Revelation of critical genes and pathways involved in hepatocellular carcinoma (HCC) and phylogenetic tree construction of CDK1 protein family An integrated multi-omics bioinformatics analysis

Jingyang Zhou
Nanchang University https://orcid.org/0000-0002-3259-7783
Ziyi Chen
Nanchang University
Weiwei Cao
Nanchang University
Weiqi Chen
Shandong University of Traditional Chinese Medicine
Wuyuan Zhou (15564100001@163.com)
https://orcid.org/0000-0001-8781-3888

Research article

Keywords: Bioinformatics analysis, CDK1, phylogenic tree, differentially expressed genes, hepatocellular carcinoma

Posted Date: June 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-31902/v1

Abstract

Background

Hepatocellular carcinoma (HCC) is one of the leading and fastest increasing causes of death in the world. It primarily happens in the people who have been exposed to high-risk factors for a long time, such as the hepatitis B virus, hepatitis C virus, etc. Some similar analysis on HCC were published years ago, and results need updating.

Methods

In this study. We analyzed the gene expression pattern obtained from the GEO database in healthy tissue and tumor tissue and identified differentially expressed genes (DEGs), then GO and KEGG pathway analyses were performed subsequently on DEGs. We also used the protein-protein interaction (PPI) network drawn from STRING to find hub genes and analyzed their effects on prognosis and expression level in different stages of HCC from GEPIA. Furthermore, we analyzed the CDK1 in different species and established the phylogenetic tree of the CDK1 family by the MEGA-X (version 5) according to the amino acid sequence of CDK1 in different species.

Results

Among all DEGs identified in this study, the CDK1 is the most statically significant hub gene in the perspective of prognosis and differential ex pression. The expression level of CDK1 varies in different stages of HCC. The expression level reaches the highest in stage two and three and reaches the lowest in Stage four. And the human CDK1 gene and ape-like CDK1 gene subtype 5 may be homologous genes. However, human CDK1 gene and the gene encoding human CDK1 protein A chain may be heterozygous genes.

Conclusion

CDK1 is a gene with great significance in the diagnosis and prognosis of HCC, more studies that investigate deeply about the CDK1 is needed in the future.

Background

Hepatocellular carcinoma(HCC) is one of the prominent and growing causes in the United States and worldwide, especially in people previously infected with hepatitis B virus, hepatitis C virus and hepatitis D virus, etc. (1, 2) There are some novel diagnostic and treatment targets have been identified and broadly used, such as histone deacetylase inhibitors (HDACi), chimeric antigen receptor T cells (CAR-T cells) and multi-targeted tyrosine kinases (TKIs) (3, 4). In this study, we aimed to discover new targets or ideas for

the diagnosis and prognosis for HCC based on the gene expression level data in healthy samples and tumor samples and provide a unique aspect for understanding the underlying mechanism of carcinogenesis and prognosis of HCC.

We firstly systemically searched the GEO database (5), and based on four different platforms, we selected four profile datasets from three studies for data extraction and analysis. They were GSE14520 (6–12), GSE101685, and GSE64041 (13). We also selected two different profile datasets for results validation. They are GSE60502 (14) and GSE36376 (15, 16). The DEGs that mentioned in all four datasets or any three datasets were considered as co-DEGs, which are shown in Table S1. The validation sets showed that the co-DEGs identified had high confidence and credibility to be authentic.

The co-DEGs identified were updated to STRING (17) for the construction of the PPI network. We used the MCODE and cytohubba (version 0.0.1) applets of Cytoscape software (version 3.7.2) (18) to modify the whole PPI network and divided the entire network into statically and medically significant modules for further analysis. And find hub genes respectively within the whole PPI network and the module 1.

We selected all co-DEGs, all genes involved in three modules of PPI network, and the top ten hub genes of the whole network for further GO (19) and KEGG (20) pathway analysis on DAVID (21). The results revealed that all co-DEGs were mainly enriched in the epoxygenase P450 pathway, organelle membrane, iron ion binding, and metabolic pathways. For the upregulated co-DEGs, they were primarily enriched in epoxygenase P450 pathway, organelle membrane, oxidoreductase activity, and act on paired donors with incorporation or reduction of molecular oxygen and metabolic pathways. In contrast, the downregulated co-DEGs, the functional enrichment analysis results were cell division, spindle, protein binding, and cell cycle. For module 1, the genes involved in this were mainly enriched in cell division, nucleus, protein binding, and cell cycle. For module 2, the genes involved in this were primarily enriched in epoxygenase P450 pathway, organelle membrane, arachidonic acid epoxygenase activity, and drug metabolism-cytochrome P450. For module 3, the genes involved in this were primarily enriched in xenobiotic metabolic process, organelle membrane, oxygen binding, and metabolic pathways. For the top 10 hub genes of the whole network, they were primarily enriched in cell division, nucleoplasm, cyclin-dependent protein serine/threonine kinase activity, and p53 signaling pathway.

After the functional enrichment analysis, we focused on the clinical significance, which is the effect on the prognosis of hub genes. We used GEPIA (22, 23) to draw the percentage of overall survival of patients with a specific hub gene differentially expressed, and set the Log-rank P-value as a criteria for determining the static correlation between the expression of this hub gene and the percentage of overall survival of patients. We found that among the ten hub genes we identified, CCNB2 is the only gene with a Log-rank P-value more than 0.05, which indicates that this gene has weak or no correlation with the prognosis of HCC. CDK1 and CCNB1 have the smallest Log-rank P-value, which were 0.00017 and 0.00015, respectively, indicating that both of these two genes have a strong correlation with the prognosis of HCC.

We validated our results with another two profile datasets searched form the GEO database, and we firstly used R (version 3.6.0) (24) to generate 25 random numbers for choosing the genes for validation.

To determine which one has more considerable medical significance for further investigation, we collected the expression data of each gene from the matrix of the four profile datasets used for data extraction and analysis. We plotted the forest plot (25, 26) (Fig. 1) for the expression level of each gene with a fixed-effect model. Based on the heterogenicity and the overall mean difference, we chose CDK1 for further analysis.

We firstly systemically searched the sequence of CDK1 in humans from the uniport (27–29) and conducted a systemic position-specific iterated (PSI) search for three times on BLAST (30–32). Then we excluded the redundant sequence and exported the final results of protein sequence and corresponding species into MEGA-X (version 5) (33, 34) for the construction of the phylogenetic tree of the CDK1 family.

In the end, we searched and analyzed the CDK1 expression pattern in different stages of patients with HCC (35, 36), which may offer a new understanding of the role of CDK1 in the carcinogenesis and proliferation of HCC.

Methods

Microarray platform

The four profile datasets for data extraction and analysis were based on different platforms. GSE101685 was based on GPL570 ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array), GSE64041 was based on the GPL6244 ([HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]) and GSE14520 contains two profile datasets, one is based on the GPL571 ([HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array), the other one was based on the GPL3921 ([HT_HG-U133A] Affymetrix HT Human Genome U133A Array).

Data assignment

To increase the credibility of our findings, we assigned the six datasets into two groups, one is a data extraction and analysis group, which is used for the identification of DEGs, and following relevant researches, the other one is used for validation, which is used for validating the results. GSE14520, GSE101685, and GSE 64041 were assigned into the data extraction and analysis group; GSE60502 and GSE36376 were assigned into the validation group.

Expression analysis of DEGs.

We used the GEO2R function on the website of the GEO database to sort the data into two groups, one is consist of the liver tissue extracted from healthy people; the other one contains the tissue obtained from patients diagnosed as the HCC, and individually analyzed each profile dataset with all the relevant sets

default and exported to Excel. For all four profile datasets, a gene with |log2-fold change|>1.5 and adj. P < 0.05 were considered as a DEG.

GO and KEGG pathway enrichment analysis.

The Database for Annotation Visualization and Integrated Discovery (DAVID) is a website that provides a wide range of functional annotation tools to investigate the effect of genes in the term of biological process, molecular function, cell components, etc. Identified DEGs were investigated further using DAVID (version 6.8), GO, and KEGG pathway enrichment analyses. While analyzing all genes in the whole network, all upregulated co-DEGs, all downregulated co-DEGs and genes involved in the three modules identified from the PPI network, P < 0.05 and gene counts of > 10 indicates a statistically significant difference in the functional enrichment analysis; For the top 10 hub genes, P < 0.05 and gene counts of > 5 indicates a statistically significant difference in the functional enrichment analysis.

Integration of the PPI network.

The protein-protein interaction network (PPI) of all proteins expressed by the DEGs was mapped by the STRING (https://string-db.org/; version 11.0) to evaluate the interactions between the DEGs. The Cytoscape plugin Molecular Complex Detection (MCODE; version 1.6) was used to detect the medically and statistically significant regions within the whole PPI network, which is called specific modules. Interactions with medium confidence (a combined score > 0.400) were defined as statistically significant. Cytoscape software (version 3.7.2) was used to visualize and adjust PPI networks. And based on the degree levels in the Cytoscape plugin cytoHubba (version 0.1), the top 10 ranked genes were defined as hub genes.

Analysis of hub genes on the GEPIA

Identified ten hub genes performed survival analysis and analysis about expression level in different stages of HCC. And in the overall survival plot, if Log-rank p < 0.05, that means the difference in the gene expression level is strictly relevant to the prognostic survival of HCC. The dotted line in the overall survival plot indicated the 95% confidence interval (95% CI). The violin plot exhibited the expression level of all hub genes in different stages of HCC.

Validation of identified DEGs

For validation, we chose another two independent profile datasets for validation, which are also called GSE60502 and GSE36376. We firstly numbered all co-DEGs in order and used R (x64 3.6.0) to generate 25 random numbers, which were used for choosing the DEGs listed in table S1 for validation. HAAO, RACGAP1, ETFDH, KCNN2, PLGLB1, SULT1C2, CYP3A43, IGFALS, KIAA0101, TOP2, CAP2, N4BP2L1, FCN2, F11, MAT1A, MT1, CYP2A6, PLG, CDC37L1, SOCS2, CENPU, LAPTM4B, SAC3D1, AZGP1P1, CYP2E1 were chosen according to the random numbers. The first 13 genes were validated according to the data from GSE60502, and the rest of genes were validated according to the data from GSE36376.

Construction of phylogenetic tree of CDK1 family

We searched the sequence of human CDK1 protein in FASTA format (37, 38) and performed a systemic position-specific iterated (PSI) search for three times on BLAST(43, 44). We imported the sequences after the de-redundancy process into MEGA-X for the construction of maximum-likelihood phylogenetic tree (39, 40) based on the JJT matrix model and 100 bootstrap replications, all other settings remained the default. This phylogenetic tree indicates the relationship and similarity of different species from the perspective of the CDK1 protein family.

Results

The identification of DEGs

In this study, 1093 DEGs were identified, of which 721 were upregulated, and 372 were downregulated, including 196 co-expressed upregulated genes and 66 co-expressed downregulated genes, (Table S1) which are roughly shown in the Venn diagram (Fig. 2a,b). Volcano plots (41, 42) (Fig. 3, a-d) and expression heatmap (Fig. 4, a-d) for all profile datasets used for data extraction and analysis were constructed. The red lines in the volcano plots represent the threshold values of the identification of DEGs.

