Proceedings of ICMMGH 2026 Symposium: Cell and Molecular Biology: Understanding Immune Cell Functions
DOI: 10.54254/2753-8818/2026.PJ30862

Drug Repositioning of Ingenol as a Potential Lead Compound for Hepatocellular Carcinoma Therapy Based on WGCNA and CMap Analyse

Siyi Ding

College of Medicine, Henan University of Chinese Medicine, Zhengzhou, China
Dsy18019137136@outlook.com

Abstract. Hepatocellular carcinoma (HCC) is one of the leading causes of cancer-related deaths worldwide. The current commonly used therapies for HCC have severely restricted the further improvement of clinical efficacy due to problems such as postoperative metastasis and recurrence, resistance to targeted drugs, and limited response rates to immunotherapy. Against this backdrop, drug retargeting, with its advantages such as a short research and development cycle and known safety, provides a feasible path for expanding the treatment strategies of HCC. This study is based on the transcriptome data of the hepatocellular carcinoma (LIHC) project in the Cancer Genome Atlas (TCGA), and uses weighted gene co-expression network analysis (WGCNA) to construct a co-expression network to identify key gene modules significantly related to the characteristics of HCC. Through topological analysis and functional enrichment of genes within the module, the core gene set involved in the malignant progression of tumors was screened out. Furthermore, using the Connectivity Map (CMap) database, small molecule compounds that could significantly reverse the expression patterns of HCC characteristic modules were screened by calculating the reversal score. The calculation results show that several candidate molecules such as ingenol, exhibit high negative correlation scores in CMap analysis, suggesting that they have the potential to reverse the malignant epitopes of HCC at the transcriptome level.

Keywords: WGCNA, CMap, Hepatocellular Carcinoma, Drug Repositioning, Ingenol.

1. Introduction

Hepatocellular Carcinoma (HCC) is one of the leading causes of cancer-related deaths worldwide. According to the latest statistics, its incidence and mortality rates remain high [1]. Although therapies such as molecular targeted therapy and immune checkpoint inhibitors have provided certain clinical benefits for patients with advanced HCC, due to the high heterogeneity of tumors, frequent postoperative recurrence and metastasis, and drug resistance issues, the overall efficacy of existing treatment regimens remains unsatisfactory, and the long-term survival rate of patients still needs to be improved [2,3]. Therefore, developing new and effective treatment strategies has become an urgent need for current HCC research.

DOI: 10.54254/2753-8818/2026.PJ30862

Drug Repurposing provides an extremely attractive approach to expanding the therapeutic Arsenal of HCC by tapping into the therapeutic potential of approved or investigational drugs in new indications, thanks to its outstanding advantages of a short research and development cycle, high cost-effectiveness and known clinical safety data. In recent years, with the rapid development of bioinformatics and computational pharmacology, systematic drug relocation methods based on large-scale omics data have demonstrated great potential [4].

The accumulation of massive genomic data in public databases, especially the hepatocellular carcinoma (LIHC) project in the Cancer Genome Atlas (TCGA), has laid a solid data foundation for in-depth analysis of the molecular characteristics and pathogenesis of HCC at the systematic level [5]. Compared with the traditional differential expression analysis that focuses on the changes of individual genes, weighted gene co-expression network analysis (WGCNA) can construct a gene interaction network from the perspective of systems biology, identify functional gene modules that are highly coordinated with specific clinical phenotypes, and provide a new perspective for understanding the molecular heterogeneity of HCC and the underlying regulatory rules of synergy [6]. Studies have shown that WGCNA demonstrates outstanding efficacy in identifying HCC driver modules, prognostic hub genes, and potential therapeutic targets [7,8].

After successfully identifying the key disease modules, how to efficiently screen small molecule compounds that can specifically reverse the expression pattern of the characteristic genes of the disease is the core link of Drug Repurposing. The Connectivity Map (CMap) database, as a pioneering pharmacogenomics resource, integrates the whole-genome expression profiles generated after perturbation of thousands of small molecule compounds in different cell lines [9]. Through pattern matching algorithms, CMap can systematically evaluate the correlation between a given set of disease characteristic genes and the expression profile of drug-permed genes, thereby calculating candidate compounds that can 'reverse' disease characteristics [10]. This strategy of reverse screening drugs based on transcriptome features has been confirmed by multiple studies to be an effective method for discovering new indications for tumor treatment [11].

Natural products, due to their diverse structures and rich biological activities, have always been an important source of lead compounds for anti-tumor drugs [12]. Ingenol, as a natural diterpene compound derived from plants of the Euphorbia genus, has increasingly attracted attention for its unique biological functions. Previous studies have shown that it can act as a regulator of protein kinase C (PKC) and exhibit the ability to regulate apoptosis and differentiation in various cancer models [13], which lays the foundation for its functional exploration and drug retargeting research in HCC.

Based on the above background, this study aims to integrate the TCGA-LIHC transcriptome data with the CMap database to construct a systematic Drug Repurposing analysis process from disease module analysis to candidate compound screening. The key gene modules most relevant to the malignant phenotype of HCC were identified through WGCNA, and candidate small molecule compounds that could significantly reverse the expression pattern of this module were screened by CMap. The aim is to provide novel lead molecules for the development of drug therapy for HCC, and at the same time demonstrate the strategic value and application prospects of bioinformatics and computational pharmacology driving the discovery of tumor drugs.

2. Analysis of GEGs

This study systematically delineates the transcriptomic characteristics of HCC by leveraging RNA-seq data from the TCGA-LIHC project, employing a rigorous differential expression analysis pipeline.

