Send Orders for Reprints to reprints@benthamscience.net

Current Genomics, **2019**, *20,* 545-555

RESEARCH ARTICLE

Deciphering the Novel Target Genes Involved in the Epigenetics of Hepatocellular Carcinoma Using Graph Theory Approach

Nimisha Roy[1], Utkarsh Raj[2], Sneha Rai[3] and Pritish K. Varadwaj[1,*]

[1]*Department of Bioinformatics and Applied Sciences, Indian Institute of Information Technology, Allahabad, Uttar Pradesh, India;* [2]*Department of Biotechnology and Bioinformatics, NIIT University, Neemrana, Rajasthan, India;* [3]*Division of Biotechnology, Netaji Subhas Institute of Technology, University of Delhi, New Delhi, India*

ARTICLE HISTORY

Received: April 25, 2019
Revised: September 25, 2019
Accepted: November 10, 2019

DOI:
10.2174/1389202921666191227100441

Abstract: ***Background*:** Even after decades of research, cancer, by and large, remains a challenge and is one of the major causes of death worldwide. For a very long time, it was believed that cancer is simply an outcome of changes at the genetic level but today, it has become a well-established fact that both genetics and epigenetics work together resulting in the transformation of normal cells to cancerous cells.

***Objective*:** In the present scenario, researchers are focusing on targeting epigenetic machinery. The main advantage of targeting epigenetic mechanisms is their reversibility. Thus, cells can be reprogrammed to their normal state. Graph theory is a powerful gift of mathematics which allows us to understand complex networks.

***Methodology*:** In this study, graph theory was utilized for quantitative analysis of the epigenetic network of hepato-cellular carcinoma (HCC) and subsequently finding out the important vertices in the network thus obtained. Secondly, this network was utilized to locate novel targets for hepato-cellular carcinoma epigenetic therapy.

***Results*:** The vertices represent the genes involved in the epigenetic mechanism of HCC. Topological parameters like clustering coefficient, eccentricity, degree, *etc*. have been evaluated for the assessment of the essentiality of the node in the epigenetic network.

***Conclusion*:** The top ten novel epigenetic target genes involved in HCC reported in this study are cdk6, cdk4, cdkn2a, smad7, smad3, ccnd1, e2f1, sf3b1, ctnnb1, and tgfb1.

Keywords: Graph theory, clustering coefficient, eccentricity, epigenome, stress centrality, hepatocellular carcinoma.

1. INTRODUCTION

C.H. Waddington's [1] definition of epigenetics illustrating it as the 'study of interactions between genes and the products thereof leading to the formation and existence of phenotype' was initially concerned with the role of epigenetics in fetal development. However, the definition of epigenetics has evolved over time as it is implicated in a wide variety of biological processes. The contemporary definition of epigenetics is 'the study of heritable and reversible changes in gene expression that occur independently of changes in the primary DNA sequence' *i.e.* somatic changes. Most of these heritable changes are established during differentiation and are stably maintained through multiple cycles of cell division, enabling cells to have distinct identities, while containing the same genetic information. This heritability of gene expression patterns is mediated by epigenetic modifications, which include methylation of CpG regions in DNA, post translational Histone modification, nucleosome positioning along with the DNA and micro-RNA modification.

Epigenome is an assembly of chemical entities that allocates and guides the genome to perform the intended function. DNA constitutes the human genome, whereas the chemical compounds and proteins constitute the epigenome. These compounds and proteins attach to the DNA and perform functions like turning genes on or off as well as controlling the rate of production of proteins. Cancers are caused as a cumulative effect of the genome and epigenome [2, 3].

For a very long time, it was assumed that genetics and epigenetics are two independent mechanisms [3, 4]. Mutations in genetic and epigenetic mechanisms lead to cancer development and promote cancer progression [5]. DNA methylation, histone modification and micro-RNA changes are found to be the important biomarkers of the initial stages in many forms of cancer [6]. The fact that epigenetic aberrations, unlike genetic mutations, are potentially reversible and can be reprogrammed to their normal state by epigenetic therapy makes such initiatives promising and therapeutically relevant [7].

The three most studied epigenetic mechanisms include (a) DNA methylation, (b) histone modifications (both covalent and non-covalent) and (c) micro RNA or miRNA. The process of addition of methyl (CH_3) groups to the DNA molecule is known as DNA methylation. It can change DNA activity without altering the sequence. Hence, it is a crucial factor in epigenetics [8]. At normal levels, DNA methylation is required for normal development and inactivation of various processes like X- chromosome inactivation. When a CpG cluster region at the promoter site of a gene is methylated, expression of the gene is turned off.

*Address correspondence to this author at the Department of Bioinformatics and Applied Sciences, Indian Institute of Information Technology, Allahabad, Uttar Pradesh, India; Tel: 9236666060; E-mail: pritish@iiita.ac.in

1389-2029/19 $58.00+.00

Post-translational histone modifications are observed in many human cancers [9]. These modifications can be of many types. Some of the common Histone modifications include Acetylation, Methylation, Phosphorylation, Ubiquitylation and Sumoylation. miRNAs play very important cellular functions like regulation of mRNA activity [10]. They have proved to be early markers of cancers [11]. miRNA control the rate of expression of various enzymes involved in epigenetics. Though the mechanism is not fully understood, the role of miRNA is an established fact.

In spite of its potential of early biomarker detection, epigenetic drug discovery poses a number of challenges (Fig. **1**) like less number of biological and chemical tools and assays for probing and screening of chemical compounds [12]. This needs to be addressed before epigenetic drugs make their way to the patients.

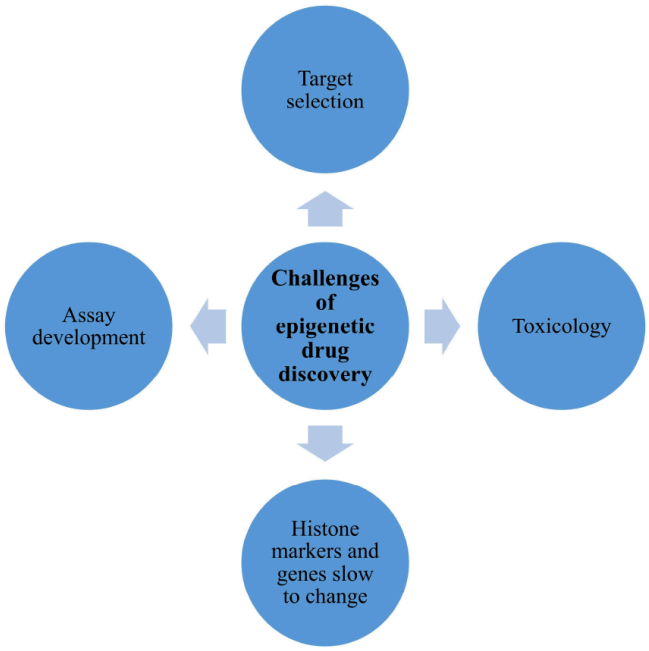


Fig. (1). Challenges in drug discovery of epigenetic drugs. It includes the lack of biological reference compounds for assay development. Currently, there is a lack of knowledge about both the short term as well as the long term repercussions of the epigenetic therapies (toxicology). The need of the hour is new models and longer duration to study tumor biology *in vivo*. Presently, limited epigenetic proteins and antibodies of high quality (target selection) are available. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