Functional Enrichment analysis

To identify the pathways which had the most biologically and statistically significant involvement with the genes identified, upregulated and downregulated DEGs were submitted into DAVID for GO and KEGG pathway analysis. For all co-DEGs, they are mainly concentrated in cyclooxygenase P450 pathway, cell membrane, iron-binding, and metabolic pathway (Fig. 5, a-d). The upregulated co-DEGs are enriched primarily on cyclooxygenase P450 pathway, cell membrane, and oxidoreductase activity, which binds or reduces molecular oxygen and metabolic pathways by acting on paired donors (Fig. 6, a-d). As for the downregulated co-DEGs, the functional enrichment analysis reveals the results that cell division, spindle, protein binding, and cell cycle. (Fig. 7, a-d) For module 1, the genes involved in this were mainly enriched in cell division, nucleus, protein binding, and cell cycle. (Fig. 8, a-d) For module 2, the genes involved in this were primarily enriched in epoxygenase P450 pathway, organelle membrane, arachidonic acid epoxygenase activity, and drug metabolism-cytochrome P450 (Fig. 9, a-d). For module 3, the genes involved in this were primarily enriched in xenobiotic metabolic process, organelle membrane, oxygen binding, and metabolic pathways (Fig. 10, a-d). For the top 10 hub genes of the whole network, they were primarily enriched in cell division, nucleoplasm, cyclin-dependent protein serine/threonine kinase activity, and p53 signaling pathway (Fig. 11, a-d).

PPI network construction and module analysis

Interactions between the identified DEGs were elucidated by the PPI network. (Fig. 12, a-c) In total, there were 226 nodes in the network, the average number of neighbors is 13.416, the clustering coefficient is

0.457, and the network heterogenicity is 1.044. The top ten hub genes were TPX2, TTK, RRM2, TOP2A, NCAPG, CCNB2, ZWINT, CCNB1, CDK1, and RFC4. And after analyzing with The Cytoscape plugin Molecular Complex Detection (MCODE; version 1.6) of Cytoscape (3.7.2), we got three statically and medically significant modules. We analyzed all three modules and selected the top five hub genes in module 1. They are TOP2A, CCNB2, TTK, PRC1, and NUSAP1 (Fig. 13a,b).

Prognostic survival analysis

We analyzed the effect of the top ten hub genes on the prognosis of HCC patients (Fig. 14, a-j). CCNB2 is the only gene with a Log-rank P value more the 0.05, which indicates that CCNB2 has almost no correlation with overall survival outcome of HCC. CDK1 and CCNB1 have the smallest Log-rank P-value among these (0.00017 and 0.00015, respectively), which indicates that both of these two genes have a strong correlation with the prognosis of HCC.

Expression level of CDK1 in different stages of HCC

The expression of CDK1 was different in different stages of HCC, with the highest expression level in stage 3 and the lowest in stage 4. The discrete degree of expression level reached the highest in the second stage. (Fig. 15)

Phylogenic tree of CDK1 family analysis

The phylogenic of the CDK1 protein family was based on the maximum-likelihood among species. From this phylogenic tree, we can easily conclude that human CDK1 gene and ape-like CDK1 gene subtype 5 may be homologous genes. However, human CDK1 gene and the gene encoding human CDK1 protein A chain may be heterozygous genes (Fig. 16).

Discussion

From the result of functional enrichment analysis, we can find that the epoxygenase P450 pathway is a pathway that is mentioned many times with a high count of genes and correlation in functional enrichment analysis of different genes. Arthur A Spector et al. reported in 2014 (45) that epoxygenase P450 pathway involved a lot in the angiogenesis and vasodilation, which are essential factors for the HCC proliferation, in other words, the proliferation of HCC may be initiated and maintained by this pathway.

In terms of molecular function, protein binding and oxygen-binding are the most prominent ones among all results. HCC is always vascularized by the upregulation of angiogenic factors, such as VEGF (46). And VEGF is a primary target of HIF-1 (47), which means HIF-1 has an angiogenic effect on HCC. In 2017, Chen et, al reported that the HIF-1 also has a positive impact on the proliferation of HCC, such as promotion of glycolysis, enhancement of critical oncogenic factors, etc. (48)

In terms of the organelle, the organelle membrane was mentioned many times with a high count of genes and correlation in functional enrichment analysis of different genes. The genes involved in the organelle membrane are all existed on the endoplasmic membrane. The relationship between the endoplasmic membrane and carcinogenesis of HCC may be involved in the ER stress (49), and more studies with high credibility are in need to validate this relationship.

In the hub gene analysis, TOP2A protein is one type of the topoisomerase in our bodies that controls the topological pattern of DNA (50), while the deletion or down-regulation of TOP2A may result in the resistance of topoi inhibitor-chemotherapy (51).

In terms of enhancing the proliferation of HCC, NCAPG shares similar mechanisms with TTK. NCAPG, which is also called non-SMC condensin I complex subunit G, and the overexpression of NCAPG leads to an enhanced proliferation and reduced apoptosis of HCC through PI3K/AKT signaling pathway (52). Liu et al. reported in 2015 that TTK activates the Akt and promotes the proliferation and migration of HCC (53).

Ying et al. .in 2018 (54) reported that ZWINT promotes the proliferation of HCC through the regulation of cell-cycle-related proteins such as CDK1, PCNA, etc. The regulation is also mentioned in the top ten hub gene network, which is corresponding to our findings.

However, there are some inevitable limitations in this study. First, the number of available profile datasets in the database was not sufficient to analyze more DEGs, which may cause the imprecision of the results. So, more microarray data that compare the gene expression level between the tumor sample and the normal sample of HCC is urgently needed. Secondly, The microarray data available now is mostly about the white people, and only a few datasets cover the yellow or black people, which means, on the one hand, the expression level difference caused by the difference of race is unable to be identified; on the other hand, the heterogenicity between studies was dramatically huge, which indirectly decreases the credibility of studies included and results.

Due to the limitation of the condition and the outbreak of the COVID-19 virus, we are unable to collect enough biopsy samples and conduct qRT-PCR for validation with our data. We have to use the data from others for validation. And the expression heatmap of the genes chosen for validation in each dataset in also generated. (Fig. 17) The reasons why we chose the chose GSE60502 and GSE36376 for validation are as follows. Firstly, these two datasets contain 469 samples in total, which will significantly increase the credibility of the results of validation. Secondly, these two studies were performed in Taiwan and Korea, thus the influence of unknown factors such as racial, country, economic development can be maximumly deduced. Thirdly, this dataset was collected in recent years. Last but not least, all genes involved in the cross-validation were randomly chosen according to the random number generated by software—all these reasons listed above contributed to the high credibility of the following cross-validation. Therefore, we can guarantee that this cross-validation will have the same reliability as a real experiment, and we have enough confidence in the results of validation.

Conclusion

CDK1 is a gene with great significance in the diagnosis and prognosis of HCC, more studies that investigate deeply about the CDK1 is needed in the future.

Abbreviations

PVALUE: -log10(p-value); HCC: hepatocellular carcinoma; DEGs: differentially expressed genes; PPI network: protein-protein interaction network; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Gene and Genomes

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by The Government of the Xuzhou City. (2018TD011).

Authors' contributions

JZ and WZ designed the study, JZ collected the data, carried out the data analysis, JZ produced the initial draft of the manuscript, WCa contributed to drafting the manuscript, ZC and WCh polished all the figures in the manuscript.

Acknowledgments

We would like to acknowledge the helpful comments on this paper received from our reviewers.

References

1. Sagnelli E, Macera M, Russo A, Coppola N, Sagnelli C. Epidemiological and etiological variations in hepatocellular carcinoma. Infection. 2020;48(1):7–17. doi:10.1007/s15010-019-01345-y.
2. Kanwal F, Singal AG. Surveillance for Hepatocellular Carcinoma: Current Best Practice and Future Direction. Gastroenterology. 2019;157(1):54–64. doi:10.1053/j.gastro.2019.02.049.
3. Neureiter D, Stintzing S, Kiesslich T, Ocker M. Hepatocellular carcinoma: Therapeutic advances in signaling, epigenetic and immune targets. World J Gastroenterol. 2019;25(25):3136–50. doi:10.3748/wjg.v25.i25.3136.
4. Mody K, Abou-Alfa GK. Systemic Therapy for Advanced Hepatocellular Carcinoma in an Evolving Landscape. Curr Treat Options Oncol. 2019;20(2):3. doi:10.1007/s11864-019-0601-1. Published 2019 Jan 11.
5. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41(Database issue):D991–5. doi:10.1093/nar/gks1193.
6. Roessler S, Jia HL, Budhu A, Forgues M, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res 2010 Dec 15;70(24):10202–12. PMID: 21159642.
7. Roessler S, Long EL, Budhu A, Chen Y, et al. Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. Gastroenterology 2012 Apr;142(4):957–966.e12. PMID: 22202459.
8. Zhao X, Parpart S, Takai A, Roessler S, et al. Integrative genomics identifies YY1AP1 as an oncogenic driver in EpCAM(+) AFP(+) hepatocellular carcinoma. Oncogene 2015 Sep 24;34(39):5095–104. PMID: 25597408.
9. Wang Y, Gao B, Tan PY, Handoko YA, et al. Genome-wide CRISPR knockout screens identify NCAPG as an essential oncogene for hepatocellular carcinoma tumor growth. FASEB J 2019 Aug;33(8):8759–8770. PMID: 31022357.
10. Sun Y, Ji F, Kumar MR, Zheng X, et al. Transcriptome integration analysis in hepatocellular carcinoma reveals discordant intronic miRNA-host gene pairs in expression. Int J Biol Sci. 2017;13(11):1438–49. PMID: 29209147.
11. Lu Y, Xu W, Ji J, Feng D, et al. Alternative splicing of the cell fate determinant Numb in hepatocellular carcinoma. Hepatology 2015 Oct;62(4):1122–31. PMID: 26058814.
12. Chen S, Fang H, Li J, Shi J, et al. Microarray Analysis For Expression Profiles of lncRNAs and circRNAs in Rat Liver after Brain-Dead Donor Liver Transplantation. Biomed Res Int 2019;2019:5604843. PMID: 31828106.
13. Makowska Z, Boldanova T, Adametz D, Quagliata L, et al. Gene expression analysis of biopsy samples reveals critical limitations of transcriptome-based molecular classifications of hepatocellular carcinoma. J Pathol Clin Res. 2016 Apr;2(2):80–92. PMID: 27499918.
14. Wang YH, Cheng TY, Chen TY, Chang KM, et al. Plasmalemmal Vesicle Associated Protein (PLVAP) as a therapeutic target for treatment of hepatocellular carcinoma. BMC Cancer 2014 Nov 6;14:815. PMID: 25376302.