DOI: 10.54254/2753-8818/2026.PJ30862

This paper made use of stringent significance p-value < 0.05 and fold-change |log2FC| > 1 cutoffs to discover 4,393 differentially expressed genes (DEGs), including 1,970 significantly upregulated and 2,423 significantly downregulated genes, as well as 30674 genes not differentially expressed significantly.

The visualization of volcano plot (Figure 1) offers the easy picture of the world distribution pattern of DEGs. Red markers are very highly upregulated genes and blue markers are very highly downregulated genes and gray markers are genes that did not pass the significance threshold. It is interesting to note that specific genes show extremely high levels of difference expression, which can imply their crucial regulatory functions in the development and evolution of HCC.

Such highly important DEGs which could be the main regulatory factors in the processes of the pathology of LIHC not only provide important information about the molecular processes underpinning the HCC but also forms a strong case to follow up on the study in the future.

Figure 1. Volcano map of the LIHC dataset (data from: TCGA-LIHC)

3. Weighted gene co-expression network analysis

The present research used the WGCNA algorithm to build a world large-scale gene-network considering transcriptomic analysis on TCGA-LIHC. This method is useful in determining higher-order regulatory interactions because it uses the topological overlap similarity to quantify similarities among the profiles of the gene expression, to overcome the drawback of a single-gene method in elucidating the complex interactions between genes, it provides new information about the multi-gene interactions to reveal the synergistic processes in the occurrence of HCC.

Firstly, hierarchical clustering analysis was done on all samples. Figure 2 meant the findings showed that normal phenotype samples were mostly concentrated in the middle part of the dendrogram, and that tumor phenotype samples are mainly scattered to the peripheral sides. Such clustering pattern means that there is a great level of consistency in the pattern of gene expression and the two phenotypes normal/tumor indicating different gene expression in normal liver tissue and hepatocellular carcinoma. These results do not only confirm that sample data is reliable with

DOI: 10.54254/2753-8818/2026.PJ30862

minimal outliers, but it also forms a critical basis to be applied to build quality co-expression network.

This study used both scale-free topology fitting index and mean connectivity to evaluate the optimal parameters of network construction in determining the optimal soft threshold parameter to be used. Figure 3 (left) clearly indicates that the wealth distribution state the scale free topology fitting Index increased to higher than 0.9 in the first time at the power of the soft threshold of 4, thus the constructed network met the scale free distribution features. At the same time, Figure 3 (right) shows that the mean connectivity of the network also decreased with a growing value of power. According to these two important pointers, a power value of 4 was eventually chosen as the best soft threshold, which would make the co-expression network developed have a biological importance as well as a statistical strength.

This was followed by the dynamic hybrid cut algorithm that hierarchically clustered the genes, and the genes with high similarity of patterns of gene expression were put into identical co-expression modules (Figure 4). The genes in these modules are usually highly functionally synergistic, which may be involved in certain signaling pathways or in the regulation of the same biological phenotypes.

The heatmap (Figure 5) used a continuous color gradient to indicate the topological overlap between modules: red spots are indicating high adjacency (adjacency values approach 1) meaning that the corresponding modules have highly similar gene expression patterns and could in fact be involved in coordinating specific biological processes or signaling pathways; blue spots are indicating low adjacency (adjacency values approach 0) therefore denoting vast difference between pattern of gene expression between modules that might then perform different biological functions; white spots indicate medium adjacency, so distinct functional relationship between modules.

As a systematic examination of the patterns of the heatmaps, this paper determined that MEblue module and the Tumor module had prominently red areas, which means that the degree of expression synergy between these modules is high and could be regarded as a fundamental functional network that controls the malignant phenotype of HCC.

In this paper, a module-phenotype association heatmap (Figure 6) was created based on computations of the Spearman coefficients of correlation between co-expression modules and clinical phenotypes. It was found that several gene modules were strongly related to the tumor phenotypes of HCC and the MEblue module has the strongest correlation features. In particular, a strong negative correlation of MEblue module was observed with normal liver tissue (correlation coefficient = -0.75, P = 3×10-78) and strong positive correlation with tumor tissue (correlation coefficient = 0.75, P = 3×10-78). This extremely important association pattern demonstrates that the genes in the MEblue module are typically lowly expressed in normal liver tissue but are highly expressed in the HCC tissue, which partially demonstrates that this group of genes can play a role in the overall malignant development of HCC. As an additional confirmation of the pivotal role of MEblue module in the development of the pathogenesis of HCC, this article formed a scatterplot (Figure 7) by the analysis of Gene Significance (GS) and Module Membership (MM) associations. The findings demonstrated that GS and MM values had a strong positive interrelations in this module whereby the genes that tend to relate more to the HCC phenotypes are highly associated in the module, again illustrating the hub role of the MEblue module in the HCC transcriptional regulatory network. On the above analysis, MEblue module was found as the functional unit that is closely related with the HCC malignant phenotypes. Not only is this module abundantly represented by abnormally expressed genes in HCC, but also the genes in this module are highly interacting with one another in the regulatory network, most likely, they are collectively involved in some of the

DOI: 10.54254/2753-8818/2026.PJ30862

essential processes involved in carcinogenesis, including cell cycle regulation, proliferation and apoptosis resistance. That is why MEblue module is deemed the most important target one that requires further in-depth examination of HCC molecular mechanisms, hub genes screening, and potential therapeutic targets exploration.

Based on the network topology (Figure 8) analysis, it is clear that genes that represent the same module constitute high color-coded clusters in the network, and extensive regions of dotted fluorescence are seen in each of these modules. It is observed that topological overlap of genes within the same module and strong functional connectivity are high. One of the most notable is the MEblue module that displays the most focal points of a strong and positive interconnection, which additionally supports its central position in the transcriptional regulatory network of HCC.

