 3.

A)

					MCM4; MCM5; MCM6; KPNA2; MCM2
	G1/S transition of mitotic cell cycle (GO:0000082)	1.49E-16	2.53E-13	1171.996	PCNA; RRM2; CDC7; CDC6; TYMS; CCNE2; POLE2; MCM3; MCM4; MCM5; MCM6; MCM2; CDKN3
	Regulation of mitotic cell cycle phase transition (GO:1901990)	9.56E-15	1.22E-11	637.9586	UBE2C; KIF14; BUB1B; HMMR; CDC6; AURKA; CDC20; ANLN; TPX2; CCNB1; CENPF; CDK1; NEK2; MAD2L1
	Mitotic sister chromatid segregation (GO:0000070)	1.47E-14	1.50E-11	1109.705	CCNB1; PRC1; KIF14; NUSAP1; NCAPG; NCAPD2; SMC4; DLGAP5; NDC80; ZWINT; SMC2
Molecular function	Kinase binding (GO:0019900)	6.59E-09	7.58E-06	152.1722	KIF14; KIF11; CDC6; FOXM1; AURKA; CCNA2; TPX2; CCNB1; RACGAP1; PRC1; CKS2; KIF20A; FAM83D
	Protein kinase binding (GO:0019901)	4.85E-08	2.79E-05	114.8905	TOP2A; KIF14; KIF11; FOXM1; AURKA; CCNA2; TPX2; CCNB1; RACGAP1; PRC1; CKS2; KIF20A; FAM83D
	Microtubule binding (GO:0008017)	8.37E-07	3.21E-04	149.1109	RACGAP1; KIF4A; KIF14; NUSAP1; BIRC5; KIF11; KIF20A; FAM83D
	Histone kinase activity (GO:0035173)	4.53E-06	0.001304	1065.3	MELK; CDK1; AURKA
	Tubulin binding (GO:0015631)	6.16E-06	0.001419	97.75949	RACGAP1; KIF4A; KIF14; NUSAP1; BIRC5; KIF11; KIF20A; FAM83D
Cellular Components	Spindle (GO:0005819)	3.13E-22	1.40E-19	1313.765	KIF14; BUB1B; TTK; CDC7; KIF11; CDC6; AURKA; CDC20; TPX2; CCNB1; CENPF; RACGAP1; PRC1; KIF4A; NUSAP1; CDK1; KIF20A; ECT2; MAD2L1
	Nuclear chromosome part (GO:0044454)	1.18E-18	2.63E-16	574.4269	GIN51; FEN1; PCNA; UHRF1; BUB1B; CENPA; NDC80; AURKA; RAD51AP1; CCNB1; POLE2; NUF2; MCM3; CDK1; MCM4; MCM5; NCAPD2; MCM6; BUB1; MCM2; EZH2
	Mitotic spindle (GO:0072686)	1.94E-14	2.89E-12	1073.88	ASPM; TPX2; RACGAP1; NUSAP1; CDK1; CDC7; KIF11; ECT2; FAM83D; AURKA; MAD2L1
	Microtubule cytoskeleton (GO:0015630)	9.35E-14	1.04E-11	341.4204	CENPU; KIF14; CKAP2; TTK; KIF11; HMMR; AURKA; CDC20; CCNB2; TPX2; CENPF; PRC1; KIF4A; KIF20A; CEP55; DLGAP5; MCM2