15. Lim HY, Sohn I, Deng S, Lee J, et al. Prediction of disease-free survival in hepatocellular carcinoma by gene expression profiling. Ann Surg Oncol. 2013 Nov;20(12):3747–53. PMID: 23800896.
16. Ha SY, Sohn I, Hwang SH, Yang JW, et al. The prognosis of hepatocellular carcinoma after curative hepatectomy in young patients. Oncotarget 2015 Jul 30;6(21):18664–73. PMID: 26093092.
17. SzklarczykD, Franceschini A, Kuhn M, Simonovic M, Roth A, MinguezP, DoerksT, Stark M, Muller J, Bork P, et al. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 2011;39(Database Issue):D561–8.
18. Smoot ME, Ono K, RuscheinskiJ, Wang PL. and IdekerT: Cytoscape2.8: New features for data integration and network visualization. Bioinformatics. 2011;27:431–2.
19. Gene Ontology Consortium. The Gene Ontology (GO) project in 2006. Nucleic Acids Res. 2006;34(Database Issue):D322–6.
20. Ogata H, GotoS, Sato K, FujibuchiW, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 1999;27:29–34.
21. Jiao X, Sherman BT, Huang da W, Stephens R, BaselerMW, Lane HC, Lempicki RA. DAVID–WS: A stateful web service to facilitate gene/protein list analysis. Bioinformatics. 2012;28:1805–6.
22. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45(W1):W98–102. doi:10.1093/nar/gkx247.
23. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res. 2019;47(W1):W556–60. doi:10.1093/nar/gkz430.
24. Shim SR, Kim SJ. Intervention meta-analysis: application and practice using R software. Epidemiol Health. 2019;41:e2019008. doi:10.4178/epih.e2019008.
25. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health. 2019;22(4):153–60. doi:10.1136/ebmental-2019-300117.
26. Hu D, O'Connor AM, Winder CB, Sargeant JM, Wang C. How to read and interpret the results of a Bayesian network meta-analysis: a short tutorial. Anim Health Res Rev. 2019;20(2):106–15. doi:10.1017/S1466252319000343.
27. The UniProt Consortium. UniProt: the universal protein knowledgebase [published correction appears in Nucleic Acids Res. 2018 Mar 16;46(5):2699]. Nucleic Acids Res. 2017;45(D1):D158-D169. doi:10.1093/nar/gkw1099.
28. UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 2015;43(Database issue):D204-D212. doi:10.1093/nar/gku989.
29. UniProt Consortium
10.1002/0471250953.bi0127s50
Pundir S, Magrane M, Martin MJ, O'Donovan C. UniProt Consortium. Searching and Navigating UniProt Databases. CurrProtoc Bioinformatics. 2015;50:1.27.1-1.27.10. Published 2015 Jun 19. doi:10.1002/0471250953.bi0127s50.

30. Cock PJ, Chilton JM, Grüning B, Johnson JE, Soranzo N. NCBI BLAST + integrated into Galaxy. Gigascience. 2015;4:39. doi:10.1186/s13742-015-0080-7. Published 2015 Aug 25.
31. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389–402. doi:10.1093/nar/25.17.3389.
32. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2018;46(D1):D8–13. doi:10.1093/nar/gkx1095.
33. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35(6):1547–9. doi:10.1093/molbev/msy096.
34. Hall BG. Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol. 2013;30(5):1229–35. doi:10.1093/molbev/mst012.
35. Chen Y, Kao SL, Tan M, et al. Feasibility of representing adherence to blood glucose monitoring through visualizations: A pilot survey study among healthcare workers. Int J Med Inform. 2018;120:172–8. doi:10.1016/j.ijmedinf.2018.09.006.
36. Chen QY, Zheng CH, Li P, et al. Which method is more suitable for advanced gastric cancer with enlarged lymph nodes, laparoscopic radical gastrectomy or open gastrectomy? Gastric Cancer. 2018;21(5):853–63. doi:10.1007/s10120-018-0800-7.
37. Binz PA, Shofstahl J, Vizcaíno JA, et al. Proteomics Standards Initiative Extended FASTA Format. J Proteome Res. 2019;18(6):2686–92. doi:10.1021/acs.jproteome.9b00064.
38. 10.1002/0471250953.bi0309s53
Pearson WR. Finding Protein and Nucleotide Similarities with FASTA. CurrProtoc Bioinformatics. 2016;53:3.9.1-3.9.25. Published 2016 Mar 24. doi:10.1002/0471250953.bi0309s53.
39. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 1992;8:275–82.
40. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.
41. Mitteer DR, Greer BD, Fisher WW, Cohrs VL. Teaching behavior technicians to create publication-quality, single-case design graphs in graphpad prism 7. J Appl Behav Anal. 2018;51(4):998–1010. doi:10.1002/jaba.483.
42. Berkman SJ, Roscoe EM, Bourret JC. Comparing self-directed methods for training staff to create graphs using Graphpad Prism. J Appl Behav Anal. 2019;52(1):188–204. doi:10.1002/jaba.522.
43. Stephen F, Altschul TL, Madden, Alejandro A, Schäffer J, Zhang Z, Zhang W, Miller, Lipman DJ. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". Nucleic Acids Res. 1997;25:3389–402.
44. Alejandro A, Schäffer L, Aravind TL, Madden S, Shavirin JL, Spouge, Yuri I, Wolf EV, Koonin, Stephen F, Altschul. "Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements". Nucleic Acids Res. 2001;29:2994–3005.

45. Spector AA, Kim HY. Cytochrome P450 epoxygenase pathway of polyunsaturated fatty acid metabolism. BiochimBiophys Acta. 2015;1851(4):356–65. doi:10.1016/j.bbalip.2014.07.020.
46. 10.1038/nrclinonc.2011.30
Zhu AX, Duda DG, Sahani DV, Jain RK. HCC and angiogenesis: possible targets and future directions [published correction appears in Nat Rev Clin Oncol. 2011 May;8(5):302]. Nat Rev Clin Oncol. 2011;8(5):292-301. doi:10.1038/nrclinonc.2011.30.
47. Zhang D, Lv FL, Wang GH. Effects of HIF-1α on diabetic retinopathy angiogenesis and VEGF expression. Eur Rev Med Pharmacol Sci. 2018;22(16):5071–6. doi:10.26355/eurrev_201808_15699.
48. Chen C, Lou T. Hypoxia inducible factors in hepatocellular carcinoma. Oncotarget. 2017;8(28):46691–703. doi:10.18632/oncotarget.17358.
49. Choi YM, Lee SY, Kim BJ. Naturally Occurring Hepatitis B Virus Mutations Leading to Endoplasmic Reticulum Stress and Their Contribution to the Progression of Hepatocellular Carcinoma. Int J Mol Sci. 2019;20(3):597. doi:10.3390/ijms20030597. Published 2019 Jan 30.
50. 10.1128/ecosalplus.ESP-0010-
Bush NG, Evans-Roberts K, Maxwell ADNA, Topoisomerases. EcoSal Plus. 2015;6(2):10.1128/ecosalplus.ESP-0010-2014. doi:10.1128/ecosalplus.ESP-0010-2014.
51. Järvinen TA, Liu ET. Topoisomerase IIalpha gene (TOP2A) amplification and deletion in cancer–more common than anticipated. Cytopathology. 2003;14(6):309–13. doi:10.1046/j.0956-5507.2003.00105.x.
52. Gong C, Ai J, Fan Y, et al. NCAPG Promotes The Proliferation Of Hepatocellular Carcinoma Through PI3K/AKT Signaling. Onco Targets Ther. 2019;12:8537–52. doi:10.2147/OTT.S217916. Published 2019 Oct 16.
53. Liu X, Liao W, Yuan Q, Ou Y, Huang J. TTK activates Akt and promotes proliferation and migration of hepatocellular carcinoma cells. Oncotarget. 2015;6(33):34309–20. doi:10.18632/oncotarget.5295.
54. Ying H, Xu Z, Chen M, Zhou S, Liang X, Cai X. Overexpression of Zwint predicts poor prognosis and promotes the proliferation of hepatocellular carcinoma by regulating cell-cycle-related proteins. Onco Targets Ther. 2018;11:689–702. doi:10.2147/OTT.S152138. Published 2018 Feb 2.

Figures

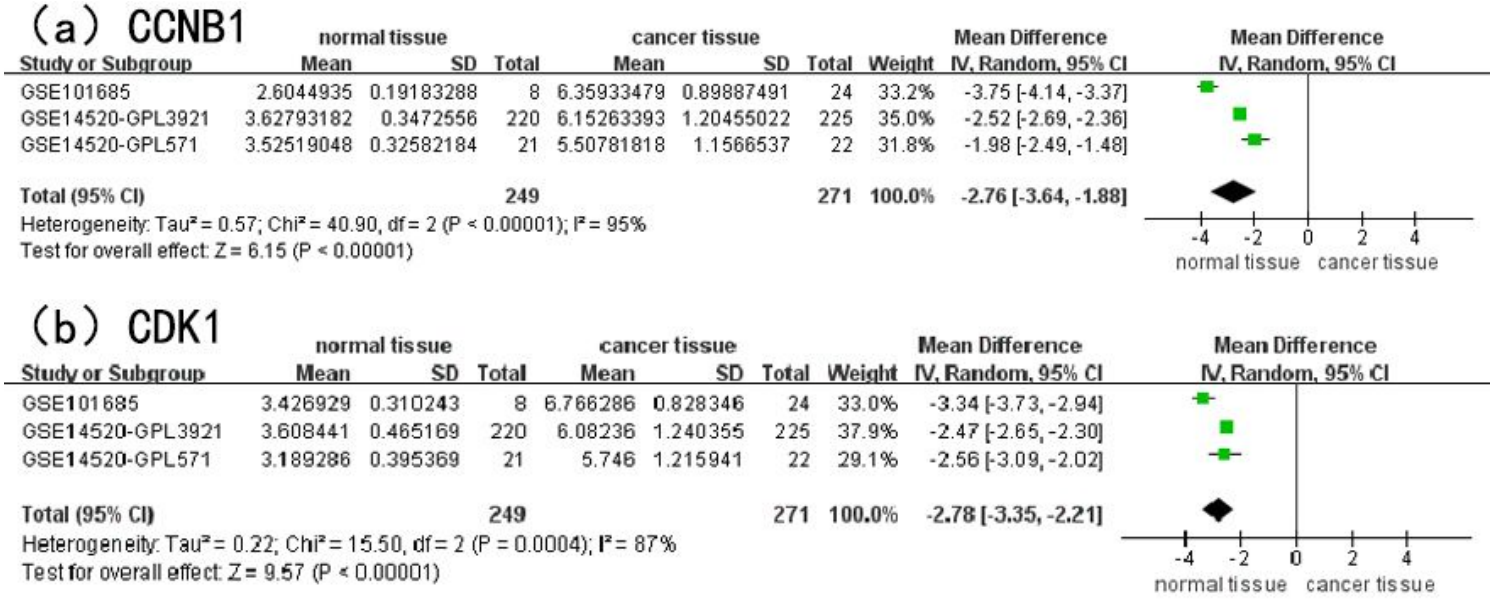

Figure 1

The forest plot of CDK1 and CCNB1 expression level

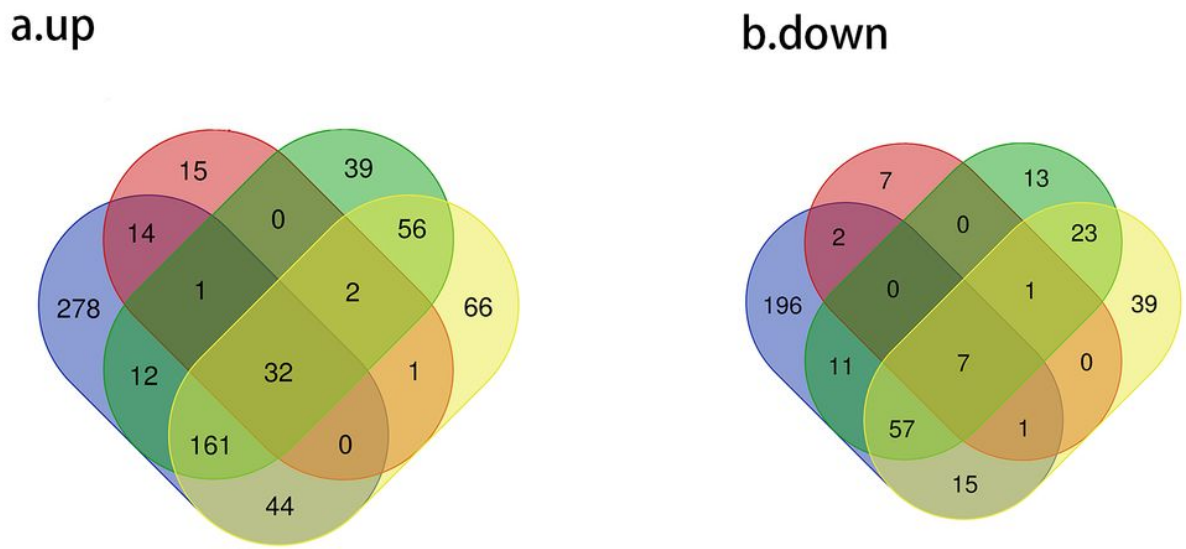

Figure 2

a) The Venn diagram for upregulated DEGs. b) The Venn diagram of the downregulated DEGs