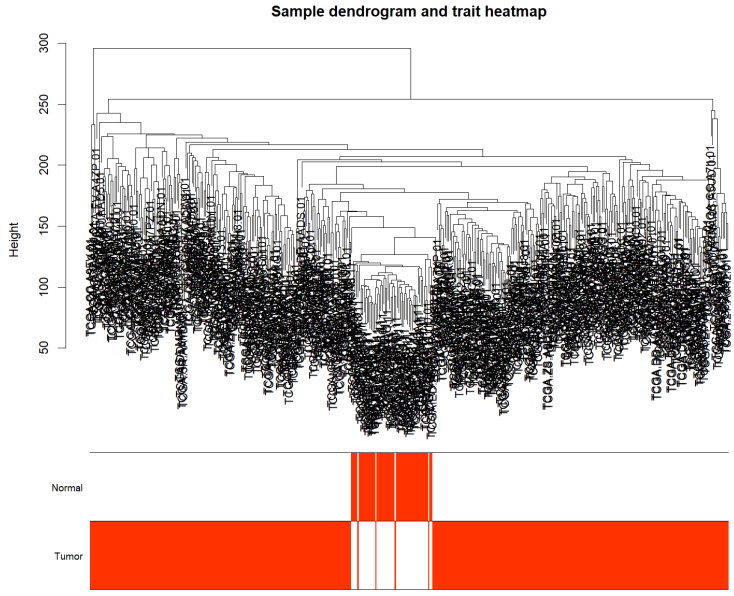


Figure 2. Sample dendrogram and trait heatmap (picture credit: original)

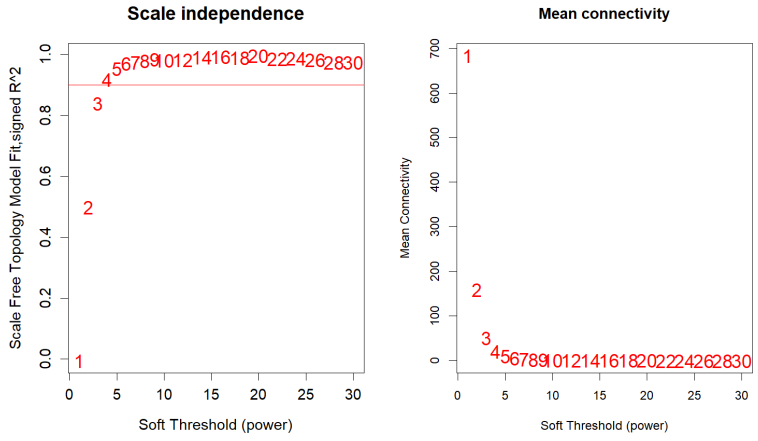


Figure 3. Optimal soft threshold (picture credit: original)

DOI: 10.54254/2753-8818/2026.PJ30862

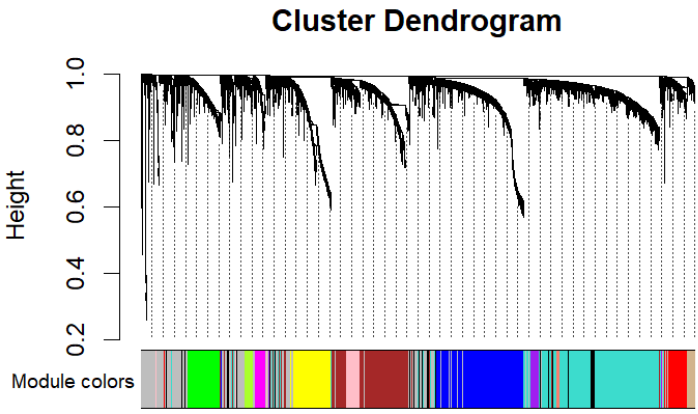


Figure 4. Cluster dendrogram (picture credit: original)

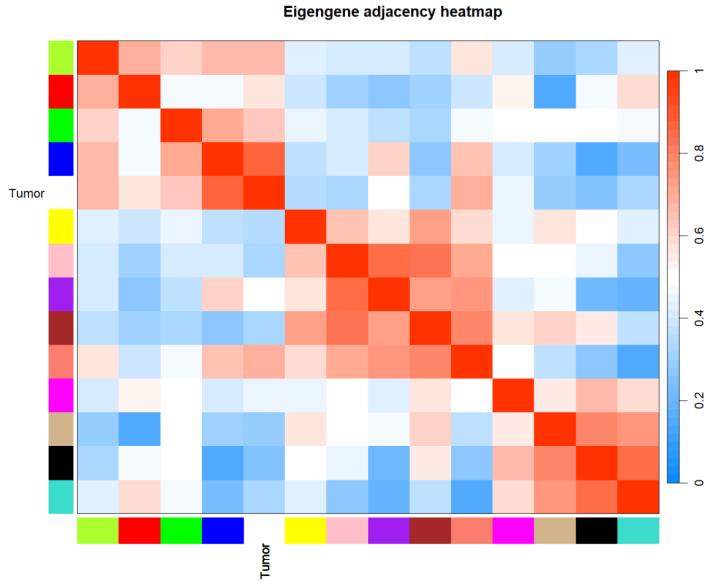


Figure 5. Eigengene adjacency heatmap (picture credit: original)

