IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 3 Issue 1, January 2016.
www.ijiset.com
ISSN 2348 – 7968

Target Genes and Core Genes of Hepatocellular Carcinoma Using Gene Force Algorithm and Gene Community Network

Jinyu Hu[1,2],Bin Hu[1,2],Weisen Pan[3],Junsong Wang[4],Xuemei Zhou[5] and Zhiwei Gao[6]

[1] Division of Immunology and Rheumatology, Department of Medicine, Stanford University
Stanford, CA, USA 94305
Email: jinyu3@stanford.edu
[2] Department of Medicine, Veterans Affairs Palo Alto Health Care System
Palo Alto, CA, USA 94306
Email: bhu616@stanford.edu
[3] Department of Computer Science and Engineering, The Ohio State University
Columbus, OH, USA 43210
Email: pan.548@osu.edu
[4]School of Biomedical Engineering, Tianjin Medical University
Tianjin 300072, China
Email: wjsong2004@126.com
[5]College of Math and Computer, Jiangxi Science & Technology Normal University
Nanchang 330038, China
zhouxue1107@hotmail.com
[6]School of Computing, Engineering and Information Sciences,Northumbria University
Newcastle upon Tyne, NE1 8ST, United Kingdom
zhiwei.gao@northumbria.ac.uk

Abstract

From the viewpoint of gene data streams, gene expression data of hepatocellular carcinoma are investigated to screen target genes and core genes, which are employed to propose a new strategy for the treatment of hepatocellular carcinoma. New concepts such as gene data streams, gene characteristic strength, gene impact factor and gene force are proposed. Together with gene community network, a novel algorithm, that is, called GF algorithm, is presented to screen feature genes, target genes and score genes. And the fifteen HCC target genes have been screened, which are further used to find three core genes including HAMP, MTs and GPC3.According to the relationship between the three core genes such as HAMP, MTs and GPC3 and the metals including copper, iron and zinc, a treatment strategy for HCC is proposed, namely, "Supplement Iron and Zinc after surgery" for HCC patients. The proposed treatment method can be used to regulate the expression levels of HCC core genes .

Keywords:*Gene Data Stream, Gene Force, Gene Community Network, target genes,Core Genes.*

1. Introduction

Hepatocellular carcinoma (HCC), with poor survival rates unless recognized and treated early, is ranked as the fifth most common cancer worldwide and the third biggest cause of cancer death[1]. Data stream has been widely used in clinical treatments such as medical imaging[2]and critical care [3], unfortunately, which has been less applied to gene researches such as gene data updating[4].The traditional feature genes approaches only consider the differences of the gene expressions between tumor and normal tissues[5]; however, the degrees of the differences are not taken into account. The intensity of the correlations between genes has not been clearly defined and interpreted. The absolute values of Pearson's correlation coefficients are usually used to assess the correlation of genes[6], which fail to distinguish the positive correlation from the negative correlation. According the graph theory, we can build networks to make Community Detection for genes [7-10].Target genes are usually obtained by using biological experiments[11], but it is noted that the experiments are difficult to implement and the coverage is often small. There seems not to be an agreed opinion on the reason caused tumor. One opinion is that oncogenes being active and anti-oncogenes being inactive will cause tumor[12]. Another one emphasizes that the mutation of the oncogenes and anti-oncogenes will lead to cancer[14]. Recently, gene therapy has been applied in gene replacement, gene correction, gene augmentation, and gene inactivation[13]. Unfortunately, the conventional gene therapy methods[15] seem to be difficult to implement and their effectiveness has yet proved significantly.

As a result, there is a strong motivation to investigate gene therapy from the viewpoint of data stream, providing a new idea for the treatment of HCC. The core gene therapy has been investigated based on HCC genes data streams.

ISSN 2348 – 7968

This innovation has been reflected in the methods and results. The novelty of the proposed research methods is summarized as follows: New concepts such as gene data streams, gene characteristic strength (CS), gene impact factor (GIF) and gene force (GF) are proposed. A novel algorithm, that is, called GF algorithm, is presented to screen feature genes, and target genes. The gene community network (GCN) proposed[16-19] can be divided into gene positive network (GPN) and gene negative network (GNN). According to the relationship between the three core genes such as HAMP, MTs and GPC3 and the metals including copper, iron and zinc, a treatment strategy for HCC is proposed, namely, "Supplement Iron and Zinc after surgery" for HCC patients. The proposed treatment method can be used to regulate the expression levels of HCC core genes.

2. Basic Concepts

2.1 Gene data stream

Gene data stream is formed by a series of data updating. From the perspective of the update, the gene data stream can be divided into horizontal data stream and vertical data stream, respectively.

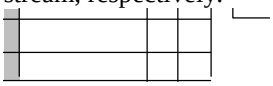


Figure 1. Vertical and Horizontal update data update

2.2 Gene Force(GF)

A GF is determined by both the total impacts between the characteristics strength (CS) and the gene impact factor (GIF). The greater the gene force is, it is more likely for this gene to become a target gene.

$$CS = |\, mean(Tumor) - mean(Normal) \,| \quad (1)$$

where mean(Tumor) and mean(Normal) denote the average of the gene expression levels in tumor, and the average of the gene expression levels in normal tissues, respectively

$$GIF = \sum |\, p_{ij} \,|, \qquad |\, p_{ij} \,| > 0.2 \quad (2)$$

where p_{ij} represents the Pearson correlation between gene i and gene j.