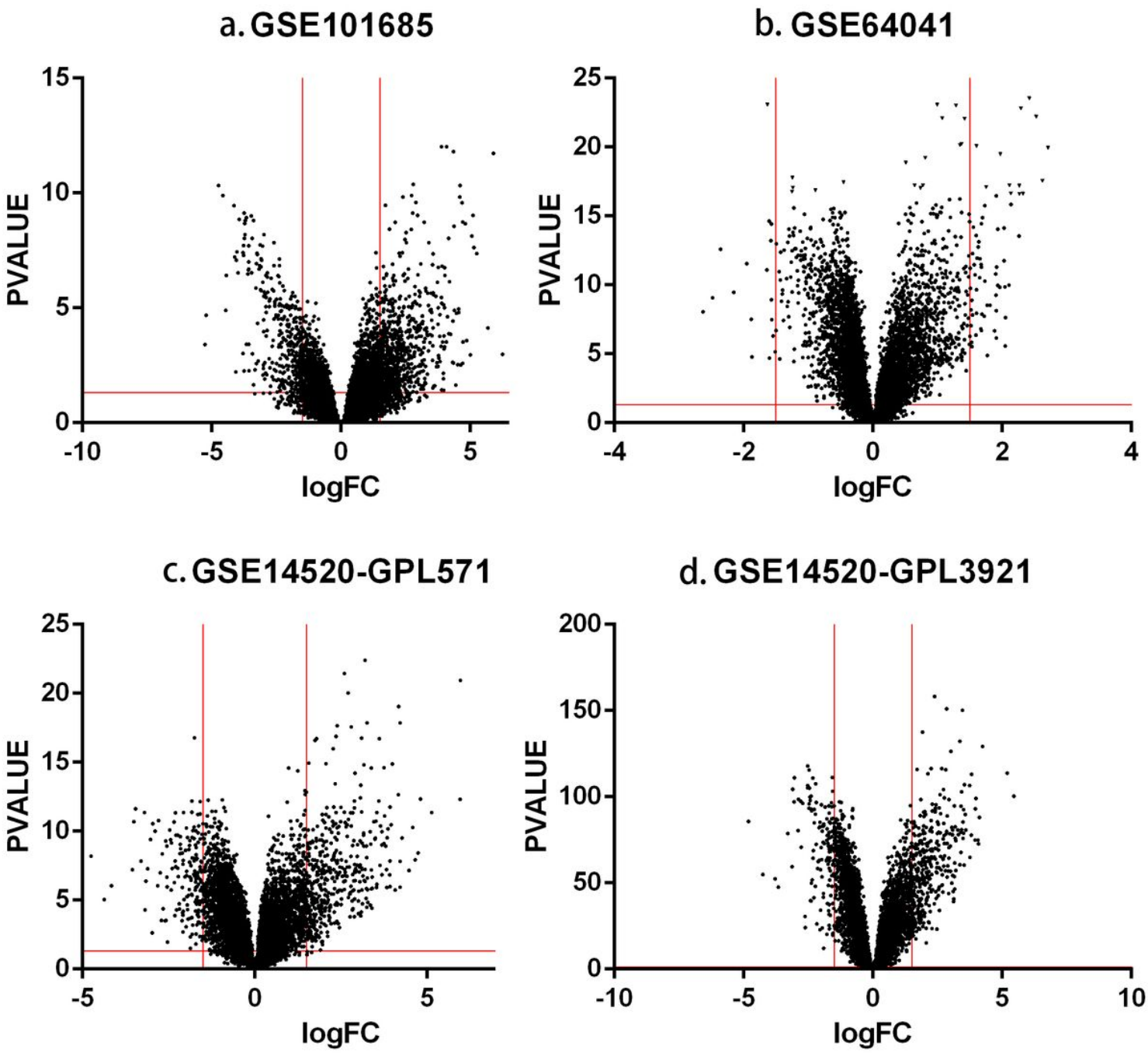

Figure 3

a) The volcano plot for GSE101685 b) The volcano plot for GSE64041 c) The volcano plot for GSE14520-GPL571 d) The volcano plot for GSE14520-GPL3921