DOI: 10.54254/2753-8818/2026.PJ30862

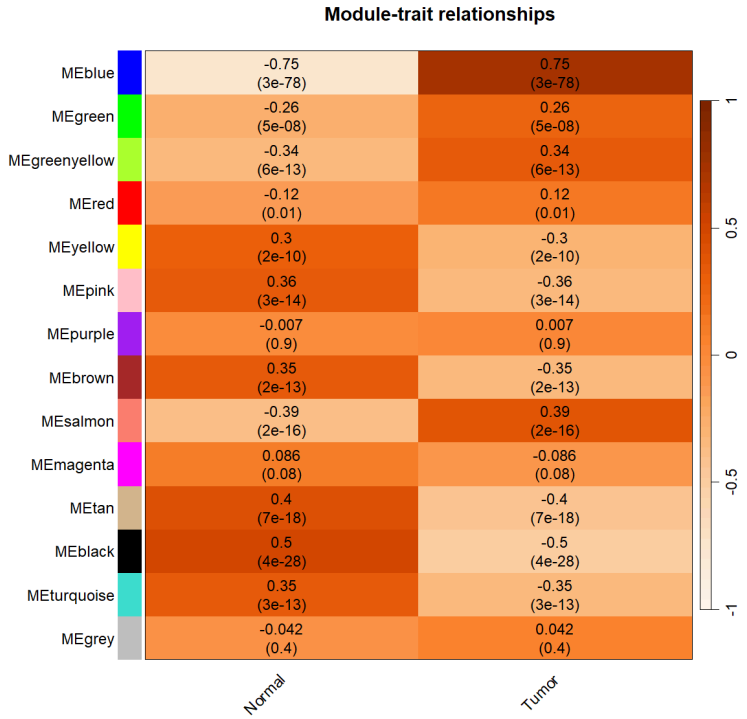


Figure 6. Module-trait relationships (picture credit: original)

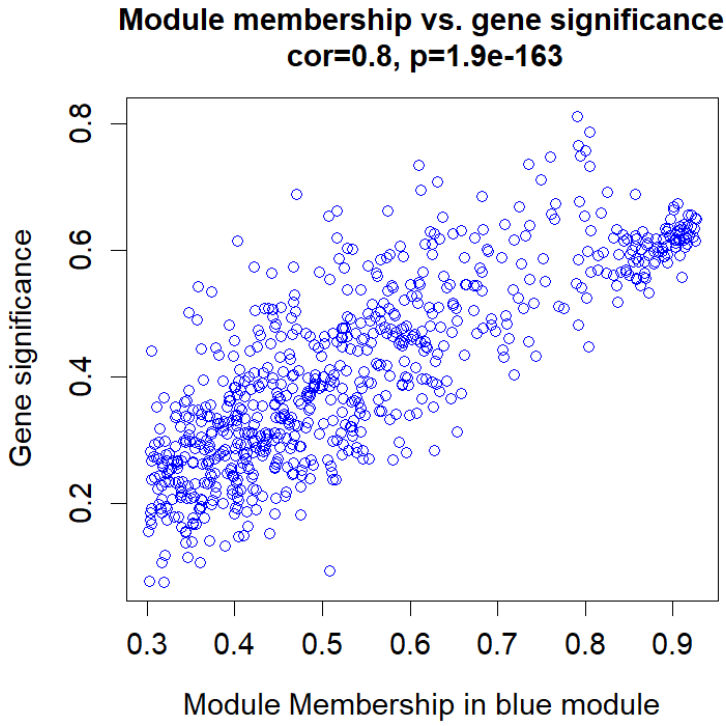


Figure 7. Gene significance (picture credit: original)

DOI: 10.54254/2753-8818/2026.PJ30862

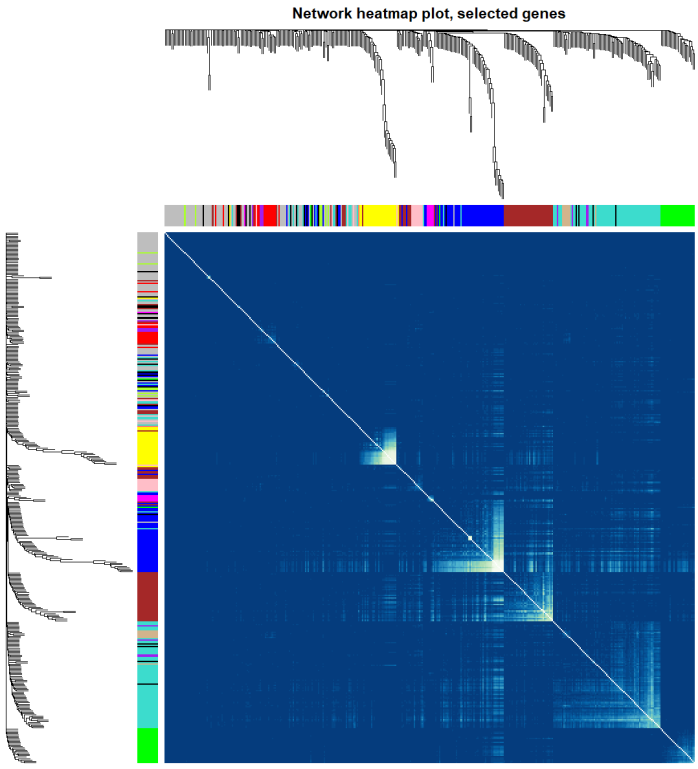


Figure 8. Network heatmap plot, selected genes (picture credit: original)