				Condensed nuclear chromosome kinetochore (GO:0000778)	1.21E-12	1.08E-10	3887.177	CCNB1; NUF2; BUB1B; CENPA; BUB1; NDC80
2	13	78	Biological process	Neutrophil degranulation (GO:0043312)	7.26E-22	3.70E-18	2032.348	CTSA; ANXA2; HEXB; GGH; LYZ; GNS; PLAC8; ACLY; FABP5; GM2A; AGA; GLA; CTSC
				Neutrophil activation involved in immune response (GO:0002283)	8.10E-22	2.07E-18	2010.983	CTSA; ANXA2; HEXB; GGH; LYZ; GNS; PLAC8; ACLY; FABP5; GM2A; AGA; GLA; CTSC
				Neutrophil mediated immunity (GO:0002446)	9.03E-22	1.54E-18	1990.007	CTSA; ANXA2; HEXB; GGH; LYZ; GNS; PLAC8; ACLY; FABP5; GM2A; AGA; GLA; CTSC
				Glycosphingolipid metabolic process (GO:0006687)	3.33E-08	4.25E-05	1962.072	CTSA; GM2A; HEXB; GLA
				Glycolipid metabolic process (GO:0006664)	7.10E-08	7.25E-05	1558.338	CTSA; GM2A; HEXB; GLA
			Molecular function	Exopeptidase activity (GO:0008238)	3.80E-04	0.437457	538.4726	CTSA; GGH
				Exo-alpha-sialidase activity (GO:0004308)	0.003894	1	1422.639	CTSA
				Alpha-sialidase activity (GO:0016997)	0.003894	1	1422.639	CTSA
				Omega peptidase activity (GO:0008242)	0.004542	1	1185.592	GGH
				Galactosidase activity (GO:0015925)	0.005189	1	1011.771	GLA
			Cellular Components	Azurophil granule lumen (GO:0035578)	1.25E-31	5.58E-29	15812.48	CTSA; ANXA2; HEXB; GGH; LYZ; GNS; PLAC8; ACLY; FABP5; GM2A; AGA; GLA; CTSC
				Azurophil granule (GO:0042582)	1.99E-28	4.44E-26	8283.607	CTSA; ANXA2; HEXB; GGH; LYZ; GNS; PLAC8; ACLY; FABP5; GM2A; AGA; GLA; CTSC
				Vacuolar lumen (GO:0005775)	3.63E-28	5.40E-26	7848.771	CTSA; ANXA2; HEXB; GGH; LYZ; GNS; PLAC8; ACLY; FABP5; GM2A; AGA; GLA; CTSC
				Secretory granule lumen (GO:0034774)	3.11E-24	3.47E-22	3414.919	CTSA; ANXA2; HEXB; GGH; LYZ; GNS; PLAC8; ACLY; FABP5; GM2A; AGA; GLA; CTSC
				Lysosome (GO:0005764)	4.32E-11	3.85E-09	696.0632	CTSA; ANXA2; GM2A; HEXB; AGA; GNS; GLA; CTSC
3	29	136	Biological process	Post-translational protein modification (GO:0043687)	3.44E-14	1.75E-10	718.6612	PSMD4; GOLM1; IGFBP3; GPC3; CALU; ADAM10; SPARCL1; LAMC1;

					APOL1; APOA5; PDIA6; CKAP4
	Cellular protein metabolic process (GO:0044267)	3.97E-14	1.01E-10	571.6101	PRKDC; GOLM1; IGFBP3; ADAM10; APOA4; LAMC1; APOA5; PDIA6; CKAP4; GPC3; CALU; SPARCL1; APOL1
	Cellular protein modification process (GO:0006464)	2.17E-11	3.69E-08	236.8335	PRKDC; GOLM1; IGFBP3; ADAM10; LAMC1; THBS1; APOA5; PDIA6; CKAP4; PSMD4; GPC3; CALU; SPARCL1; APOL1
	Regulation of fibrinolysis (GO:0051917)	2.52E-09	3.22E-06	4200.982	F11; SERPINE1; THBS1; KLKB1
	Extracellular matrix organization (GO:0030198)	8.04E-07	8.21E-04	253.5776	ELN; SERPINE1; ADAM10; LAMC1; THBS1; KLKB1
Molecular function	Phosphatidylcholine-sterol O-acyltransferase activator activity (GO:0060228)	3.03E-05	0.034922	2391.501	APOA4; APOA5
	Cholesterol transporter activity (GO:0017127)	2.72E-04	0.156756	665.988	APOA4; APOA5
	Sterol transporter activity (GO:0015248)	7.50E-04	0.287603	354.4801	APOA4; APOA5
	Peptidase inhibitor activity (GO:0030414)	0.00124	0.356931	256.4102	GPC3; FETUB
	Endopeptidase activity (GO:0004175)	0.001615	0.371838	50.09349	F11; ADAM10; KLKB1; HGFAC
	Endoplasmic reticulum lumen (GO:0005788)	2.07E-17	9.24E-15	1275.603	GOLM1; IGFBP3; ADAM10; APOA4; LAMC1; THBS1; APOA5; PDIA6; CKAP4; GPC3; CALU; SPARCL1; APOL1
Cellular Components	Very-low-density lipoprotein particle (GO:0034361)	1.23E-06	2.75E-04	1876.755	APOA4; APOL1; APOA5
	Membrane attack complex (GO:0005579)	3.03E-05	0.004511	2391.501	C6; C8A
	Rough endoplasmic reticulum (GO:0005791)	7.50E-04	0.083582	354.4801	HGFAC; CKAP4
	Platelet alpha granule lumen (GO:0031093)	0.004234	0.377665	112.4988	SERPINE1; THBS1