The mathematical discipline concerned with the study of complex networks is graph theory. A graph consists of points called vertices which are connected to each other *via* edges. This can be used to model many practical problems like biological systems [13]. Shifting from mainstream mathematics, it is now actively being used in fields like genomics, biological network analysis, electrical engineering, *etc*. The infinite scope of the combinatorial methods of graph theory leading to significant results in mathematics and other applications has been well-documented [14, 15]. In the systems biological approach, the analysis of the interaction between various components is very important as it provides a deeper insight into the overall working of pathway. A centrality measure quantifies the importance of vertices by allotting ranks [16]. Various centrality measures are there like closeness centrality, degree centrality, motif-based centralities, *etc*. Depending upon the type of network, the type of centrality measure is chosen. Network analysis provides a powerful tool in order to understand the structure, function and the evolutionary patterns involved in the biological process [17].

Almost 90% of liver cancers result from Hepatocellular carcinoma (HCC) [18]. Approximately, 7.5 Lakhs of new cases of HCC per year occurs globally which makes it the 5th common cause of cancers affecting human [18]. According to the report published by the International Agency of Research on Cancer (WHO), the male: female ratio for HCC in India is 4:1.

In this paper, we have merged the network of differentially expressed genes with the pathways most frequently affected in HCC to form an integrated network. This network was further validated, and topological parameters were calculated to quantify each node (Fig. **2**).

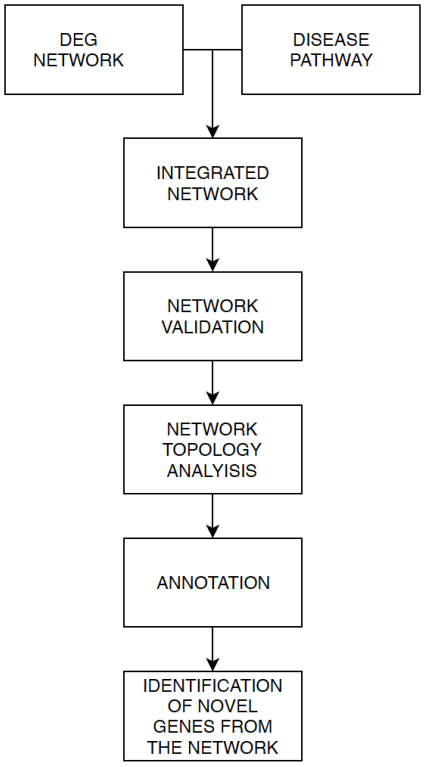


Fig. (2). The methodology implemented to identify novel targets.

2. METHODOLOGY

2.1. Data Mining

The initial data for epigenetic network construction was incorporated from two resources.

2.1.1. GEO Datasets

Expression profiles of GSE18081 [19], GSE37988 [20], GSE44970 [21], GSE54503 [22] and GSE57956 [23] were obtained from the GEO database [24]. GPL570 platform was used for the analysis. GSE18081 data represents CpG site methylation of HCV-cirrhotic, HCV-HCC and normal liver

tissues. GSE37988 represents the profile of methylation in Taiwanese HCC tumor and adjacent tissues. GSE44970 comprises genome-wide DNA methylation profiles of human hepatocellular carcinoma. GSE54503 consists of genome-wide DNA methylation profiles altered in hepatocellular carcinoma using Infinium Human Methylation 450 Bead Chips. GSE57956 is the methylation profile of hepatocellular carcinoma. All the datasets were checked for median centricity to see if the datasets chosen are comparable or not.

These datasets were basically micro-array data of Homo sapiens. The cutoff value of adjusted p-value <0.05 and |log FC| value > 1.5 was chosen as the benchmark. The DEGs were then used for the construction of PPI network from STRING [25].

2.1.2. Literature Search

There are various epigenetic markers involved in cancer epigenetics. One of the most critical is DNA methylation. It could be both hypomethylation and hypermethylation. We searched for various pathways that are most frequently affected by HCC, incorporated genes and merged it with the DEG network. Some of the pathways are Wnt-beta catenin, p53, hedgehog, *etc*. The PPI network of the differentially expressed genes was obtained from the STRING database [25].

2.2. Formation of an Integrated Protein-protein Interaction (PPI) Network

Both the PPI network constructed from DEGs and the pathways searched from literature, were imported in Cytoscape 3.5.1 [26]. These networks were merged using the merge tool implemented in Cytoscape 3.5.1 using the union operation.

2.3. Network Validation

Epigenomic network was validated by comparing it to random network and examining if it followed the power-law degree distribution. A comparison with random network was done using the Network Randomizer 1.1.2 [27] plugin in Cytoscape 3.5.1. A modular approach was followed so that it was easy to add additional random network models. The random networks may either be created by randomizing real networks implementing the edge shuffling algorithm or by generating new random networks by using the available models (we used the Barabasi-Albert model [28] as the networks generated by using this model are scale-free *i.e.* they follow power-law degree distribution).

2.4. Network Topology Analysis

The topology of the network was analyzed using NetworkAnalyzer 3.3.1 [29] and CentiScape 1.2.1 [30] plugin fully implemented in Cytoscape 3.5.1. Various topological features calculated were:

- ***Degree***: Number of connections emerging from a node. It is one of the most elementary characteristics of a node. It is expressed as:

$$D_i = \sum_{j=1}^{n} X_{ij}$$

Where, D_i is the degree of node D and X_{ij} is the adjacency matrix whose value is 1, if there exists an edge between the nodes i and j. If there is no edge between the two nodes then its value is zero. In our study, we have used this parameter to identify hubs and bottlenecks.

- ***Diameter*** [31] is the measure of the distance between the two most distant nodes. It is indicative of compactness of the network. A high value of diameter would suggest that the graph is not compact with respect to the nodes being considered. So, the value of diameter is a function of the two nodes that are chosen for distance calculation. Therefore, we see that a high value of diameter is not so much reliable as compared to a low value of diameter because such a value suggests that the graph is compact.
- ***Density:*** It has been observed that biological networks are mostly sparse thus preserving robustness. The property used to calculate how dense or sparse a network is called network density.
- ***Clustering Coefficient***, (C) [31] is the measure of the cluster formation tendency of the graph. A cluster can be visualized as a subset of the vertices of the entire graph having a large number of edges of a node having 'm' neighbors is calculated by the formula:

$$C = \frac{m(m-1)}{2}$$

- ***Neighborhood Connectivity*** [31] of a node is the average connectivity of all the neighbors where the connectivity of a single node is given by the number of neighbors of that node.
- ***Hub*** is a highly connected node and has a critical function in the network. Hubs are the nodes that have functional significance.
- ***Hub Bottlenecks*** are nodes having high degree and betweenness centrality values and are significant in scale-free biological networks. Therefore, these nodes are termed as Hub-Bottlenecks (HBNs) [32]. These nodes are likely to be essential to the system and often coincide with high degree hub nodes [33].
- ***Closeness Centrality*** is the reciprocal of the summation of shortest distances between a vertex and every other vertex present in the network [34].