4. Identification of key genes in HCC

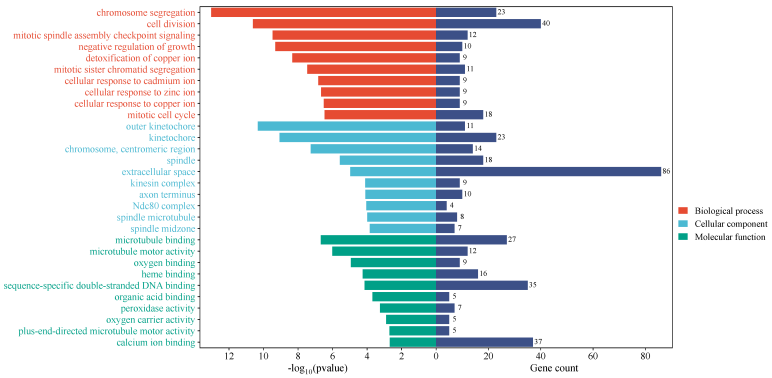
The analysis revealed that this study actually determined a fundamental set of genes of 696 elements via Venn diagram analysis that paired with the criteria of differentiation expression as well as was in the MEblue module (Figure 9). In this gene set, there are 365 up-regulated genes and 331 down-regulated genes constituting an essential target of exploring the HCC mechanisms at the molecular level. These genes do not only show consistent patterns of differentiation expression in various individuals but also have highly synchronized regulatory networks in the co-expression network, implying that they may has some collective role in powering some important pathological events of HCC.

Figure 9. Intersection map of key module genes and DEGs (picture credit: original)

DOI: 10.54254/2753-8818/2026.PJ30862

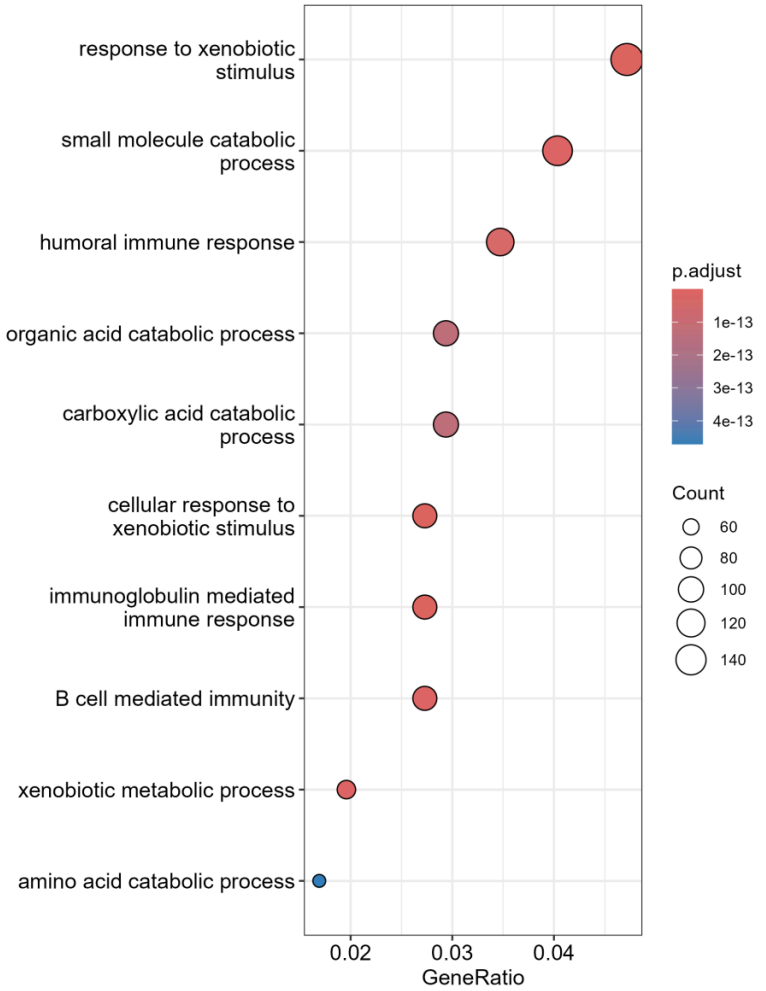
5. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis

The biological foundation of MEblue module was logically explained under both functional and pathway through GO and KEGG enrichment analysis of 696 core candidate genes identified in HCC screening. Figure 10, which was GO analysis, showed that these genes were significantly in enrichment of proliferation related biological processes (cell mitosis and chromosome segregation) and cellular structures (the spindle apparatus) and molecular functions, particularly DNA binding and microtubule binding, meaning that the MEblue module mainly plays a role in regulating the cell cycle a mechanism that is now central to HCC malignant phenotypes. Response to xenobiotic stimulus was the most significant biological process and co-enrichment with such entries as small molecule catabolic process unveiled both important peculiarities of metabolic re-synthesis in HCC. KEGG pathway analysis (Figure 11) also further validated that the cell cycle pathway was ranked topmost with respect to the number of differentially expressed genes and enrichment significance which supports the high correlation between MEblue module in WGCNA and tumor phenotypes. In the meantime, the most important pathway, namely the cytotoxin -cytotoxin receptor interaction, and the co-enrichment of pathways, including the drug metabolism - cytochrome P450, wrote an HCC molecular network between the control of immune microenvironment and xenobiotic metabolism. The aggregate outcomes of these enrichment activities explain the multidimensional contributions to the development of socioeconomic theories of the MEblue module that innovates HCC malignant disease.