Table 6. KEGG enrichment analysis of top 3 modules.

module	node	edges	Term	P-value	Adjusted P-value	Combined Score	Genes
1	77	2670	Cell cycle	1.02E-25	3.13E-23	2290.387	PCNA; BUB1B; TTK; CDC7; CDC6; CDC20; CCNA2; CCNB2; CCNB1; PTTG1; CCNE2; MCM3; CDK1; MCM4; MCM5; MCM6; BUB1; MCM2; MAD2L1
			DNA replication	9.92E-15	1.53E-12	2093.77	FEN1; PCNA; RFC4; POLE2; MCM3; MCM4; MCM5; MCM6; MCM2
			Oocyte meiosis	1.23E-09	1.26E-07	383.732	CDC20; CCNB2; CCNB1; PTTG1; CCNE2; CDK1; BUB1; AURKA; MAD2L1
			Progesterone-mediated oocyte maturation	1.06E-07	8.19E-06	294.8769	CCNA2; CCNB2; CCNB1; CDK1; BUB1; AURKA; MAD2L1
			p53 signaling pathway	3.41E-07	2.10E-05	322.3209	CCNB2; CCNB1; RRM2; CCNE2; CDK1; GTSE1
2	13	78	Lysosome	4.66E-13	1.43E-10	2486.122	CTSA; GM2A; HEXB; AGA; GNS; GLA; CTSC
			Other glycan degradation	5.93E-05	0.009135	1663.683	HEXB; AGA
			Glycosaminoglycan degradation	6.63E-05	0.006804	1558.168	HEXB; GNS
			Glycosphingolipid biosynthesis	3.80E-04	0.029265	538.4726	HEXB; GLA
			Renin-angiotensin system	0.01485157	0.914857	281.5819	CTSA
3	29	136	Complement and coagulation cascades	9.32E-08	2.87E-05	706.5978	C6; SERPINE1; F11; C8A; KLKB1
			p53 signaling pathway	3.42E-06	5.27E-04	482.1627	CDKN2A; IGFBP3; SERPINE1; THBS1
			Prion diseases	1.74E-05	0.001785	647.869	C6; LAMC1; C8A
			Systemic lupus erythematosus	3.90E-05	0.003004	210.5579	C6; HIST1H2BK; C8A; HIST1H2AC
			Cellular senescence	0.00157588	0.097074	83.44321	CDKN2A; IGFBP3; SERPINE1

Table 7. Top 15 hub genes in regard to their related degree.