It is calculated as,

$$CC_i = \frac{1}{\sum_J d(i,j)}$$

Where, d (i, j) is the shortest distance between node i and j.

- ***Stress Centrality***: For a node 'a', the stress centrality is the number of shortest paths passing through 'a'. Higher the value of stress centrality the greater is the importance of a node in holding various communicating nodes together. It can thus help us identify connecting proteins. A node with a high-stress centrality value can potentially be a good target.
- ***Topological Coefficient*** is a relative measure of the tendency of the nodes to share neighbors with other nodes.
- ***Eccentricity*** is the minimum distance between a node,

say b, and all other nodes. It is an index of the centrality of the node. In biological networks, nodes with a higher value of eccentricity are functionally accessible by other nodes of the network, and hence the changes in the concentration of the molecules are reflected readily. A protein having an eccentricity greater than the average eccentricity of the network has the capacity to influence other proteins.

2.5. Functional Annotation and Disease Gene Identification

KOBAS 3.0 [35] webserver was employed to ascertain the overexpressed genes and disease pathways in the network. In order to identify the statistically relevant hits, hypergeometric test was applied and p-value cut of ≤ 0.05 was kept as the threshold. KOBAS 3.0 is a web server for functional annotation of proteins or genes along with their functional enrichment. It integrates information from 9 gene set enrichment (GSE) methods. These methods can either be net-based or set-based. More than 4000 species and multiple pathways databases are supported by KOBAS 3.0. It supports both microarray and RNA-Seq data.

The results from KOBAS 3.0 helped in identifying the overexpressed pathways and genes. These pathways were superimposed on the integrated network. This helped in visualizing the genes involved in the pathogenesis of cancer both directly as well as indirectly.

2.6. Selection of the Novel Drug Targets from the Network

The approved drug targets in case of cancer epigenetics were searched and their interactions, as well as other topological characteristics in our network, were analyzed. The cytoHubba [36] plugin was utilized to rank the nodes and identify which nodes are capable of being classified as hubs. This plugin uses 11 topological methods among which we selected maximal clique centrality (MCC) owing to its better performance in terms of results and evaluation. The MCC of approved drug targets and rest of the network proteins were compared to identify potential drug targets.

For any node a, the MCC of a is defined as:

$$MCC(a) = \sum_{C \in S(a)} (|C|-1)!$$

Where,

$S(a)$ is the collection of maximal cliques which contain v, and $(|C|-1)!$ is the product of all positive integers less than $|C|$. If there is no edge between the neighbors of the node a, then $MCC(a)$ is the same as the degree [37].

3. RESULTS AND DISCUSSION

3.1. Cancer Epigenetics Network Structure

In order to develop and study HCC epigenetic pathway, a network consisting of all the DEG was developed. These DEGs were acquired from publicly available microarray study data of normal *versus* cancerous subjects. 250 DEGs were selected from each of the datasets *i.e.* GSE18081 [19], GSE37988 [20], GSE44970 [21], GSE54503 [22] and GSE57956 [23] expression profile using the cutoff of p-value < 0.05 and |log FC| value >1.5. Furthermore, these DEGs were integrated with pathways involved in regulation of HCC. PPI network of DEGs and those searched from literature was constructed using STRING database [25]. Both the PPI networks were merged using the merge tool of Cytoscape 3.5.1 to develop the final epigenetics network consisting of 1423 nodes and 5282 edges (Fig. **3**).

3.2. Network Validation

The network validation was conducted with the help of Network Randomizer plugin of Cytoscape 3.5.1. As compared to random network, cancer epigenetic networks have a higher value of average clustering coefficient depicting its modular nature. In our case, the average clustering coefficients for the random and epigenetic networks were 0.004 and 0.609 respectively. Besides, a higher value of average clustering coefficient also indicates redundancy and cohesiveness of the neighbors in the network. Random graphs have a small value of average clustering coefficient. Since proteins do not work in isolation, therefore, high cluster forming tendency of epigenetic network nodes indicates the biological significance of interactions between nodes. Secondly, the epigenetic network followed Power Law (Fig. **4**) with a degree exponent 'γ' value of -0.778 and R^2 value of 0.817. Small 'γ' value denotes essentiality of highly connected nodes in the network and R^2 value closer to 1, indicating strong correlation between network nodes [38].

3.3. Topological Characteristics of the Network

Network characteristics as calculated from Network Analyzer 3.1.1 and Centiscape 2.2 are given in Table **1**.

CytoHubba [36] plugin of Cytoscape 3.5.1, aided in identifying the central element of the cancer epigenetic network. Hubs are those proteins that interact with many partners and occupy the central region of a PPI network. HBNs are important linkers as they connect the sub-networks and have a high value of betweenness centrality. In our network, 25 and 13 nodes were identified as hubs and HBNs, respectively. Furthermore, the shortest path length distribution of the epigenetics network (Fig. **5A**) also showed that the network is scale-free. This metric is widely used for analysis of disease-associated genes as functionally related genes tend to remain closer to each other. Similarly, the shared neighbor distribution (Fig. **5B**) showed that the nodes in the network do not show the tendency of forming shared clusters rather every cluster is attached with each other by means of a single gene. These connecting genes are very important in the context of being considered as novel targets as these are crucial links between various pathways which eventually lead to epigenetic alterations [39].

3.4. Overrepresented Pathways in the Epigenetic Network

With corrected p-value <=0.05, thirty-five signaling pathways were found to be over-represented in our epigenetic network. These included PI3K-Akt signaling pathway, Thyroid signaling pathway, Wnt signaling pathway, p53 signaling pathway, FoXO signaling pathway, ErbB signaling pathway, Hippo signaling pathway, estrogen signaling,