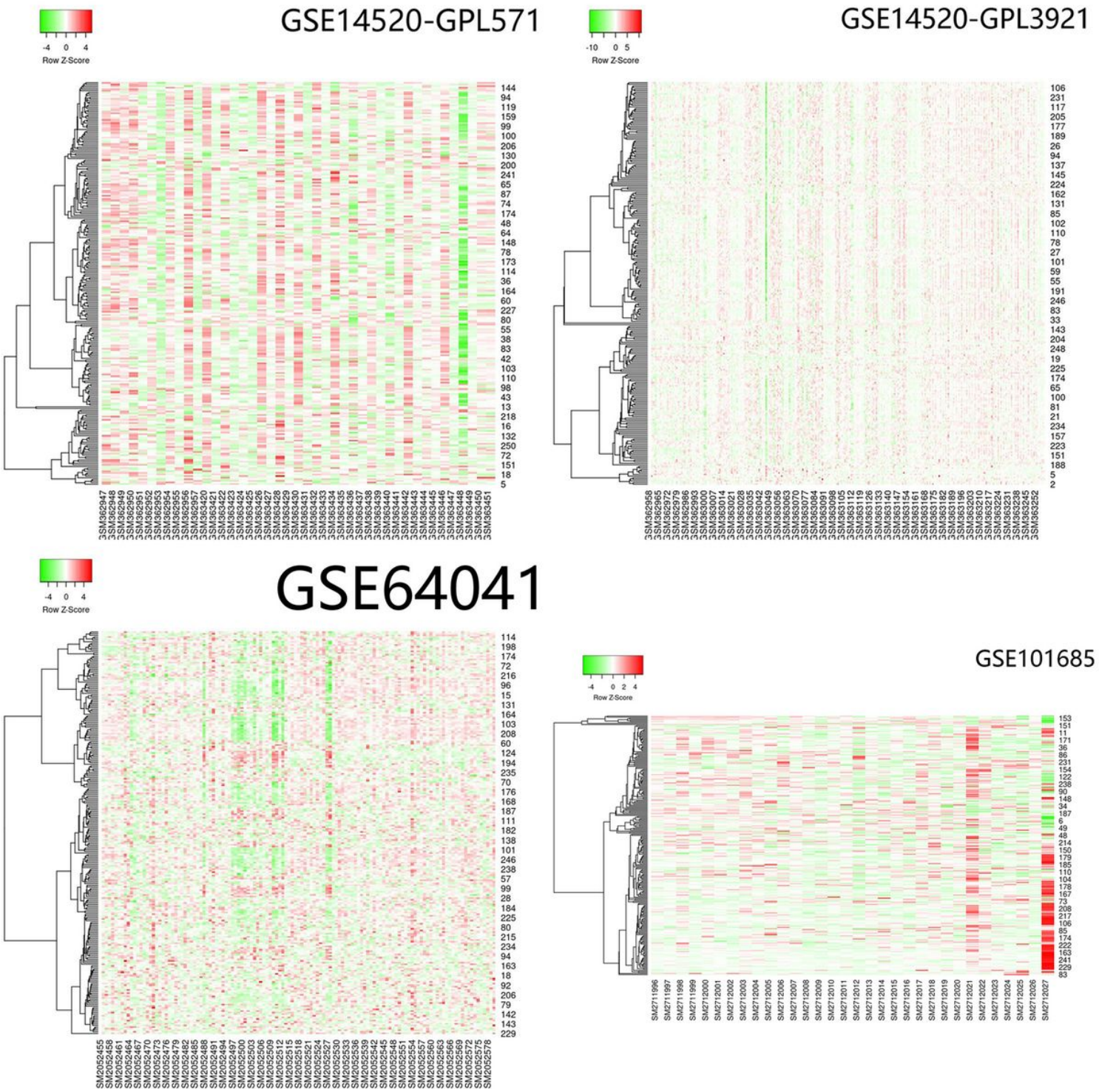

Figure 4

The expression heatmap for each dataset

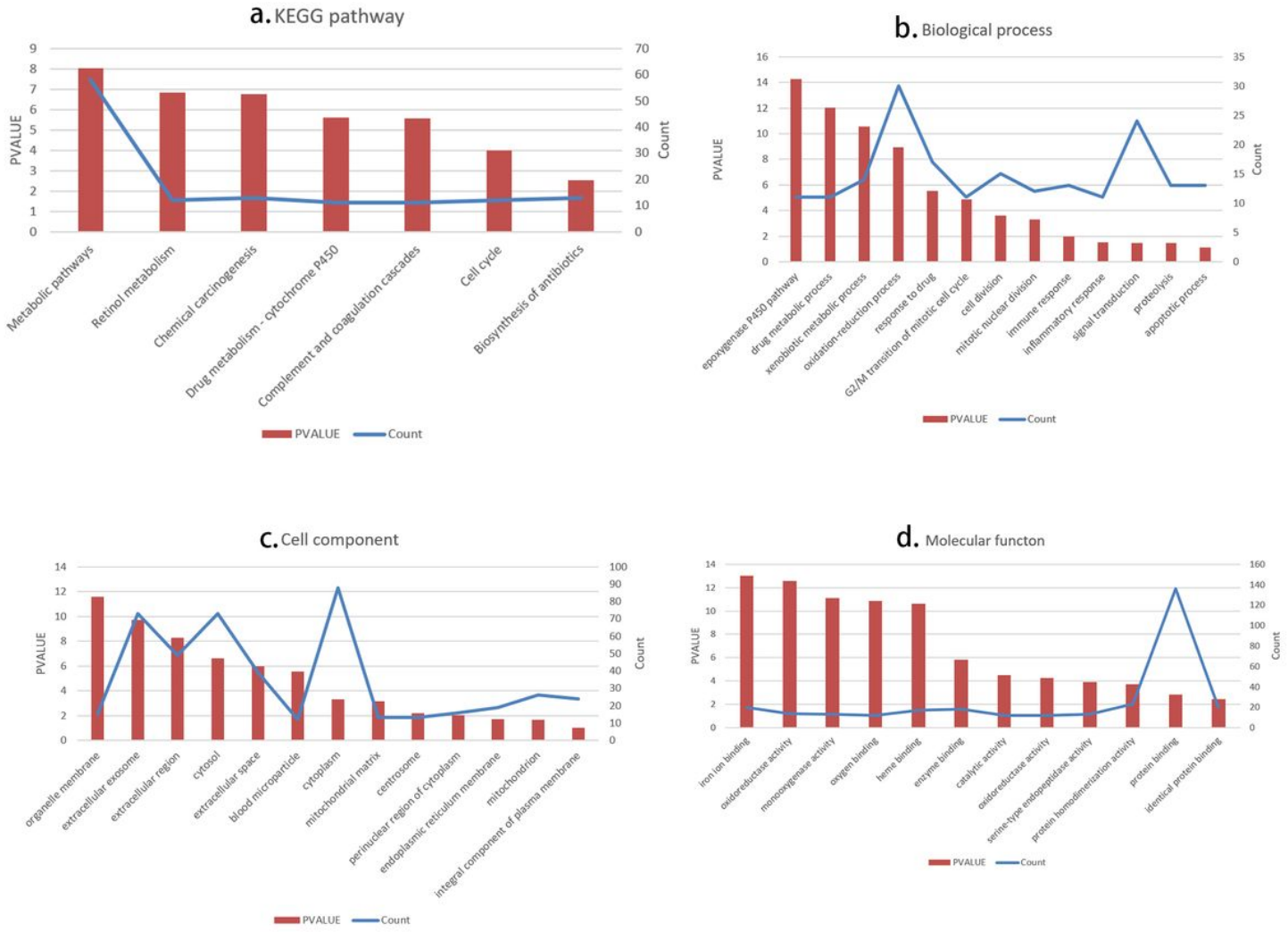

Figure 5

The result of the functional enrichment analysis of all co-DEGs

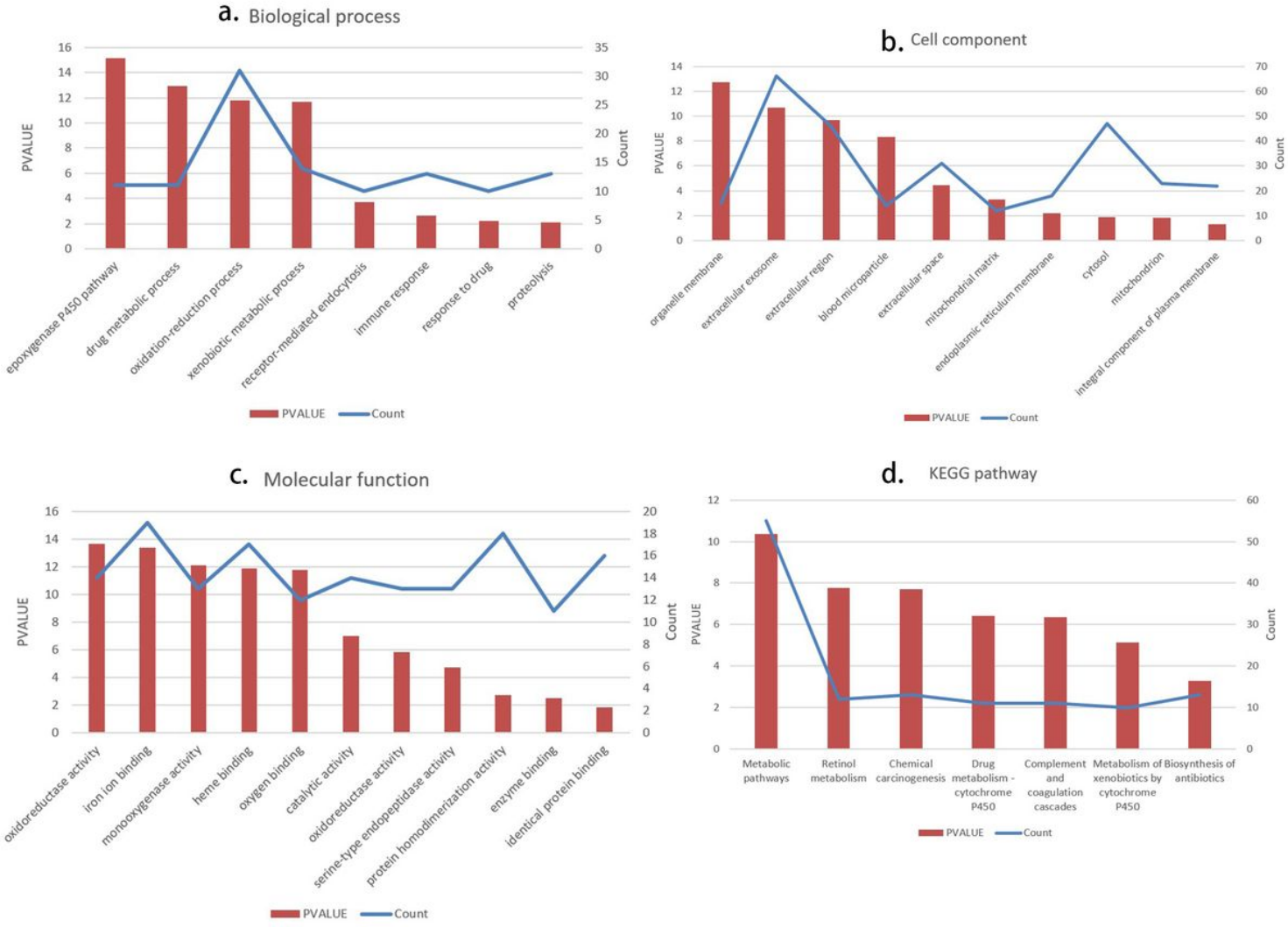

Figure 6

The result of the functional enrichment analysis of upregulated DEGs

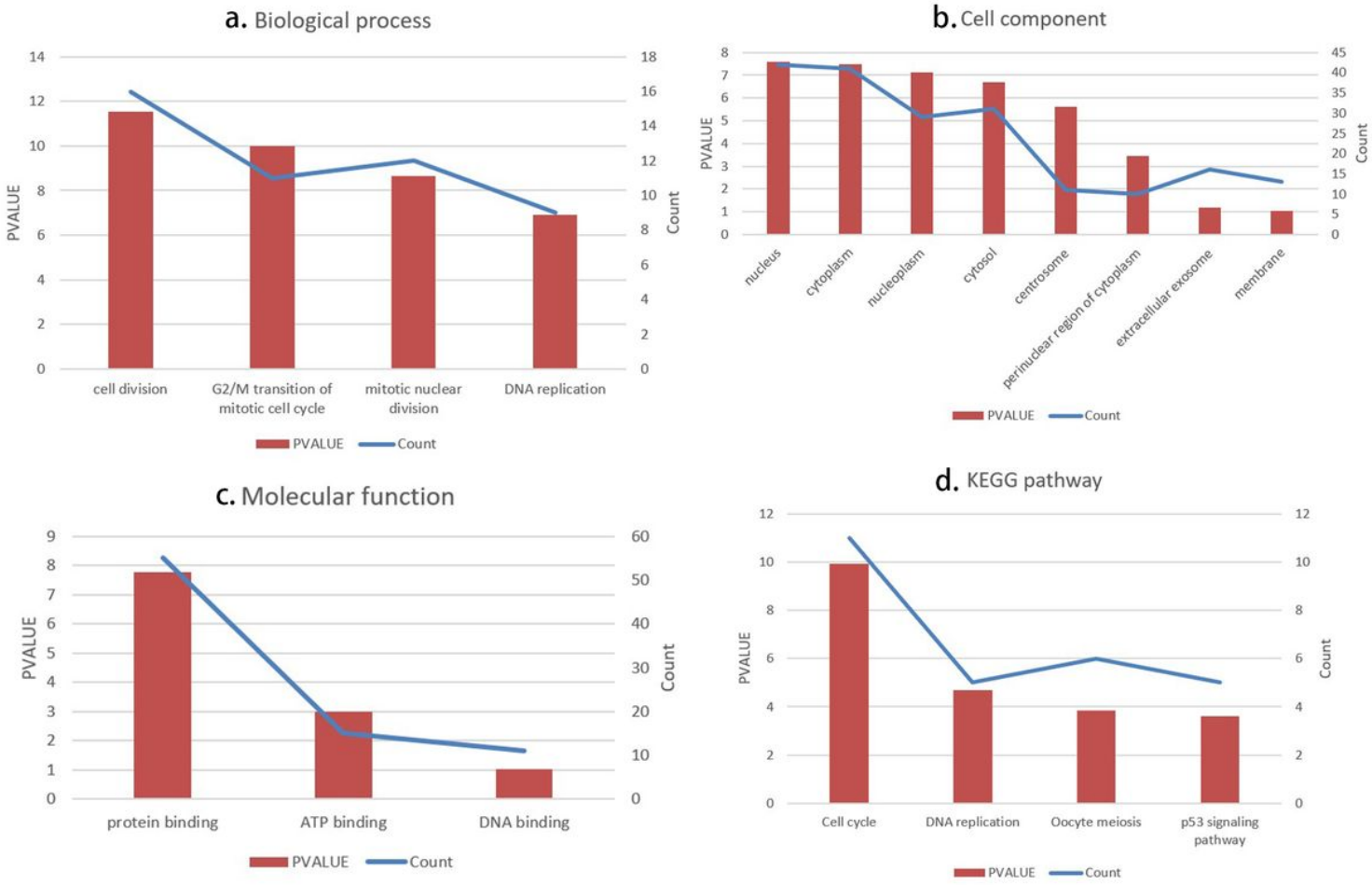

Figure 7

The result of the functional enrichment analysis of downregulated DEGs