Gene	Degree
ALB	187
CDK1	156
CCNB1	136
CCNA2	128
CDC20	124
AURKA	122
MAD2L1	118
TOP2A	117
KIF11	116
CCNB2	115
BUB1B	115
TYMS	110
EZH2	109
BUB1	108
ESR1	107

Hub genes PPI network and enrichment analysis

Top 15 hub genes considering their degree are ALB (albumin), CDK1 (Cyclin Dependent Kinase 1), CCNB1 (Cyclin B1), CCNA2 (Cyclin A2), CDC20 (Cell Division Cycle 20), AURKA (Aurora Kinase A), MAD2L1 (Mitotic Arrest Deficient 2 Like 1), TOP2A (DNA Topoisomerase II Alpha), KIF11 (Kinesin Family Member 11), CCNB2 (Cyclin B2), BUB1B (BUB1B Mitotic Checkpoint Serine/Threonine Kinase B), TYMS (Thymidylate Synthetase), EZH2 (Enhance Of Zeste 2 Polycomb Repressive Complex 2 Subunit), BUB1 (BUB1 Mitotic Checkpoint Serine/Threonine Kinase), ESR1 (Estrogen Receptor 1). (table 7)

PPI network of 15 hug genes reconstructed using STRING (Figure 6). Additionally, GO term (Table 8) and KEGG pathway (Table 9) enrichment analysis for 15 hub genes performed by enrichr server.

Figure 6. PPI network of 15 hub genes.

Table 8. Top 5 GO term enrichment analysis of 15 Hub genes.

GO enrichment analysis	Term	P-value	Adjusted P-value	Combined Score	Genes
Biological process	anaphase-promoting complex-dependent catabolic process (GO:0031145)	1.52E-11	7.77E-08	2522.3532	CDC20; CCNB1; CDK1; BUB1B; AURKA; MAD2L1
	negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle (GO:0051436)	1.43E-09	3.64E-06	1912.4721	CDC20; CCNB1; CDK1; BUB1B; MAD2L1
	regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle (GO:0051439)	1.53E-09	2.61E-06	1879.2843	CDC20; CCNB1; CDK1; BUB1B; MAD2L1
	positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition (GO:0051437)	2.02E-09	2.58E-06	1756.1385	CDC20; CCNB1; CDK1; BUB1B; MAD2L1
	negative regulation of ubiquitin protein ligase activity (GO:1904667)	2.16E-09	2.20E-06	1727.5547	CDC20; CCNB1; CDK1; BUB1B; MAD2L1
Molecular function	cyclin-dependent protein kinase activity (GO:0097472)	9.37E-09	1.08E-05	2899.6898	CCNA2; CCNB2; CCNB1; CDK1
	protein kinase binding (GO:0019901)	9.23E-07	5.31E-04	224.57388	TOP2A; CCNA2; CCNB1; KIF11; ESR1; AURKA

Cellular Components	cyclin-dependent protein serine/threonine kinase activity (GO:0004693)	1.84E-06	7.05E-04	1600.894	CCNB2; CCNB1; CDK1
	kinase binding (GO:0019900)	9.84E-06	0.00283	183.88355	CCNA2; CCNB1; KIF11; ESR1; AURKA
	histone kinase activity (GO:0035173)	1.88E-05	0.004338	3223.5228	CDK1; AURKA
	spindle (GO:0005819)	3.25E-11	1.45E-08	1211.8807	CDC20; CCNB1; CDK1; BUB1B; KIF11; AURKA; MAD2L1
	nuclear chromosome part (GO:0044454)	5.92E-09	1.32E-06	451.06908	CCNB1; CDK1; BUB1B; ESR1; BUB1; AURKA; EZH2
	condensed nuclear chromosome kinetochore (GO:0000778)	5.61E-08	8.34E-06	6071.289	CCNB1; BUB1B; BUB1
	condensed nuclear chromosome, centromeric region (GO:0000780)	7.48E-08	8.34E-06	5469.6028	BUB1B; BUB1; AURKA
	mitotic spindle (GO:0072686)	3.81E-07	3.40E-05	938.36615	CDK1; KIF11; AURKA; MAD2L1