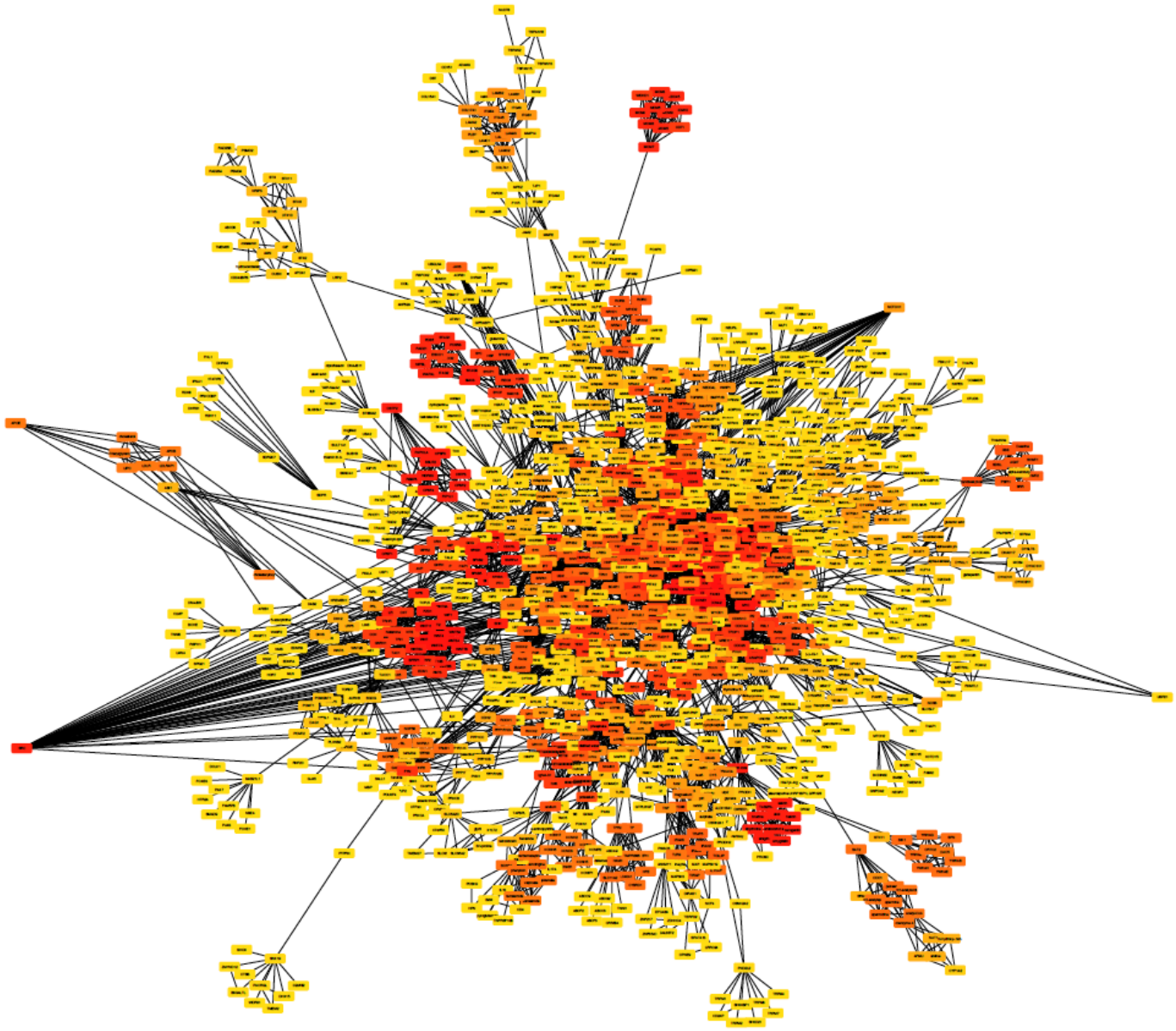


Fig. (3). The Epigenetic Network consisting of 1423 nodes and 5282 edges as obtained in Cytoscape. The vertices are the genes connected *via* edges. The variation in color of the nodes shows its essentiality with red being the most crucial followed by orange and yellow. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

AGE-RAGE signaling pathway, Ras signaling pathway, TGF-beta signaling pathway, mTOR signaling pathway, Neutrophil signaling pathway, Rap1 signaling pathway, HIF1 signaling pathway, Hedgehog signaling pathway, Notch signaling pathway, GnRH signaling pathway, T cell receptor signaling pathway, prolactin signaling pathway, MAPK signaling pathway, VEGF signaling pathway, Insulin signaling pathway, adipocytokine signaling, B-cell receptor signaling, JAK STAT signaling pathway, TNF signaling, NOD like receptor signaling, Oxytocin signaling pathway, Fc epsilon signaling, sphingolipid signaling pathway, phospholipase D signaling pathway, RIG-I like receptor signaling pathway, PPAR signaling pathway, NF-kappa B signaling pathway.

Among these overrepresented pathways, 32 pathways have already been reported for their role in HCC epigenetics. The novel ones identified are NOD like receptor signaling pathway, Fc epsilon signaling pathway and Hippo signaling pathway.

3.5. Identification of Potential Drug Targets for Cancer Epigenetics

The identification of potential drug targets was done on the basis of MCC. Therefore, the network nodes were ranked from the top 700 nodes which were selected based on their MCC score. From the list of 700 genes, we performed a rigorous literature survey to find out which of these genes have not yet been reported for their role in cancer epigenetics. We finally narrowed down our list to 10 genes based on their topological properties (Table **2**, Fig. **6**) which play a major role in HCC epigenetics and thus prove to be potential for epigenetic therapy. Furthermore, we also searched targets for

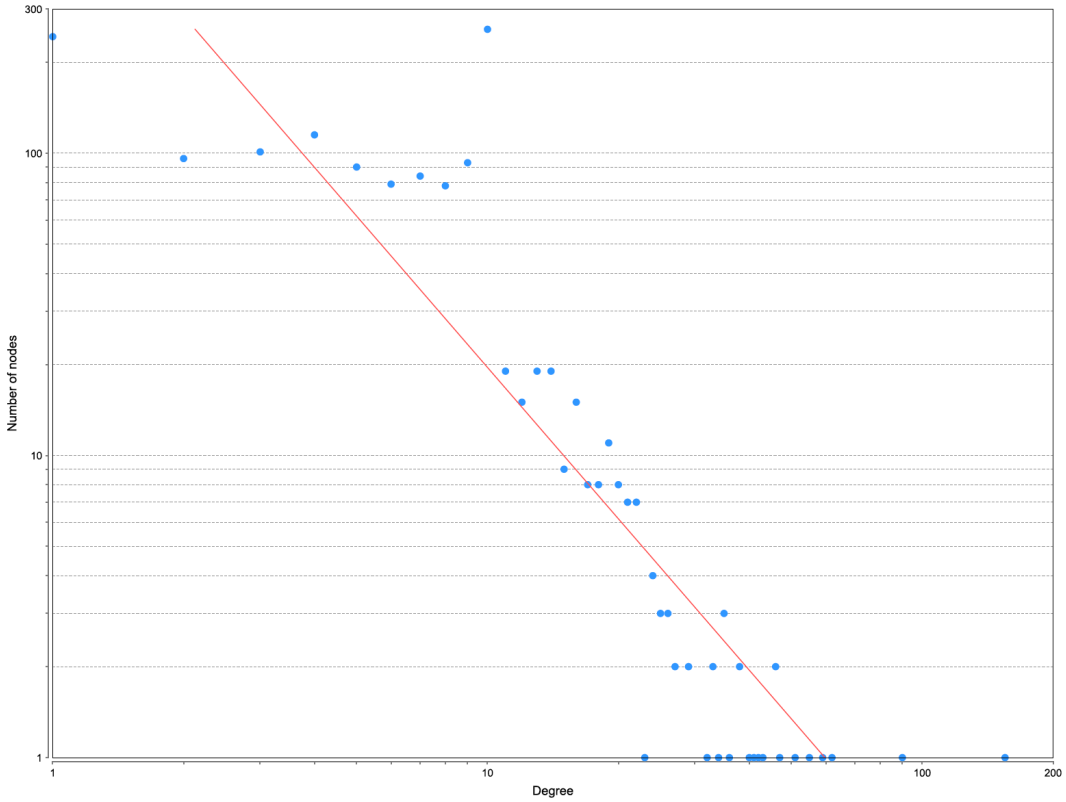


Fig. (4). Cancer epigenetics network follows Power Law with degree exponent 'γ' value of -0.778 and R^2 value of 0.817. The line shows power law fitting. Both the axes are logarithmic with base 10. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

investigational drugs at various stages of clinical trials used for cancer treatment and calculated their MCC values (Tables **3** and **4**) [40].