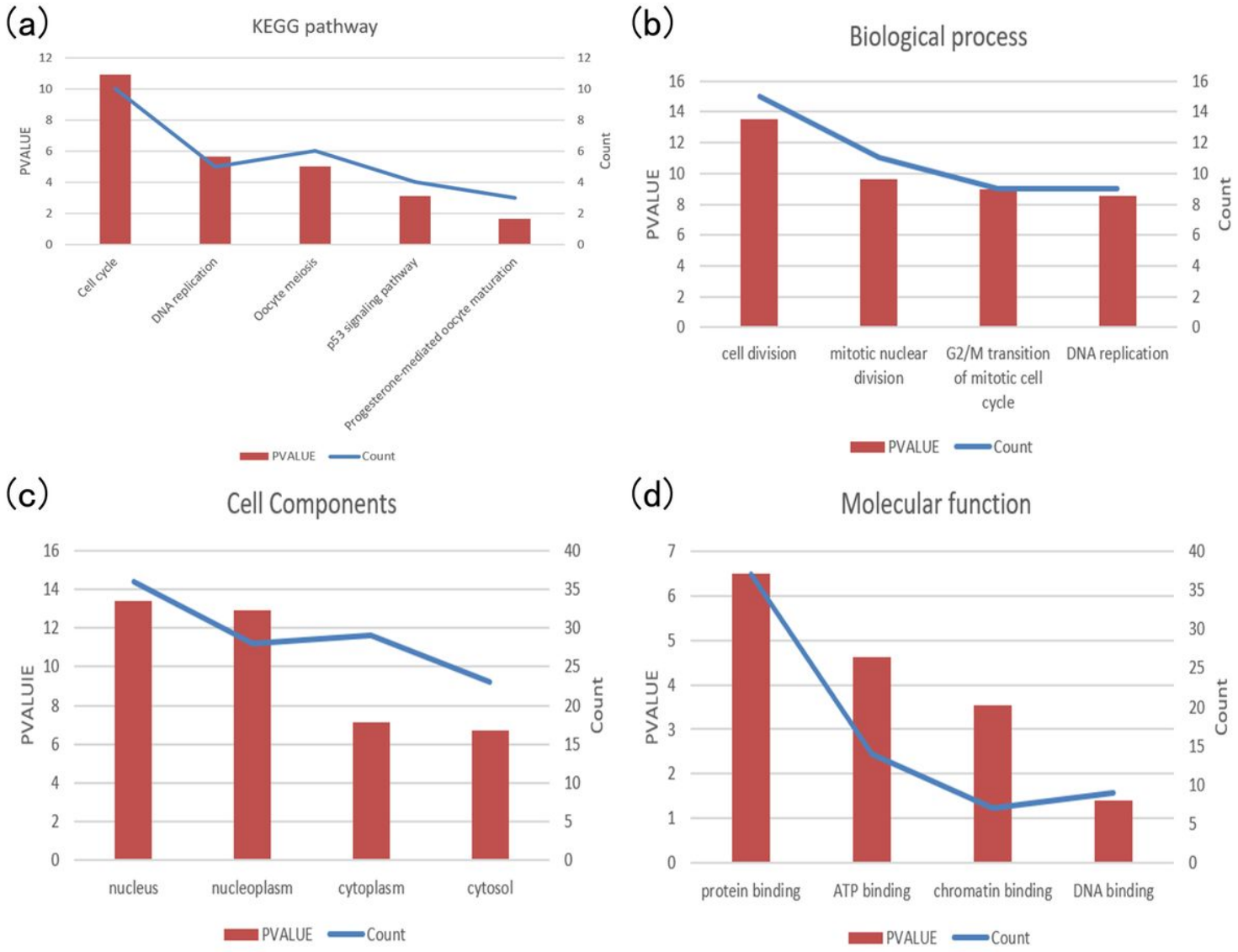

Figure 8

The result of the functional enrichment analysis of the genes in module 1

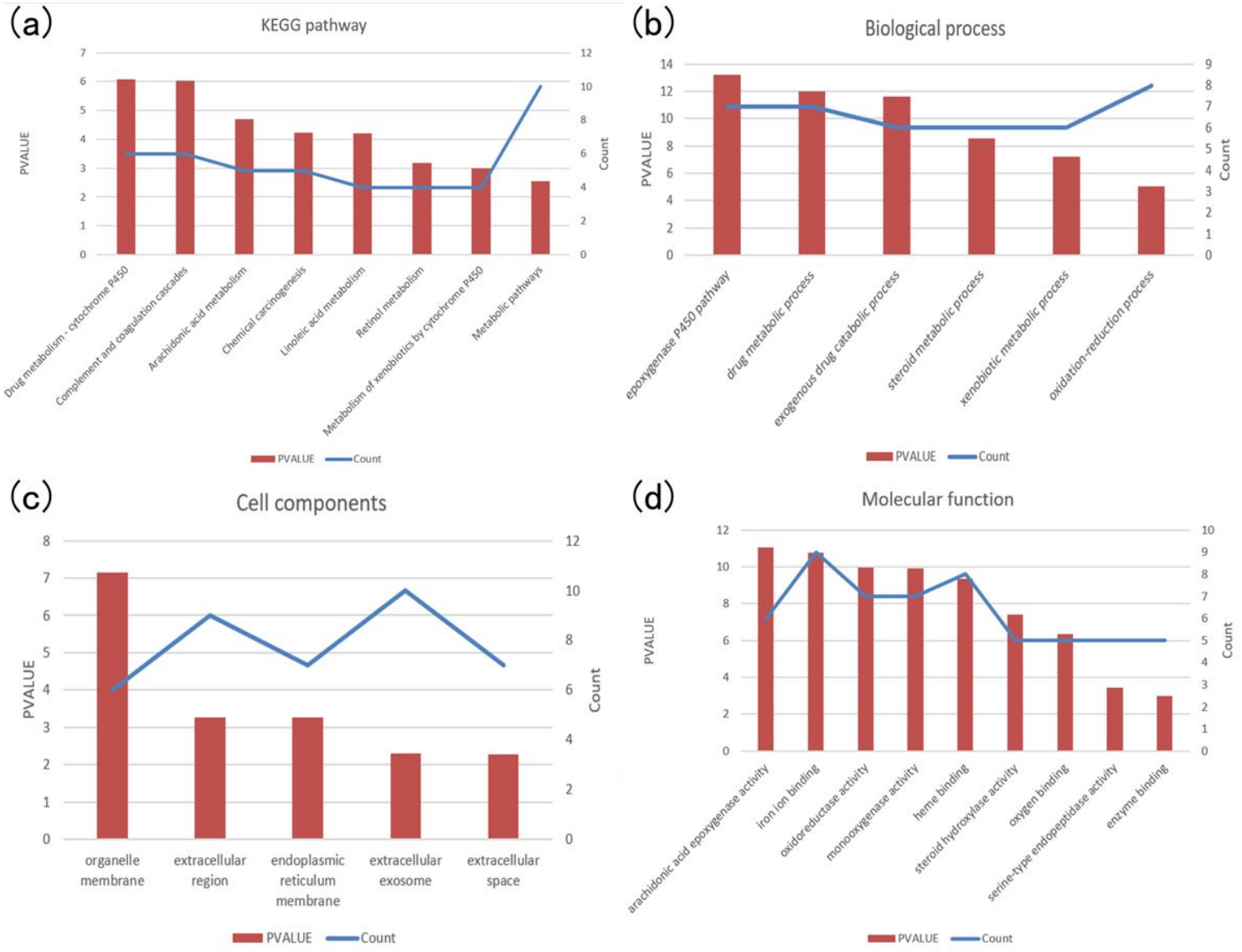

Figure 9

The result of the functional enrichment analysis of the genes in module 2

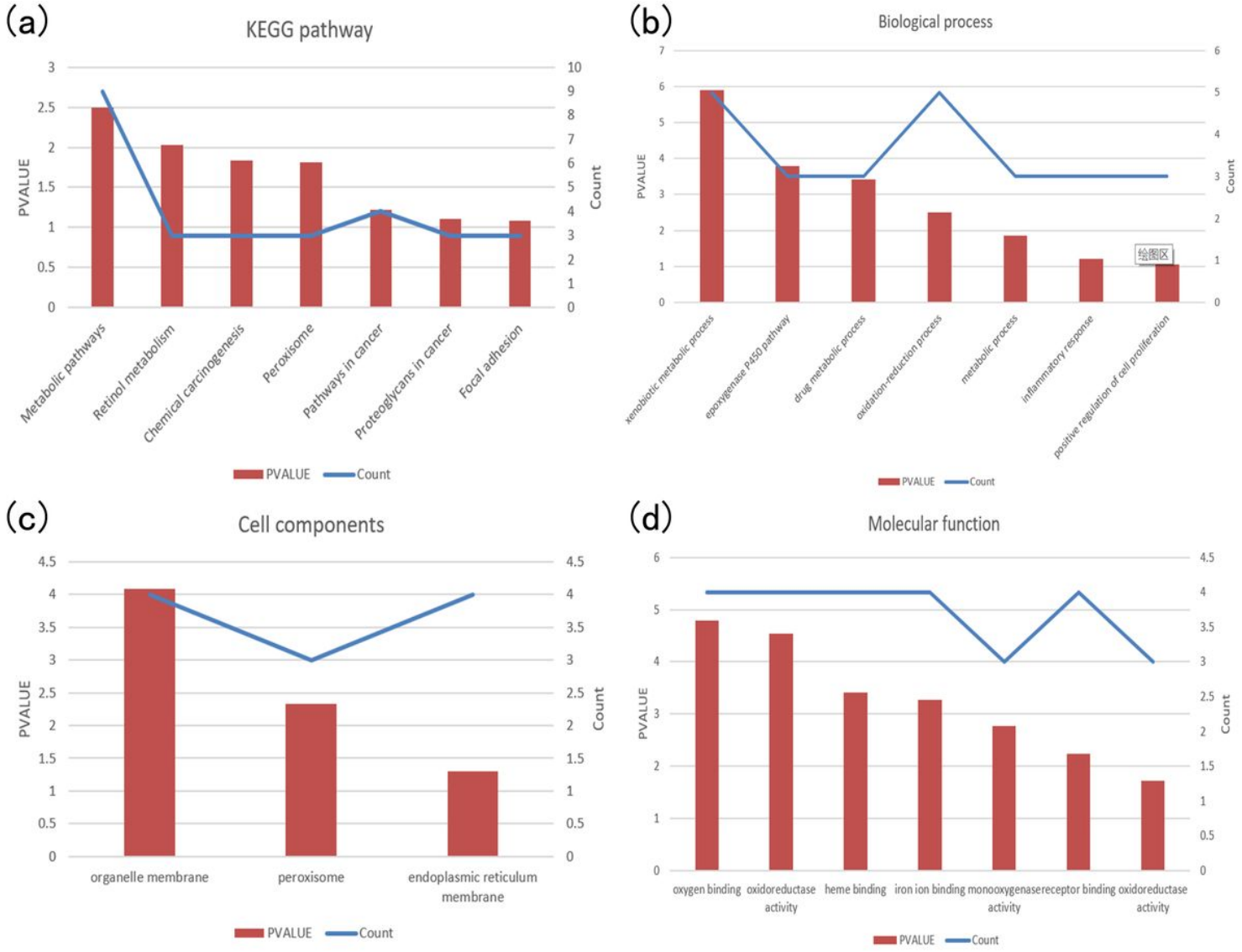

Figure 10

The result of the functional enrichment analysis of the genes in module 3

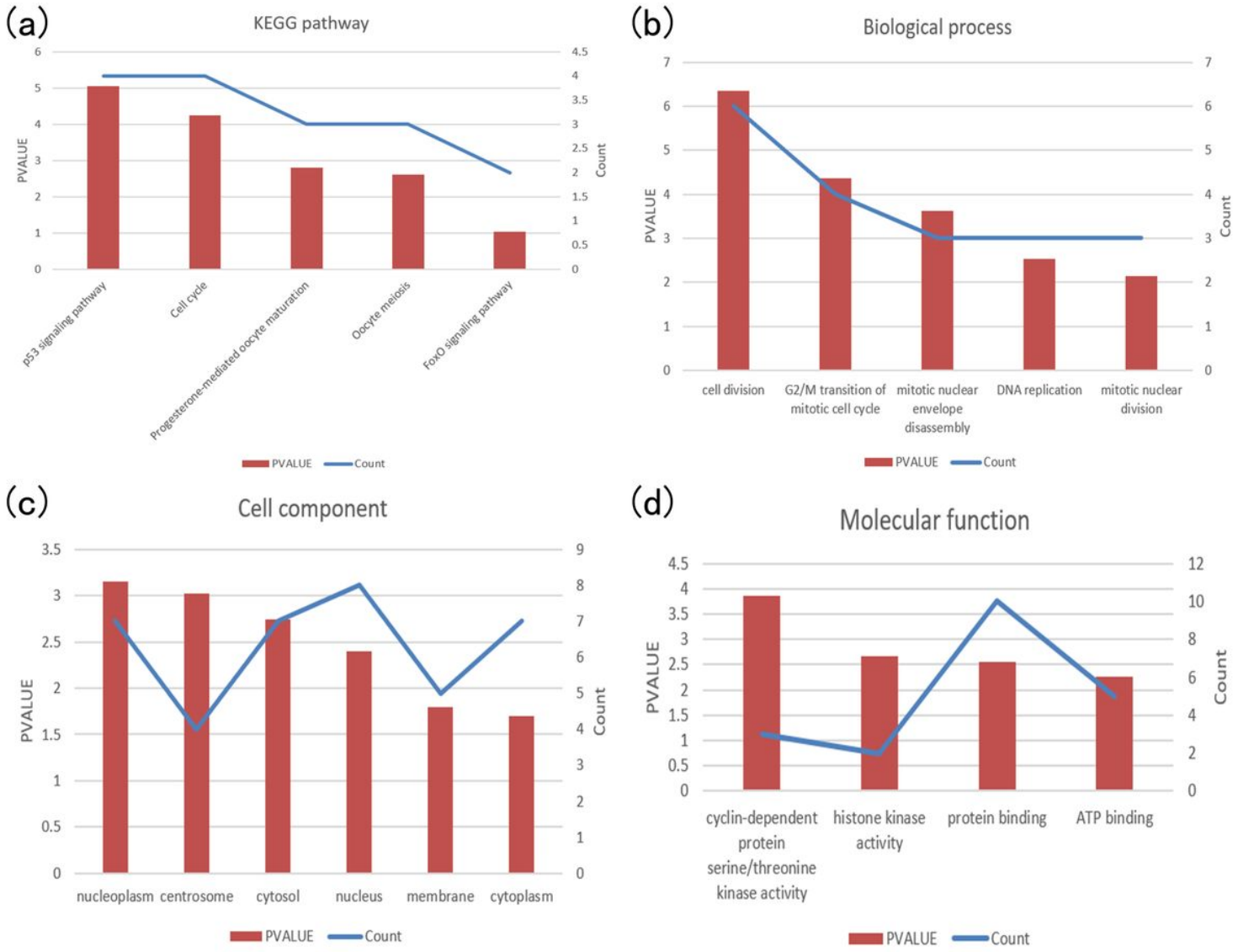

Figure 11