Table 9. top 5 KEGG pathway enrichment analysis of 15 Hub genes

Term	P-value	Adjusted P-value	Combined Score	Genes
Cell cycle	1.08E-14	3.32E-12	2766.618	CCNA2; CDC20; CCNB2; CCNB1; CDK1; BUB1B; BUB1; MAD2L1
Progesterone-mediated oocyte maturation	3.65E-13	5.63E-11	2699.841	CCNA2; CCNB2; CCNB1; CDK1; BUB1; AURKA; MAD2L1
Oocyte meiosis	1.94E-12	1.99E-10	2013.641	CDC20; CCNB2; CCNB1; CDK1; BUB1; AURKA; MAD2L1
Human T-cell leukemia virus 1 infection	4.13E-07	3.18E-05	447.4824	CCNA2; CDC20; CCNB2; BUB1B; MAD2L1
Cellular senescence	5.03E-06	3.10E-04	406.6895	CCNA2; CCNB2; CCNB1; CDK1

Survival analysis of hub genes

Each of hub genes was submitted to GEPIA in order to obtain survival plots.

Figure 7. The overall survival of 15 hub genes in LIHC patients with median cutoff. (a) ALB, (b) CDK1, (c) CCNB1, (d) CCNA2, (e) CDC20, (f) AURKA, (g) MAD2L1, (h) TOP2A, (i) KIF11, (j) CCNB2, (k) BUB1B, (l) TYMS, (m) EZH2, (n) BUB1, (o) ESR1.

Figure 8. Disease-free Survival (DFS) of 15 hub genes in LIHC patients with median cutoff. (a) ALB, (b) CDK1, (c) CCNB1, (d) CCNA2, (e) CDC20, (f) AURKA, (g) MAD2L1, (h) TOP2A, (i) KIF11, (j) CCNB2, (k) BUB1B, (l) TYMS, (m) EZH2, (n) BUB1, (o) ESR1.

Discussion

Hepatocellular carcinoma (HCC) is one of the leading causes of death by cancer (16). Due to the important role of biomarkers in diagnosis of HCC, various biomarkers have been introduced in the recent years (17). Despite all efforts in developing diagnosis methods and treatment strategy for HCC patients, there is still shortcomings in this area (18). Consequently, biomarkers with high sensitivity and specificity need in precise detection of HCC.

Integration of network-based approach with microarray has been resulted in the emergence of a robust strategy to find potential biomarkers in various cancers specially HCC (19). In our study, firstly, a gene expression profile from gene expression omnibus (GEO) with accession number GSE62232 was downloaded. The downloaded profile consisted of 81 HCC samples and 10 normal samples. Following normalization of CELL files with Transcriptome Analysis Console (TAC) software, 1996 DEGs were identified including 995 up-regulated and 1001 down-regulated genes.

GO term and KEGG pathway enrichment analysis were also performed to achieve more information about DEGs. Afterward PPI network reconstruction and determining structural modules with Cytoscape, based on highest degree, 15 hub genes were selected as the HCC candidates' biomarkers. Ultimately, OS and DFS plots using GEPIA were applied to validate hub genes.

Comparing enrichment analysis of DEGs and top 3 modules showed that module 1 is more important than other modules considering its two common pathways with top 10 KEGG pathways of DEGs (cell cycle, and p53 signaling pathway). There was no common GO biological process and GO molecular function between top 5 GO bp and GO mf of the selected modules and top 10 GO bp and GO mf of DEGs. However, module 1 based on GO cellular component results showed to be more important in compare to other modules (common GO cellular components are spindle, mitotic spindle, condensed nuclear chromosome kinetochore).

Final hub genes based on the highest degree are ALB, CDK1, CCNB1, CCNA2, CDC20, AURKA, MAD2L1, TOP2A, KIF11, CCNB2, BUB1B, TYMS, EZH2, BUB1, ESR1. Interestingly, all hub genes are present in module 1 except ALB and ESR1. Common GO terms between hub genes and DEGs are: I) 'cell cycle' in KEGG pathway; II) 'spindle', 'condensed nuclear chromosome kinetochore', 'condensed nuclear chromosome, centromeric region', and 'mitotic spindle' in GO cellular components. No common in GO bp and GO mf was seen.