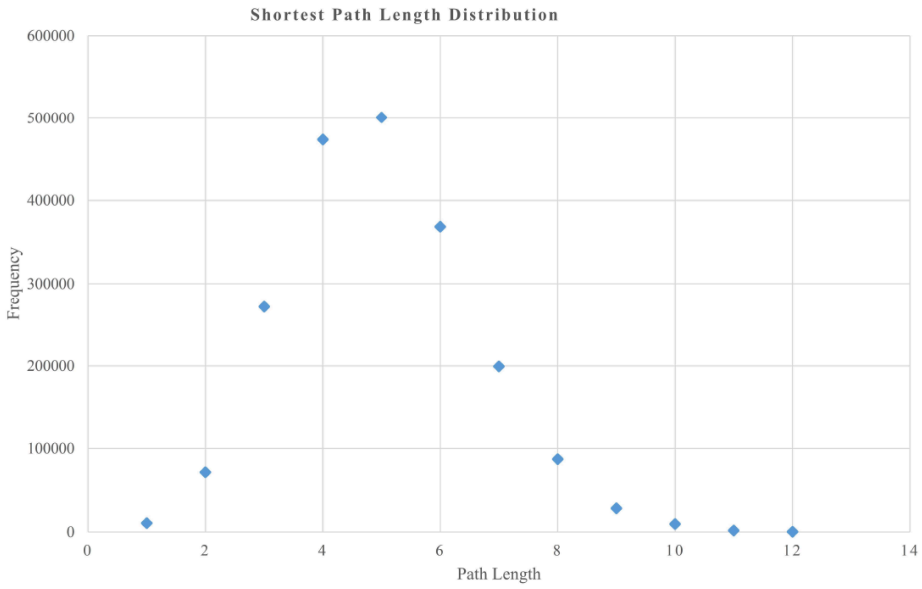
Table 1. Cancer epigenetics network characteristics.

Parameters	Value
Average clustering coefficient	0.614
Network diameter	12
Network radius	7
Shortest paths	2012142 (100%)
Network centralization	0.104
Network density	0.005
Characteristic path length	4.988

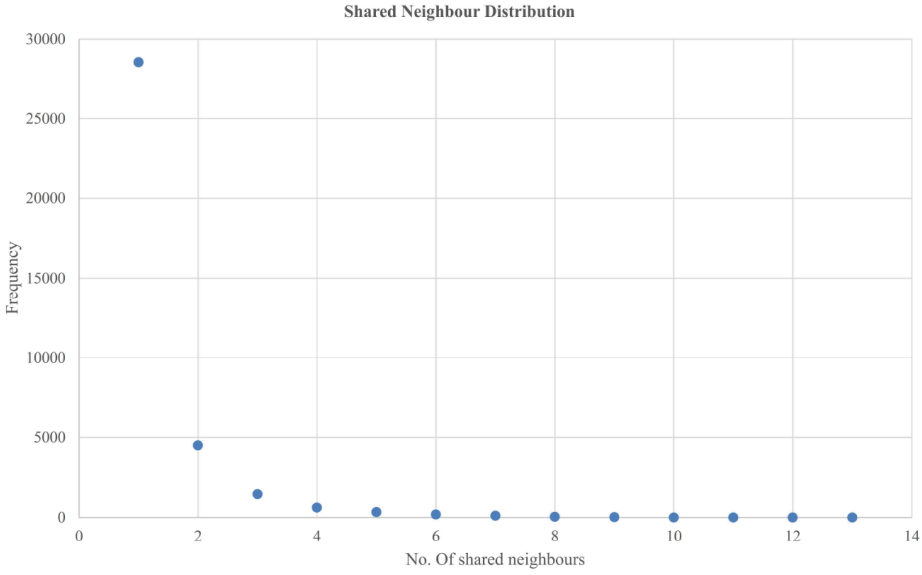
The comparison between MCC values of approved and potential drug targets revealed that the novel targets *cdk6, cdk4, cdkn2a, smad3, smad7, ccnd1, e2f1, sf3b1, ctnnb1* and *tgfb1* have MCC score greater than the approved drug targets. Comparison of Tables **2** and **3** showed that our potential drug targets had features similar to that of approved drug targets. The value of eccentricity, which measures the easiness of a protein to influence the activities of several other proteins was greater than 1 for both approved and potential drug targets. Furthermore, both approved and potential drug targets had almost similar degree except *cdkn2a* that had a degree of 21. Similarly, the clustering coefficient and radiality of both approved and potential drug targets were close to 1. The radiality value is a measure of functional relevance of a protein for several other proteins but with the possibility of functionally irrelevant for few other proteins. The value of topological coefficient that describe the tendency of a node to have shared neighbors was close to 1 for approved drug targets, while our potential drug targets had much smaller values for topological coefficient.

The present clinical significance of the genes reported here are:

- *cdk6* activates cell proliferation. Overexpression of *cdk6* is associated with resistance to hormone therapy in breast cancer. Upregulation of *cdk6* has been observed in almost one-third of medulloblastoma cases. Changes in *cdk6* can effect various hallmarks of cancer like induction of angiogenesis and evasion of growth suppressors *etc*.
- *cdk4* encodes a protein which is a member of the Ser/Thr protein kinase family. Mutations in this gene along with some closely related proteins have shown association with the tumorigenesis in variety of cancers. Multiple polyadenylation sites of this gene have been reported. Cyclin D regulates its functioning.
- *cdkn2a* provides instructions for synthesis of crucial proteins like p16(INK4A) and p14(ARF) which play an important role in tumor suppression. In older cells, these help in stopping cell division and *p14* helps stop the breakdown of *p53*.