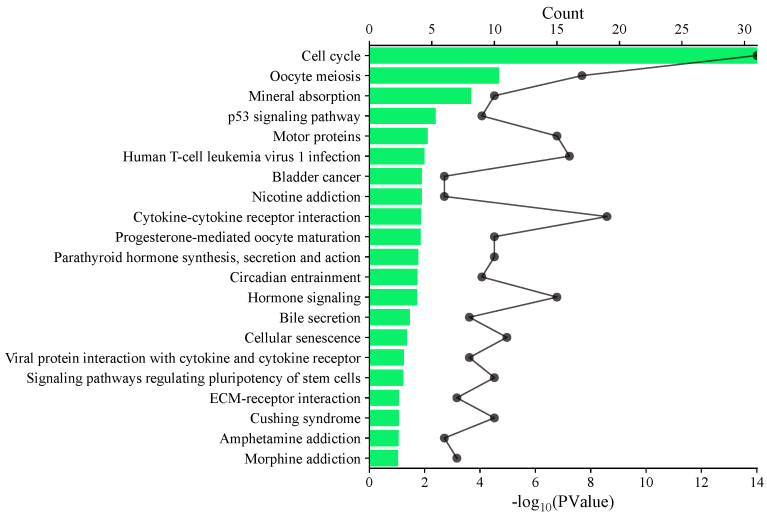
(a)

DOI: 10.54254/2753-8818/2026.PJ30862



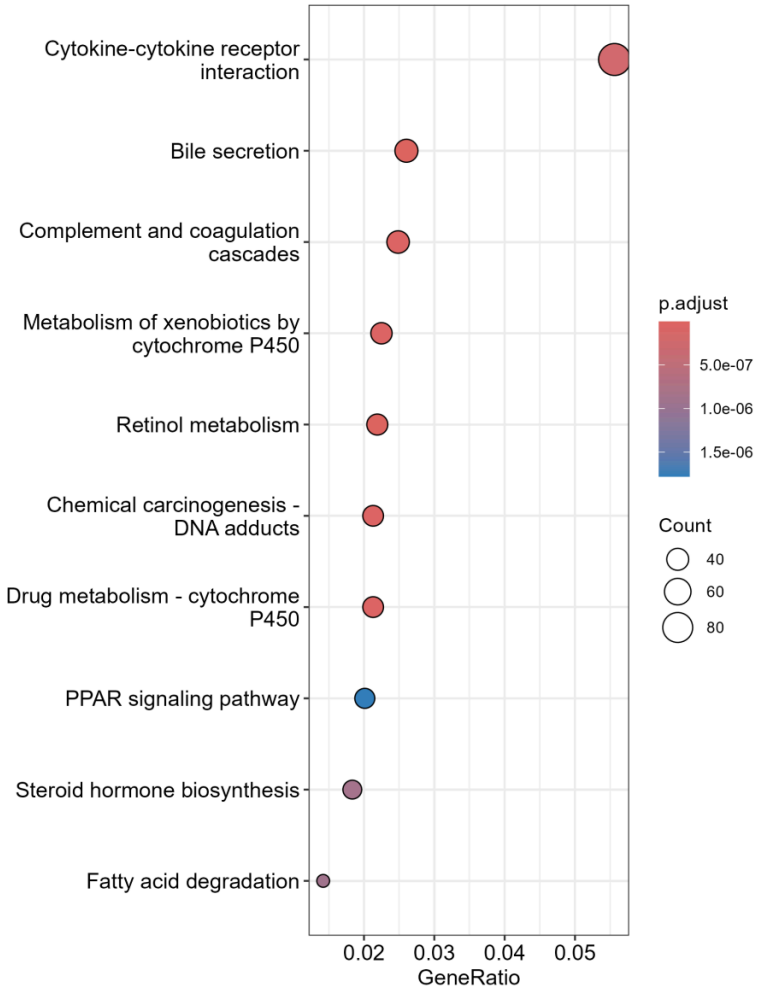
(b)

Figure 10. GO enrichment analysis (picture credit: original)



(a)

DOI: 10.54254/2753-8818/2026.PJ30862



(b)

Figure 11. KEGG enrichment analysis (picture credit: original)

6. The connectivity map

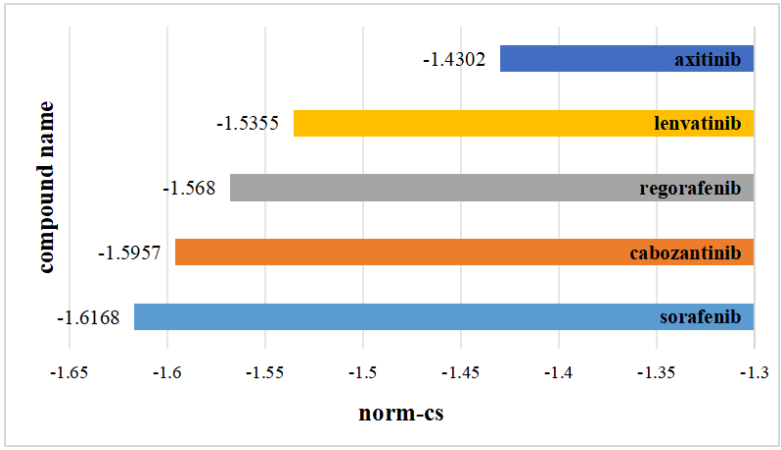
Based on the core signature of gene expression of HCC developed earlier, this paper is attempting a systematic analysis of drug repositioning with the help of the Connectivity Map (CMap) database. A subset of small-molecule compounds with a potential of reversing the HCC gene expression patterns were discovered by computing the negative connectivity scores between the drug-induced and HCC disease signatures (Figure 12).