Overall survival (OS) and disease-free survival (DFS) studies were performed using GEPIA in order to validate final hub genes. Common genes between OS and DFS which their up-regulation is related to poor prognosis of HCC were considered as final biomarkers. These genes are CDK1, CCNB1, CCNA2, CDC20, AURKA, MAD2L1, TOP2A, KIF11, BUB1B, TYMS, EZH2, BUB1.

Therefore, a literature review was done to obtain more details about these genes in different cancers.

CDK1 (Cyclin Dependent Kinase 1) gene belongs to serine/threonine protein kinase family (20). CCNB1 encodes G2/mitotic-specific cyclin-B1 and based on our result its up-regulation is related with low survival in HCC patients. The potential role of CDK1 and CCNB1 inhibition in increasing the efficacy of HCC treatment have also been investigated (21, 22). CCNA2 encodes Cyclin A2 and the regulation of CCNA2 using RNA interference has been shown recently (23). A protein produced from CDC20 (Cell Division Cycle 20) interacts with anaphase-promoting complex/cyclosome (APC/C) in the cell cycle. Up-regulation of CDC20 has been reported in various cancers, including oral squamous cell carcinoma (24), and gastric cancer (25). Additionally, the effects of CDC20 up-regulation in the progression of HCC have been examined (26). Overexpression of BUB1B (also termed BUBR1) which encodes mitotic checkpoint serine/threonine-protein kinase BUB1 beta are associated with poor prognosis in HCC (27). BUB1 encodes mitotic checkpoint serine/threonine-protein kinase and recently its role to maintain breast cancer stem cell has been studied (28). MAD2L1 encodes mitotic spindle assembly checkpoint protein MAD2A, its suppression using RNA interferences has been shown to control the proliferation and metastasis in HCC (29). CDK1, CCNB1, CCNA2, CDC20, MAD2L1, BUB1B, and BUB1 are enriching cell cycle. Interestingly, cell cycle is among top KEGG pathways which is enriched with both module 1 and DEGs.

AURKA encodes Aurora Kinase A protein and due to our findings, AURKA enriches progesterone-mediated oocyte maturation and oocyte meiosis in KEGG pathways. A study suggests that variation in AURKA gene is important to predict early-stage of HCC, and it is a reliable biomarker for HCC (30). DNA topoisomerase 2-alpha is a protein produced from TOP2A;

it has been suggested that TOP2A overexpression in HCC patients is a potential candidate for therapeutic purposes (31) and based on GO molecular function it is related to protein kinase binding. KIF11 encodes Kinesin-like protein KIF11 and due to results of GO cellular component it is related to spindle and mitotic spindle. Recently, a study demonstrated that kinesin family members, including KIF11 are potential markers to predict poor prognosis and cell proliferation in HCC (32). TYMS encode thymidylate synthetase protein and through a study it was shown that TYMS related polymorphisms are valuable factors in predicting clinical outcomes of HCC (33). EZH2 encode Histone-lysine N-methyltransferase protein and considering GO cellular component enrichment analysis, it is related to nuclear chromosome part. It has shown that EZH2 through modulation of miR-22/galectin-9 axis can cause progression of HCC (34).

Conclusion

In conclusion, our findings alongside recent studies validate the importance of CDK1, CCNB1, CCNA2, CDC20, AURKA, MAD2L1, TOP2A, KIF11, BUB1B, TYMS, EZH2, BUB1 to be considered as potential biomarkers in hepatocellular carcinoma. Based on our findings, we demonstrate that network-based approach integrated with microarray data is a promising method to obtain potential therapeutic and diagnostic biomarkers in various cancers, in our case, hepatocellular carcinoma. However, further experimental analysis is needed to confirm our findings.