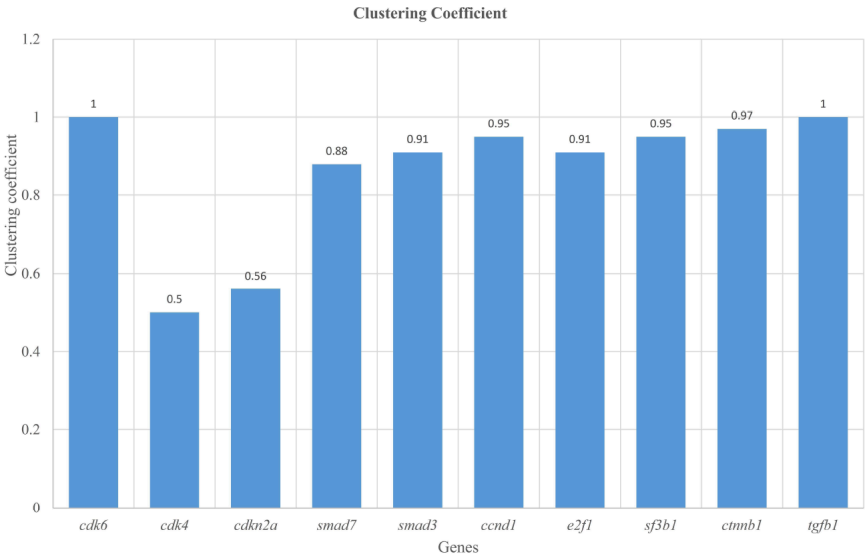
(A)



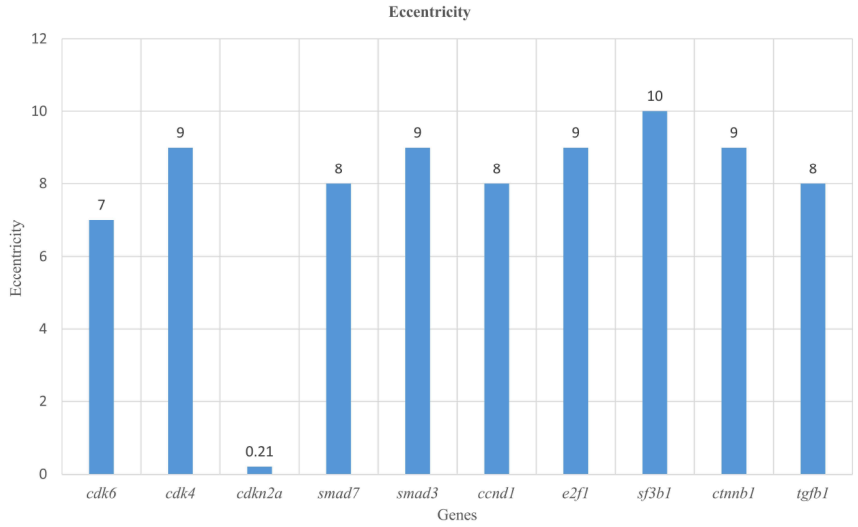
(B)

Fig. (5). (**A**) Analysis of the distribution of various path lengths of the integrated epigenetic network (left). The shortest path length distribution depicts that it is a small world network. (**B**) Distribution of shared neighbors. Both x and y axis are in logarithmic scale. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

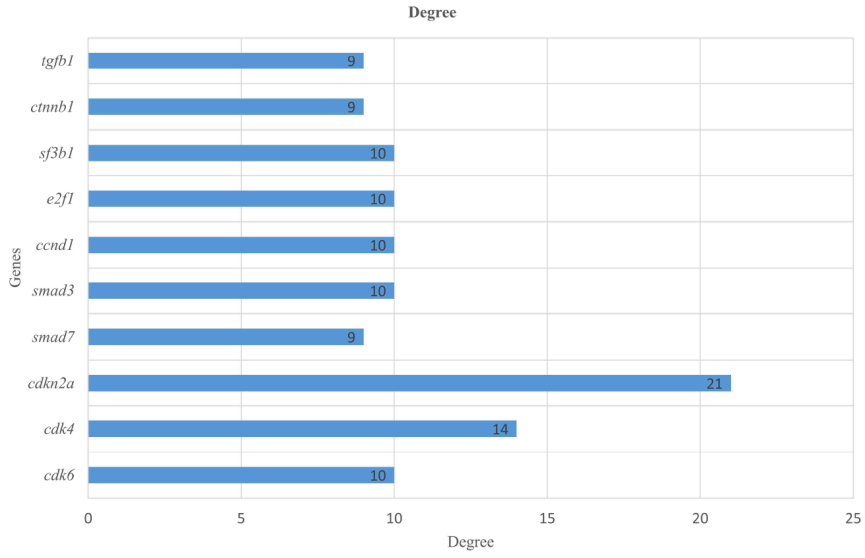
- *smad7* exerts an inhibitory effect on the EGF signaling pathway which is involved in breast cancer and ovarian cancer invasion and metastasis. A mutation located in *smad7* gene is a cause of susceptibility to colorectal cancer (CRC) type 3. It is also overexpressed in pancreatic cancer.
- *smad3* plays an important role in the regulation of genes involved in growth differentiation and death implying that an alteration in its activity or repressing its activity can lead to the formation or development of cancer. It has an important role in colorectal, breast and pancreatic cancer.
- *ccnd1* codes for the protein cyclin D1. The overexpression of this protein is strongly correlated to ER+ breast cancer and deregulation of cyclin D1 is associated with hormone therapy resistance in breast cancer.
- *e2f1*encodes protein that play a crucial role in the controlling cell cycle and tumor suppressor protein. Elevated expression of *E2F1* protein in breast cancer cell lines and head and neck carcinoma cell lines, and overexpression of *E2F1* in invasive ductal breast cancer and non-small cell lung cancer.
- *sf3b1* gene mutations have been recurrently seen in cases of advanced chronic lymphocytic leukemia, myelodysplastic syndromes and breast cancer. *sf3b1* is one of several genes involved in RNA splicing that has been identified as recurrently mutated in MDS and other malignancies.
- *ctnnb1* gene encodes for the protein β-catenin. This protein has dual function of managing transcription and intercellular adhesion. Mutations and overexpression of this protein has been found to be associated with many cancers like endometrial cancer, ovarian cancer and colorectal cancer.



(A)

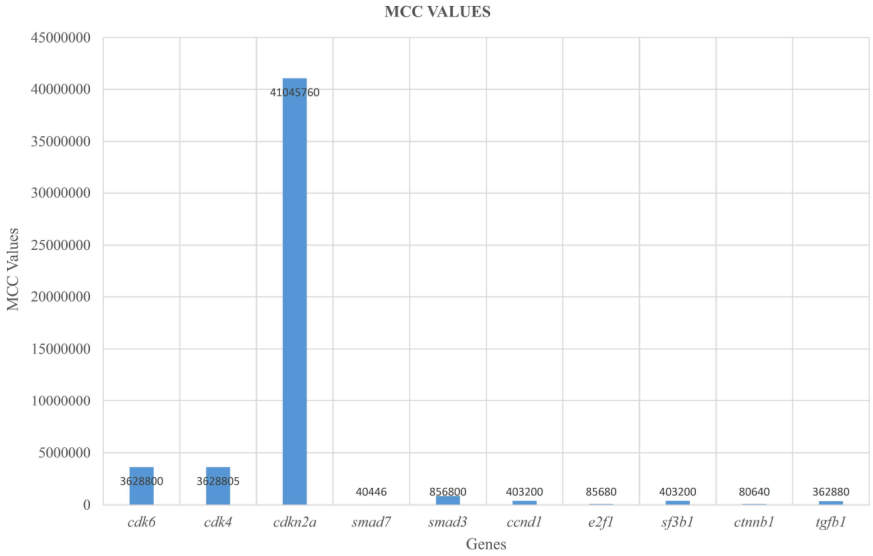


(B)



(C)

Fig. (6) contd….