The results of the analytic showed that ingenol had a very high negative score of connectivity (-1.9934) in the CMap screening, which shows that it could potentially be useful in reversing the expression patterns of HCC characteristic gene modules. A systematic comparison of ingenol against the therapeutic agents of HCC, which are currently in clinical use, was done to determine the therapeutic potential of ingenol. The scores of sorafenib, cabozantinib, regorafenib, lenvatinib and axitinib with the negative connectivity were -1.6168, -1.5957, -1.56, -1.5355 and -1.4302 respectively. It was apparent in the comparative analysis that ingenol had better reversed HCC disease gene signatures.

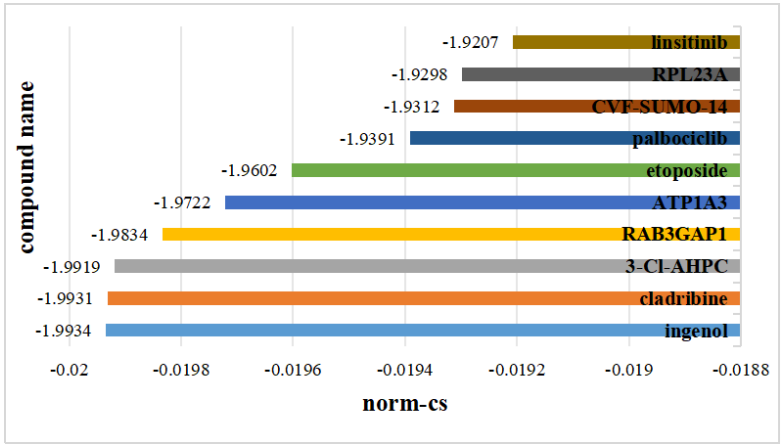
DOI: 10.54254/2753-8818/2026.PJ30862

This finding can be of importance in translationational medicine. Ingolenol is a natural and diterpenoid molecule and has been previously known to have anti-proliferative and pro-differentiation properties in tumor polymorphs as a protein kinase C (PKC) modulator. The outcomes of the CMap analysis in the current research also widen its possible applications, which recommends that this compound can be used as an anti-HCC agent by modulating the MEblue module-related signaling that can involve cytokine-receptor interaction network in particular.

On this basis of this computational pharmacological evidence, ingenol has been re-positioned as a first hand lead compound in treating HCC. The repositioning of drugs discovery is important as in addition to introducing a novel candidate drug in the treatment of HCC, it justifies the utility of the combined WGCNA/CMap strategy of computational pharmacology in drug discovery research as a viable methodology that can be used to systematically discover new treatment indications in cancer treatment.



(a)



(b)

Figure 12. CMap negative score sheet (picture credit: original)

7. Protein-protein interaction network and core gene screening

This study constructed a protein-protein interaction (PPI) network using the STRING database and performed topological analysis with Cytoscape software, ultimately identifying five hub genes:

DOI: 10.54254/2753-8818/2026.PJ30862

CCNB1, CDK1, CCNA2, TOP2A, and PLK1 (Figure 13). The analysis of network topology showed that the five genes possessed even more values of node degree (degree > 25) and occupied the central position in the PPI network, which means that the genes are of high regulatory importance in the hepatocellular carcinoma (HCC) protein interaction network. Remarkably, general pattern, these hub-genes belonged to the already discovered MEblue module and their expression was more significantly up-regulated in the HCC tissues (logFC>4, FDR< 0.001), which also provided further evidence of their critical roles in the HCC pathogenesis.

At the functional level, these five hub genes primarily participate in core processes of cell cycle regulation: CCNB1, CDK1, and CCNA2 constitute key regulatory components of the G2/M checkpoint; TOP2A is involved in DNA replication and chromosome segregation; and PLK1 regulates centrosome maturation and mitotic progression. The coordinated overexpression of these genes in HCC may collectively drive abnormal tumor cell proliferation and genomic instability.

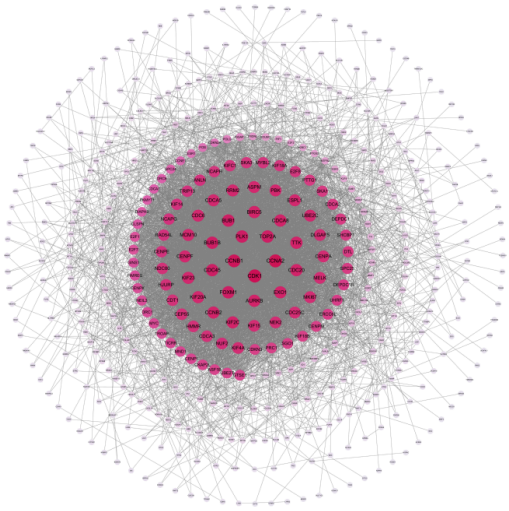


Figure 13. Protein-protein interaction network (picture credit: original)

8. Integrated analysis of drug-target-pathway interaction network

This study constructed a multidimensional "drug-target-pathway" interaction network (Figure 14), which systematically demonstrates the complete regulatory network of ingenol's intervention in HCC pathological progression through modulation of key genes and downstream signaling pathways. The core architecture of this network comprises four hierarchical levels: red diamond nodes represent the target drug ingenol (DRUG), green rectangular nodes denote the disease (DISEASE), purple circular nodes indicate HCC-upregulated key genes (e.g., CDK1, CCNB1, TOP2A), and yellow circular nodes represent HCC-downregulated key genes. Green arrow-shaped labels surrounding the network annotate relevant biological pathways and disease phenotypes.