The result of the functional enrichment analysis of the top ten hub genes

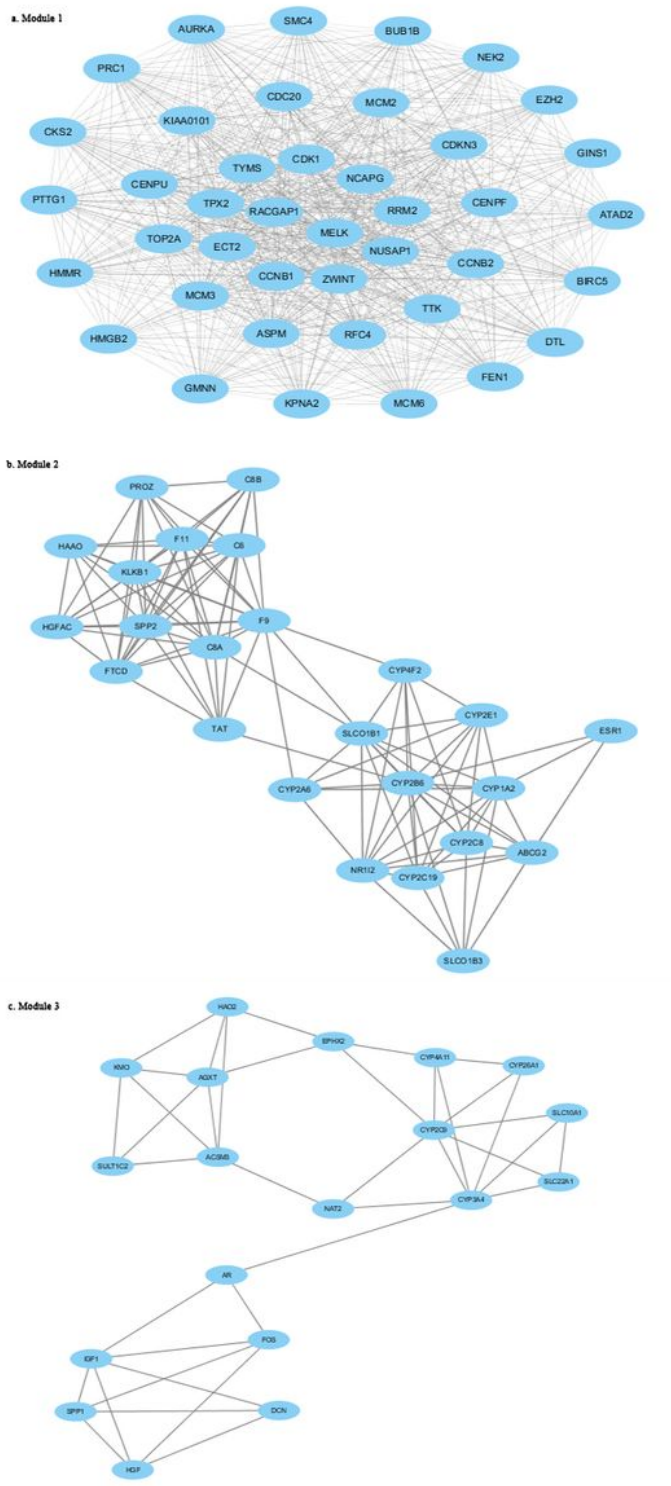

Figure 12

The module 1, module 2 and module 3

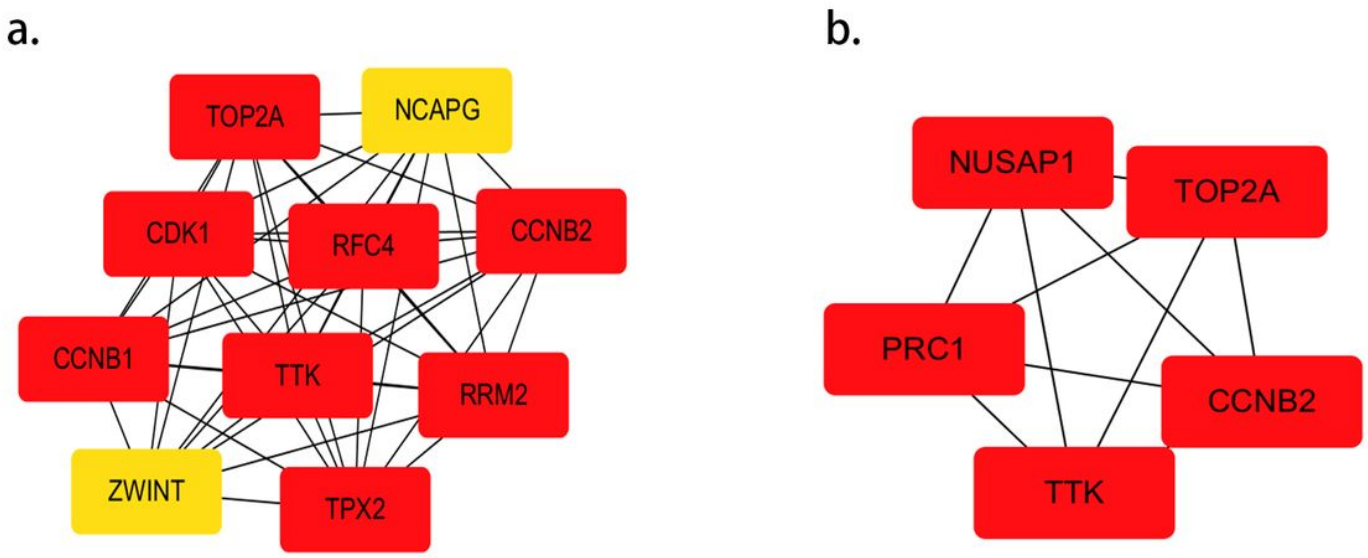

Figure 13

a) The top ten hub genes identified from the whole PPI network. b) The top five hub genes identified from module 1

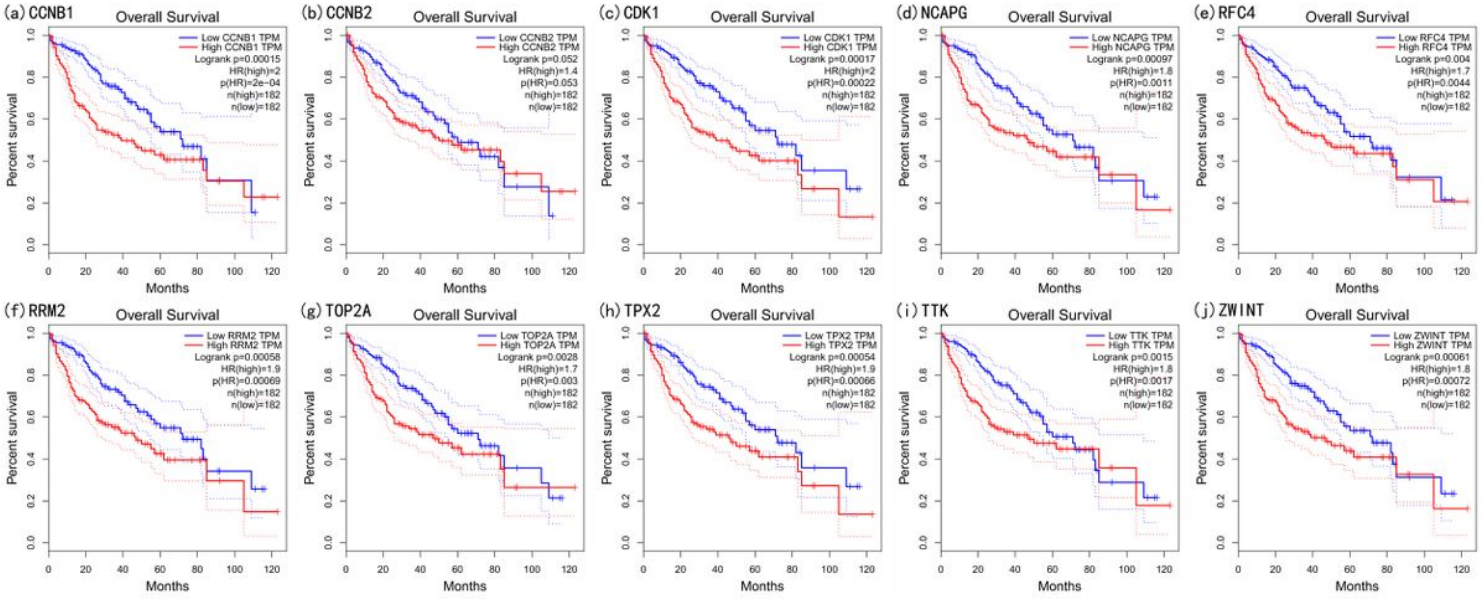

Figure 14

The overall survival outcome of the patients with top ten hub genes differentially expressed