(D)

Fig. (6). Characteristics of potential drug targets (**A**) Clustering coefficient. It is clear that *cdk6 and tgfb1* show the highest value of clustering coefficients. This is a very important topological property as it provides information about the cluster forming tendency of a node. (**B**) Eccentricity. *sf3b1* exhibits the maximum eccentricity indicating its accessibility to other nodes and *cdk6* has the minimum eccentricity value indicating its marginal functional role. (**C**) Degree. (**D**) MCC. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

Table 2. Novel targets for cancer epigenetic therapy involved in methylation.

S. No.	Gene	Degree	Radiality	Clustering Coefficient	Eccentricity	Topological Coefficient	MCC Value
1.	*cdk6*	10	0.79	1.0	7	0.22	3628800
2.	*cdk4*	14	0.73	0.5	9	0.27	3628805
3.	*cdkn2a*	21	3.21	0.560	0.21	0.00	41045760
4.	*smad7*	9	0.76	0.88	8	0.18	40446
5.	*smad3*	10	0.72	0.91	9	0.29	856800
6.	*ccnd1*	10	0.72	0.95	8	0.44	403200
7.	*e2f1*	10	0.72	0.91	9	0.29	85680
8.	*sf3b1*	10	0.63	0.95	10	0.69	403200
9.	*ctnnb1*	9	0.70	0.97	9	0.33	80640
10.	*tgfb1*	9	0.72	1.0	8	0.44	362880

Table 3. MCC values of approved drug targets for HCC.

S. No.	Drug Targets	Degree	Radiality	Clustering Coefficient	Eccentricity	Topological Coefficient	MCC Value
1.	HDAC1	10	1	0.89	1	0.86	21840
2.	HDAC3	10	1	0.75	1	0.78	11550
3.	HDAC6	10	1	0.77	1	0.8	6480
4.	MET	10	1	0.822	1	0.84	130
5.	VEGFR	10	1	0.26	1	0.36	31

- *tgfb1* gene encodes for multifunctional peptide set TGF-β. It acts negative autocrine growth factor and controls the immune system response. Besides, it performs functions like cell growth, differentiation and apoptosis.

Table 4. List of HCC investigational drugs with their targets.

Drug	Target	Status
Resminostat	HDAC1, HDAC3 and HDAC6	Phase II
Tivantinib	MET/Tubulin	Phase III
Axitinib	VEGFR	Phase II
Apatinib	VEGFR	Phase III

CONCLUSION

Graph theory has proven to be a promising tool in understanding various biological networks in the past. In the current work, this powerful tool implemented by means of Cytoscape 3.5.1 was analyzed to understand HCC epigenetic pathways. The network approach provides a deeper insight into the interactions taking place.

Epigenetic therapy is still in its infancy with only a few epigenetic drugs available. The novel targets reported here have the potential for opening new avenues in HCC therapy. Some of the reported targets are not directly involved in the pathways yet they play an indispensable role. One of the major reasons for interest in developing drugs based on epigenetics (epidrugs) is the reversible nature of epigenetic changes.

The major challenges epigenetic drug research is facing today and likely in the future prominently include interpreting the relevance of pharmacodynamics and somatic mutation level in the patient to hone the response to these agents.

Epigenetic alterations can lead to various other diseases like cardiovascular disorders, metabolic disorders, and neurological disorders. Since epigenetic changes are reversible, a better understanding of these mechanisms can be utilized to regulate the genetic switch and bring back the cell to the normal state.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of the article is available in the [GEO] at [URL: https://www.ncbi.nlm.nih.gov/geo/], reference number [Manuscript Ref:19,20,21,22,23,24].

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Waddington, C.H. The epigenotype. *Endeavour,* **1942**, *1*, 18-20.
[2] Egger, G.; Liang, G.; Aparicio, A.; Jones, P.A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature,* **2004**, *429*(6990), 457-463. [http://dx.doi.org/10.1038/nature02625] [PMID: 15164071]
[3] Jones, P.A.; Baylin, S.B. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.,* **2002**, *3*(6), 415-428. [http://dx.doi.org/10.1038/nrg816] [PMID: 12042769]
[4] Miranda, T.B.; Jones, P.A. DNA methylation: the nuts and bolts of repression. *J. Cell. Physiol.,* **2007**, *213*(2), 384-390. [http://dx.doi.org/10.1002/jcp.21224] [PMID: 17708532]
[5] Jones, P.A.; Laird, P.W. Cancer epigenetics comes of age. *Nat. Genet.,* **1999**, *21*(2), 163-167. [http://dx.doi.org/10.1038/5947] [PMID: 9988266]
[6] Feinberg, A.P.; Ohlsson, R.; Henikoff, S. The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.,* **2006**, *7*(1), 21-33. [http://dx.doi.org/10.1038/nrg1748] [PMID: 16369569]
[7] Jones, P.A.; Martienssen, R. A blueprint for a human epigenome project: the AACR human epigenome workshop. *Cancer Res.,* **2005**, *65*(24), 11241-11246. [http://dx.doi.org/10.1158/0008-5472.CAN-05-3865] [PMID: 16357125]
[8] Raj, U.; Kumar, H.; Gupta, S.; Varadwaj, P.K. Novel DOT1L receptor natural inhibitors involved in mixed lineage leukemia: a virtual screening, molecular docking and dynamics simulation study. *Asian Pac. J. Cancer Prev.,* **2015**, *16*(9), 3817-3825. [http://dx.doi.org/10.7314/APJCP.2015.16.9.3817] [PMID: 25987043]
[9] Raj, U.; Varadwaj, P.K.. Epigenetics and its role in human cancer. In: Translational Bioinformatics and Its Application. Translational Medicine Research; Wei DQ., Ma Y., Cho W., Xu Q., Zhou F. (eds).; Springer, Dordrecht, pp. 249-267. [http://dx.doi.org/10.1007/978-94-024-1045-7_11]
[10] He, L.; Hannon, G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.,* **2004**, *5*(7), 522-531. [http://dx.doi.org/10.1038/nrg1379] [PMID: 15211354]
[11] Lu, J.; Getz, G.; Miska, E.A.; Alvarez-Saavedra, E.; Lamb, J.; Peck, D. MicroRNA expression profiles classify human cancers. *Nature,* **2005**, *435*(7043), 834-838.
[12] Campbell, R.M.; Tummino, P.J. Cancer epigenetics drug discovery and development: the challenge of hitting the mark. *J. Clin. Invest.,* **2014**, *124*(1), 64-69. [http://dx.doi.org/10.1172/JCI71605] [PMID: 24382391]
[13] Huber, W.; Carey, V.J.; Long, L.; Falcon, S.; Gentleman, R. Graphs in molecular biology. *BMC Bioinformatics,* **2007**, *8*(6)(Suppl. 6), S8. [http://dx.doi.org/10.1186/1471-2105-8-S6-S8] [PMID: 17903289]
[14] Lesne, A. Complex networks: from graph theory to biology. *Lett. Math. Phys.,* **2006**, *78*(3), 235-262. [http://dx.doi.org/10.1007/s11005-006-0123-1]
[15] Mason, O.; Verwoerd, M. Graph theory and networks in Biology. *IET Syst. Biol.,* **2007**, *1*(2), 89-119. [http://dx.doi.org/10.1049/iet-syb:20060038] [PMID: 17441552]
[16] Hahn, M.W.; Kern, A.D. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.,* **2005**, *22*(4), 803-806. [http://dx.doi.org/10.1093/molbev/msi072] [PMID: 15616139]
[17] Ma'ayan, A. Introduction to network analysis in systems biology. *Sci. Signal.,* **2011**, *4*(190), tr5. [http://dx.doi.org/10.1126/scisignal.2001965] [PMID: 21917719]