Analytical findings indicate that ingenol triggers in the HCC fundamental signalling structures by controlling several vital gene nodes. The drug interferes with the important cyclins like CDK1 and CCNB1 in terms of controlling the cell cycle directly, thus, influencing the pathway of the cell cycle. Concurrently, the network shows that ingenol adjusts the gene nodes such as TP73 and CDKN2A which regulate the "p53 signaling pathway" indicating its possible role as an anti-tumor agent by acting to restore p53-regulated functions of cell cycle arrest and apoptosis.

Not only does this "drug-target-pathway" network diagram confirm the direct effects of ingenol on the core cell cycle targets, it increases the systematic pharmacological action of ingenol using

DOI: 10.54254/2753-8818/2026.PJ30862

multi-target and multi-pathway interaction. This network gives a holistic insight on the molecular mechanisms of the anti-HCC action of ingenol.

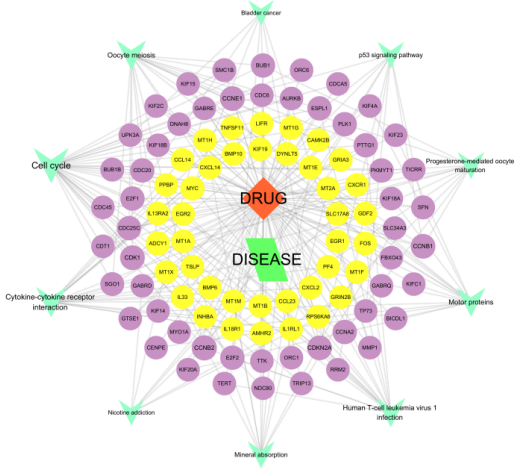


Figure 14. Drug-target-pathway interaction network (picture credit: original)

9. Conclusion

This research proposal systematically developed a research framework that incorporates the approaches of bioinformatics and computational pharmacology, including molecular characterization of diseases to drug repositioning screening. TCGA-LIHC transcriptomic data was analyzed and containing 4,393 differentially expressed genes related to HCC, WGCNA was constructed to obtain a gene co-expression network, among which MEblue module exhibited very high positive correlation with tumor phenotype. This module was independently functional-analysed in depth to identify important enrichment of genes in core biological functions such as cell cycle control, DNA replication and metabolism stress responses to offer a systematic view of the HCC pathogenesis. Ingenol was selected with CMap analysis as the most promising therapeutic candidate of HCC as it is more effective in reversing disease phenotypes than currently available clinical drugs. Analysis of protein-protein interaction networks also revealed five hub genes (CCNB1, CDK1, CCNA2, TOP2A and PLK1), which were significantly up regulated and had a role to play in the process of cell cycle regulation. This study demonstrated that Ingenol can affect the key signaling pathways in HCC by a multi-target mechanism of action by building an integrated network linking drugs and their targets.

The study has created a stepwise analytical methodology of identifying disease modules to the elucidation of drug mechanism that can comprehensively annotate the functional aspects to the elucidation of the biological basis of MEblue module-driven HCC malignant progression, and also reposition Ingenol as a therapeutic candidate to the treatment of HCC and the clarification of its multi-target mechanism of action.

References

[1] Sung, H., Ferlay, J., Siegel, R.L., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 71, 209 - 249.

[2] Llovet, J. M., Kelley, R. K., Villanueva, A., Singal, A. G., Pikarsky, E., Roayaie, S., Lencioni, R., Koike, K., Zucman-Rossi, J., & Finn, R. S. (2021). Hepatocellular carcinoma. Nature reviews. Disease primers, 7(1), 6.

DOI: 10.54254/2753-8818/2026.PJ30862

[3] Llovet, J. M., Castet, F., Heikenwalder, M., et al. (2022). Immunotherapies for hepatocellular carcinoma. Nature reviews. Clinical oncology, 19(3), 151–172.
[4] Pushpakom, S., Iorio, F., Eyers, P. A., et al. (2019). Drug repurposing: progress, challenges and recommendations. Nature reviews. Drug discovery, 18(1), 41–58.
[5] Wu, M., Liu, Z., Zhang, A., & Li, N. (2019). Identification of key genes and pathways in hepatocellular carcinoma: A preliminary bioinformatics analysis. Medicine, 98(5), e14287.
[6] Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics, 9, 559.
[7] Li, B., Pu, K., & Wu, X. (2019). Identifying novel biomarkers in hepatocellular carcinoma by weighted gene co-expression network analysis. Journal of cellular biochemistry, 120(7), 11418–11431.
[8] Liu, P., Li, H., Liao, C., Tang, Y., Li, M., Wang, Z., Wu, Q., & Zhou, Y. (2022). Identification of key genes and biological pathways in Chinese lung cancer population using bioinformatics analysis. PeerJ, 10, e12731.
[9] Subramanian, A., Narayan, R., Corsello, S. M., et al. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1, 000, 000 Profiles. Cell, 171(6), 1437–1452.e17.
[10] Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J. P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S., & Golub, T. R. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science (New York, N.Y.), 313(5795), 1929–1935.
[11] Musa, A., Ghoraie, L. S., Zhang, S. D., et al. (2018). A review of connectivity map and computational approaches in pharmacogenomics. Briefings in bioinformatics, 19(3), 506–523.
[12] Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., International Natural Product Sciences Taskforce, & Supuran, C. T. (2021). Natural products in drug discovery: advances and opportunities. Nature reviews. Drug discovery, 20(3), 200–216.
[13] Kedei, N., Lundberg, D. J., Toth, A., Welburn, P., Garfield, S. H., & Blumberg, P. M. (2004). Characterization of the interaction of ingenol 3-angelate with protein kinase C. Cancer research, 64(9), 3243–3255.