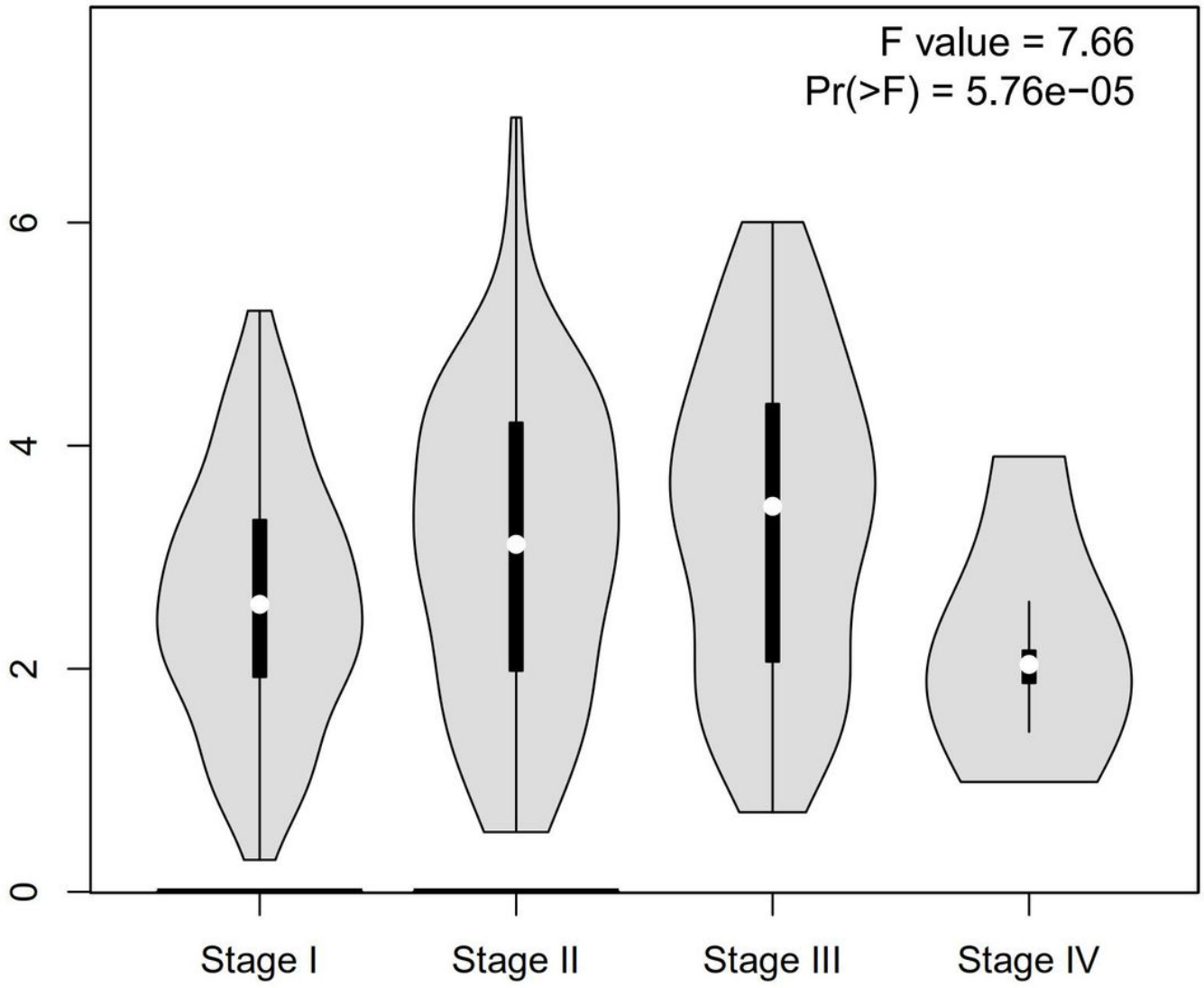

Figure 15

The expression level of CDK1 in different stages of HCC

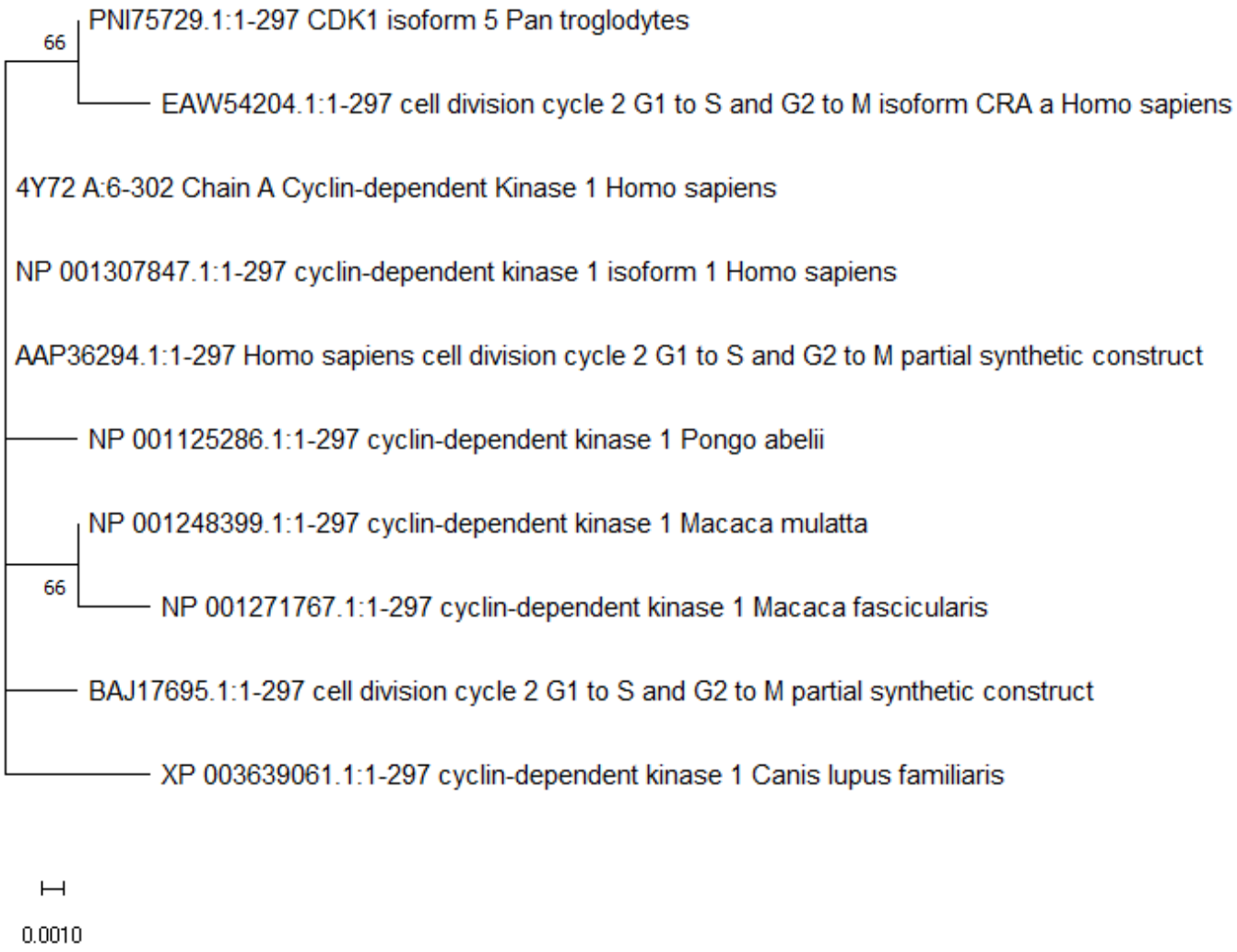

Figure 16

The phylogenetic tree of CDK1 protein family

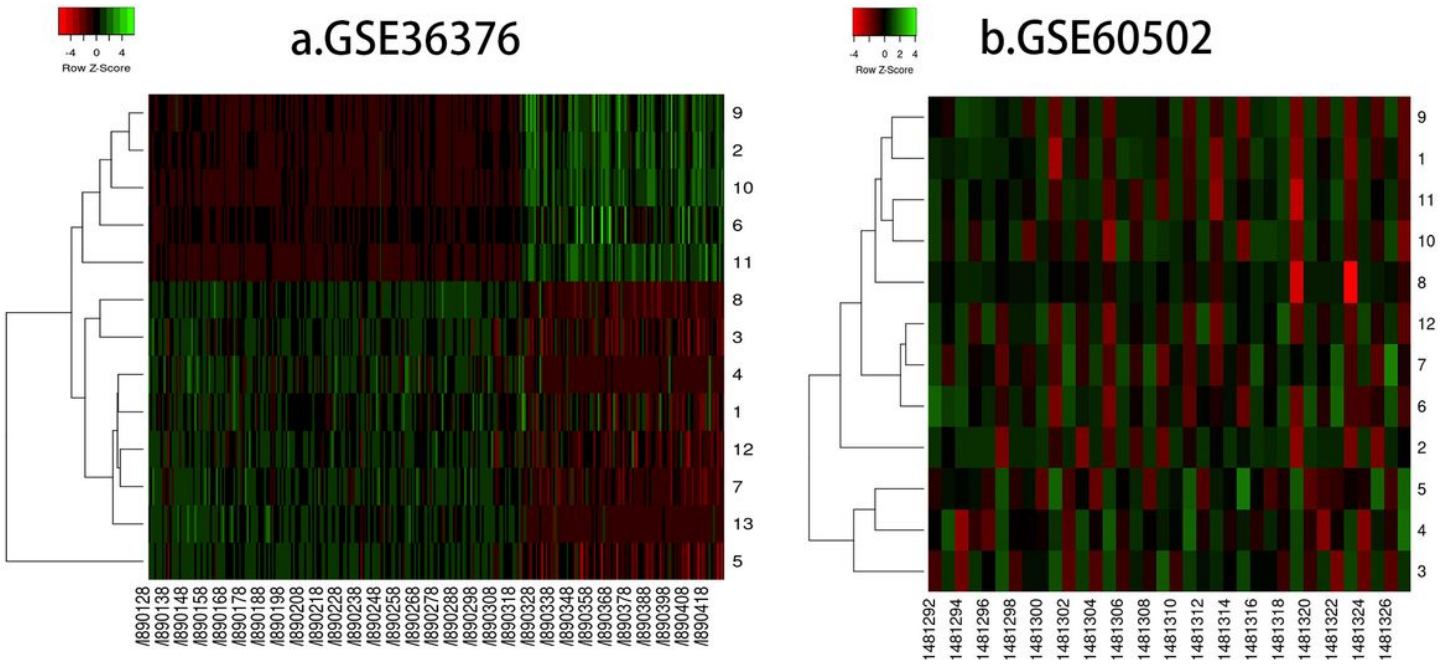

Figure 17

The expression heatmap of the genes chosen for validation in GSE36376 and GSE60502

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplymentaryfile.docx