[18] EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J. Hepatol.,* **2012**, *56*(4), 908-943. [http://dx.doi.org/10.1016/j.jhep.2011.12.001] [PMID: 22424438]

[19] Archer, K.J.; Mas, V.R.; Maluf, D.G.; Fisher, R.A. High-throughput assessment of CpG site methylation for distinguishing between HCV-cirrhosis and HCV-associated hepatocellular carcinoma. *Mol. Genet. Genomics,* **2010**, *283*(4), 341-349. [http://dx.doi.org/10.1007/s00438-010-0522-y] [PMID: 20165882]

[20] Shen, J.; Wang, S.; Zhang, Y.J.; Kappil, M.; Wu, H.C.; Kibriya, M.G.; Wang, Q.; Jasmine, F.; Ahsan, H.; Lee, P.H.; Yu, M.W.; Chen, C.J.; Santella, R.M. Genome-wide DNA methylation profiles in hepatocellular carcinoma. *Hepatology,* **2012**, *55*(6), 1799-1808. [http://dx.doi.org/10.1002/hep.25569] [PMID: 22234943]

[21] Revill, K; Wang, T; Lachenmayer, A; Kojima, K. Genome-wide methylation analysis and epigenetic unmasking identify tumor suppressor genes in hepatocellular carcinoma. *Gastroenterology,* **2013**, *145*(6), 1424-1435. [http://dx.doi.org/10.1053/j.gastro.2013.08.055]

[22] Shen, J.; Wang, S.; Zhang, Y.J.; Wu, H.C.; Kibriya, M.G.; Jasmine, F.; Ahsan, H.; Wu, D.P.; Siegel, A.B.; Remotti, H.; Santella, R.M. Exploring genome-wide DNA methylation profiles altered in hepatocellular carcinoma using Infinium HumanMethylation 450 BeadChips. *Epigenetics,* **2013**, *8*(1), 34-43. [http://dx.doi.org/10.4161/epi.23062] [PMID: 23208076]

[23] Mah, W.C.; Thurnherr, T.; Chow, P.K.; Chung, A.Y.; Ooi L.L.; Toh, H.C.; The, B.T.; Sauntharajah, Y.; Lee, C.G. Methylation profiles reveal distinct subgroup of hepatocellular carcinoma patients with poor prognosis. *PLoS One*. **2014**, *9*(8), e104158. [http://dx.doi.org/10.1371/journal.pone.0104158]

[24] Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.,* **2002**, *30*(1), 207-210. [http://dx.doi.org/10.1093/nar/30.1.207] [PMID: 11752295]

[25] Jensen, L.J.; Kuhn, M.; Stark, M.; Chaffron, S.; Creevey, C.; Muller, J.; Doerks, T.; Julien, P.; Roth, A.; Simonovic, M.; Bork, P.; von Mering, C. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.,* **2009**, *37*(Database issue), D412-D416. [http://dx.doi.org/10.1093/nar/gkn760] [PMID: 18940858]

[26] Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.,* **2003**, *13*(11), 2498-2504. [http://dx.doi.org/10.1101/gr.1239303] [PMID: 14597658]

[27] Tosadori, G.; Bestvina, I.; Spoto, F.; Laudanna, C.; Scardoni, G. Creating, generating and comparing random network models with NetworkRandomizer. *F1000 Res.,* **2017**, 5, 2524.

[28] Barabasi, A.L.; Albert, R. Emergence of scaling in random networks. *Science,* **1999**, *286*(5439), 509-512.

[29] Assenov, Y.; Ramírez, F.; Schelhorn, S.E.; Lengauer, T.; Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics,* **2008**, *24*(2), 282-284. [http://dx.doi.org/10.1093/bioinformatics/btm554] [PMID: 18006545]

[30] Scardoni, G.; Tosadori, G.; Faizan, M.; Spoto, F.; Fabbri, F.; Laudanna, C. Biological network analysis with CentiScaPe: centralities and experimental dataset integration. *F1000 Res.,* **2014**, *3*, 139. http://dx.doi.org/10.12688/f1000research.4477.1 PMID: 26594322

[31] Lima-Mendez, G.; van Helden, J. The powerful law of the power law and other myths in network biology. *Mol. Biosyst.,* **2009**, *5*(12), 1482-1493. [http://dx.doi.org/10.1039/b908681a] [PMID: 20023717]

[32] Maslov, S.; Sneppen, K. Specificity and stability in topology of protein networks. *Science,* **2002**, *296*(5569), 910-913. [http://dx.doi.org/10.1126/science.1065103] [PMID: 11988575]

[33] Barabási, A.L.; Gulbahce, N.; Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.,* **2011**, *12*(1), 56-68. [http://dx.doi.org/10.1038/nrg2918] [PMID: 21164525]

[34] Yu, H.; Kim, P.M.; Sprecher, E.; Trifonov, V.; Gerstein, M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLOS Comput. Biol.,* **2007**, *3*(4), e59. [http://dx.doi.org/10.1371/journal.pcbi.0030059] [PMID: 17447836]

[35] Newman, M.E. A measure of betweenness centrality based on random walks. *Soc. Networks,* **2005**, *27*(1), 39-54. [http://dx.doi.org/10.1016/j.socnet.2004.11.009]

[36] Wu, J.; Mao, X.; Cai, T.; Luo, J.; Wei, L. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.,* **2006**, *34*(suppl_2), W720-W724.

[37] Chen, S.H.; Chin, C.H.; Wu, H.H.; Ho, C.W.; Ko, M.T.; Lin, C.Y. **2009**. cyto-Hubba: a Cytoscape plug-in for hub object analysis in network biology. *20th International Conference on Genome Informatics.*

[38] Sengupta, U.; Ukil, S.; Dimitrova, N.; Agrawal, S. Expression-based network biology identifies alteration in key regulatory pathways of type 2 diabetes and associated risk/complications. *PLoS One,* **2009**, *4*(12), e8100. http://dx.doi.org/10.1371/journal.pone.0008100 PMID: 19997558

[39] Jones, P.A.; Issa, J.P.J.; Baylin, S. Targeting the cancer epigenome for therapy. *Nat. Rev. Genet.,* **2016**, *17*(10), 630-641. [http://dx.doi.org/10.1038/nrg.2016.93] [PMID: 27629931]

[40] Thillai, K.; Ross, P.; Sarker, D. Molecularly targeted therapy for advanced hepatocellular carcinoma - a drug development crisis? *World J. Gastrointest. Oncol.,* **2016**, *8*(2), 173-185. [http://dx.doi.org/10.4251/wjgo.v8.i2.173] [PMID: 26909132]