Abbreviations

HCC: Hepatocellular carcinoma

EGFR: Epidermal growth factor receptor

TGF- β : Transforming growth factor-beta 1

DEGs: Differentially expressed genes

GEO: Gene expression omnibus

GO: Gene ontology

KEGG: Kyoto encyclopedia of genes and genomes

PPI-network: Protein-protein interaction network

TAC: Transcriptome analysis console

BP: Biological process

MF: Molecular function

CC: Cellular component

GEPIA: Gene expression profiling interactive analysis

LIHC: Liver hepatocellular carcinoma

OS: Overall survival

DFS: Disease-free survival

PCA: Principal components analysis

ALB: albumin

CDK1: Cyclin Dependent Kinase 1

CCNB1: Cyclin B1

CCNA2: Cyclin A2

CDC20: Cell Division Cycle 20

AURKA: Aurora Kinase A

MAD2L1: Mitotic Arrest Deficient 2 Like 1

TOP2A: DNA Topoisomerase II Alpha

KIF11: Kinesin Family Member 11

CCNB2: Cyclin B2

BUB1B: BUB1B Mitotic Checkpoint Serine/Threonine Kinase B

TYMS: Thymidylate Synthetase

EZH2: Enhance of Zeste 2 Polycomb Repressive Complex 2 Subunit

BUB1: BUB1 Mitotic Checkpoint Serine/Threonine Kinase

ESR1: Estrogen Receptor 1

Acknowledgements

The authors wish to thank Shiraz University of Medical Sciences for supporting the conduct of this research.

Contributions

Mehrdad Ameri, Haniye Salimi, Sedigheh Eskandari, Navid Nezafat participated in data analysis and writing the manuscript, discussing data and supervised the study; and all authors performed data analysis and interpretation and read and approved the final manuscript.

Funding

No funding.

Ethics approval and consent to participate**Availability of data and materials**

The datasets analyzed during the current study are available in the GEO repository, (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62232>).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References:

1. El-Serag HB. Epidemiology of hepatocellular carcinoma. *The Liver: Biology and Pathobiology*. 2020;758-72.
2. Ho DW, Lo RC, Chan LK, Ng IO. Molecular Pathogenesis of Hepatocellular Carcinoma. *Liver Cancer*. 2016;5(4):290-302.
3. Jang JW, Song Y, Kim SH, Kim JS, Kim KM, Choi EK, et al. CD133 confers cancer stem-like cell properties by stabilizing EGFR-AKT signaling in hepatocellular carcinoma. *Cancer Lett*. 2017;389:1-10.
4. Huang J, Qiu M, Wan L, Wang G, Huang T, Chen Z, et al. TGF- β 1 Promotes Hepatocellular Carcinoma Invasion and Metastasis via ERK Pathway-Mediated FGFR4 Expression. *Cellular Physiology and Biochemistry*. 2018;45(4):1690-9.
5. Nevzorova YA, Hu W, Cubero FJ, Haas U, Freimuth J, Tacke F, et al. Overexpression of c-myc in hepatocytes promotes activation of hepatic stellate cells and facilitates the onset of liver fibrosis. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. 2013;1832(10):1765-75.
6. Ocker M. Biomarkers for hepatocellular carcinoma: What's new on the horizon? *World J Gastroenterol*. 2018;24(35):3974-9.
7. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207-10.
8. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*. 2013;14:128.
9. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*. 2016;44(W1):W90-W7.
10. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*. 2000;25(1):25-9.
11. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*. 2018;47(D1):D330-D8.
12. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. 2000;28(1):27-30.
13. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*. 2018;47(D1):D607-D13.
14. 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 research*. 2003;13(11):2498-504.
15. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*. 2003;4:2.
16. Rawla P, Sunkara T, Muralidharan P, Raj JP. Update in global trends and aetiology of hepatocellular carcinoma. *Contemp Oncol (Pozn)*. 2018;22(3):141-50.
17. Lou J, Zhang L, Lv S, Zhang C, Jiang S. Biomarkers for Hepatocellular Carcinoma. *Biomark Cancer*. 2017;9:1-9.

18. Okajima W, Komatsu S, Ichikawa D, Miyamae M, Ohashi T, Imamura T, et al. Liquid biopsy in patients with hepatocellular carcinoma: Circulating tumor cells and cell-free nucleic acids. *World J Gastroenterol*. 2017;23(31):5650-68.
19. Wang J, Zuo Y, Man Y-G, Avital I, Stojadinovic A, Liu M, et al. Pathway and network approaches for identification of cancer signature markers from omics data. *J Cancer* [Internet]. 2015 2015; 6(1):[54-65 pp.].
20. Malumbres M. Cyclin-dependent kinases. *Genome Biol*. 2014;15(6):122.
21. Wu CX, Wang XQ, Chok SH, Man K, Tsang SHY, Chan ACY, et al. Blocking CDK1/PDK1/ β -Catenin signaling by CDK1 inhibitor RO3306 increased the efficacy of sorafenib treatment by targeting cancer stem cells in a preclinical model of hepatocellular carcinoma. *Theranostics*. 2018;8(14):3737-50.
22. Gu J, Liu X, Li J, He Y. MicroRNA-144 inhibits cell proliferation, migration and invasion in human hepatocellular carcinoma by targeting CCNB1. *Cancer Cell International*. 2019;19(1):15.
23. Yang F, Gong J, Wang G, Chen P, Yang L, Wang Z. Waltonitone inhibits proliferation of hepatoma cells and tumorigenesis via FXR-miR-22-CCNA2 signaling pathway. *Oncotarget*. 2016;7(46):75165-75.
24. Moura IM, Delgado ML, Silva PM, Lopes CA, do Amaral JB, Monteiro LS, et al. High CDC20 expression is associated with poor prognosis in oral squamous cell carcinoma. *J Oral Pathol Med*. 2014;43(3):225-31.
25. Ding ZY, Wu HR, Zhang JM, Huang GR, Ji DD. Expression characteristics of CDC20 in gastric cancer and its correlation with poor prognosis. *Int J Clin Exp Pathol*. 2014;7(2):722-7.
26. Li J, Gao J-Z, Du J-L, Huang Z-X, Wei L-X. Increased CDC20 expression is associated with development and progression of hepatocellular carcinoma. *International journal of oncology*. 2014;45(4):1547-55.
27. Liu AW, Cai J, Zhao XL, Xu AM, Fu HQ, Nian H, et al. The clinicopathological significance of BUBR1 overexpression in hepatocellular carcinoma. *J Clin Pathol*. 2009;62(11):1003-8.
28. Han JY, Han YK, Park G-Y, Kim SD, Geun Lee C. Bub1 is required for maintaining cancer stem cells in breast cancer cell lines. *Scientific Reports*. 2015;5(1):15993.
29. Li Y, Bai W, Zhang J. MiR-200c-5p suppresses proliferation and metastasis of human hepatocellular carcinoma (HCC) via suppressing MAD2L1. *Biomed Pharmacother*. 2017;92:1038-44.
30. Wang B, Hsu CJ, Chou CH, Lee HL, Chiang WL, Su CM, et al. Variations in the AURKA Gene: Biomarkers for the Development and Progression of Hepatocellular Carcinoma. *Int J Med Sci*. 2018;15(2):170-5.
31. Panvichian R, Tantiwetrueangdet A, Angkathunyakul N, Leelaudomlipi S. TOP2A amplification and overexpression in hepatocellular carcinoma tissues. *Biomed Res Int*. 2015;2015:381602.
32. Li X, Huang W, Huang W, Wei T, Zhu W, Chen G, et al. Kinesin family members KIF2C/4A/10/11/14/18B/20A/23 predict poor prognosis and promote cell proliferation in hepatocellular carcinoma. *Am J Transl Res*. 2020;12(5):1614-39.
33. Wang X, Sun X, Du X, Zhou F, Yang F, Xing J, et al. Thymidylate synthase gene polymorphisms as important contributors affecting hepatocellular carcinoma prognosis. *Clinics and Research in Hepatology and Gastroenterology*. 2017;41(3):319-26.

34. Chen S, Pu J, Bai J, Yin Y, Wu K, Wang J, et al. EZH2 promotes hepatocellular carcinoma progression through modulating miR-22/galectin-9 axis. *J Exp Clin Cancer Res.* 2018;37(1):3.