The GF is defined as the multiplication of the CS and GIF, that is,

$$GF = CS * GIF \quad (3)$$

2.3 Gene community network (GCN)

It is used to describe the relationship between genes. A connection matrix A is used to store gene community networks (GCN),which is divide into two types: GPN and GNN.

2.3.1Gene positive network (GPN)

Extract a network from the GCN where the values of all the edges in the network are greater than 0, that is,

$$a_{ij} = \begin{cases} p_{ij} & i > j \ \ and \ \ p_{ij} > 0 \\ 0 & i \le j \end{cases} \quad (4)$$

2.3.2Gene negative network (GNN)

Remove the edges with the weights greater than 0 to form a network.

$$a_{ij} = \begin{cases} p_{ij} & i > j \ \ and \ \ p_{ij} < 0 \\ 0 & i \le j \end{cases} \quad (5)$$

where r_{ij} is the Pearson correlation coefficient of the nodes i and j, and $r_{ij} \in [-1,1]$.

3. Gene Force Algorithm

Now it is ready to address the main method of the present paper, that is, GF algorithm. The flow chart of the algorithm is described by Figure 2.

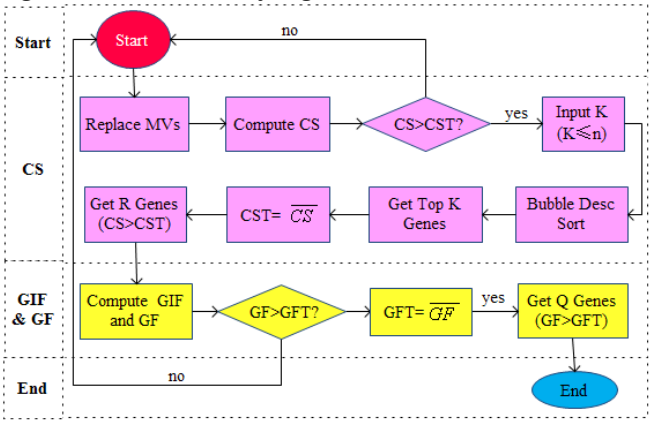


Figure 2. The flowchart of gene force (GF) algorithm

ISSN 2348 – 7968

The procedure of the gene force (GF) algorithm can be concluded as follows.
(1)Initialization: Initial gene expression data are from *m* tissues and *n* genes.
(2)Replacing missing values (MVs): The missing data are replaced by the mean of the other data of this gene in tumor issues or normal issues. Note that the missing gene expression data of the tumor tissues and normal tissue are replaced separately.
(3)Calculate the CS values of all the genes.
(4)Determine if CS values are greater than CS threshold (CST). Initially CST=0.
(5)Input the parameter K (K≤n), where K is the number of candidate feature genes.
(6)Sort the CS values in descending order to give the top K genes by using bubble sort algorithm.
(7)Update the CS threshold (CST) by selecting $CST = \overline{CS}$, where $\overline{CS}$ is the mean of the CS values.
(8)Produce R (R≤K) genes with the CS values larger than the CS threshold (CST).
(9)Calculate the Pearson correlation strength (PF), gene impact factors (GIF) and the gene forces (GF) of the R genes.
(10)Update the GF threshold (GFT) by selecting $GFT = \overline{GF}$, where $\overline{GF}$ denotes the mean of the gene forces (GF).
(11)Obtain Q (Q≤R) genes with their GF values larger than the GF threshold (GFT). The Q genes are the target genes.

4. Results and Discussions

4.1 Data

The liver cancer microarray data are taken from Stamford genetic database, which is available at http://genome-www.stanford.edu/hcc/supplement.shtml. The 3964 genes are differentially expression in 156 liver tissues (74 non-tumor liver tissues and 82 HCC tissues).

4.2 Screen the set of target genes

The GF algorithm is presented to screen feature genes, target genes. The 15 target genes are sorted in descending order according to the GF values(see table I). If the transcribedlocus and the duplication of genes are extracted, the set of the target genes can be given by TGS= {HAMP, RNAHP, MT1H, MT1G, GPC3, MT1E, MT1L, AQP4, VIPR1, DNASE1L3, MT1B}.

Table 1:Fifteen Target Genes

No.	*GName*	*CS*	*GIF*	*GF*
41	HAMP	5.28	21.16	111.7
51	RNAHP	3.78	22.22	83.94
49	MT1H	3.73	22.48	83.77
52	MT1G	3.67	21.87	80.17
50	RNAHP	3.48	22.68	78.89
47	MT1L	3.41	21.99	74.99
40	Trans	3.34	21.76	72.61
42	AQP4	3.43	20.53	70.52
39	Trans	3.24	21.52	69.76
12	GPC3	4.19	15.88	66.6
53	MT1E	3.16	20.81	65.77
38	Trans	3.08	21.21	65.28
58	VIPR1	2.96	22.02	65.23
55	DNASE1L3	2.98	21.03	62.6
48	MT1B	2.85	21.88	62.39

No.: the order number; GName: gene name; CS: characteristic strength; Red: oncogenes; Green: suppressor genes; Trans: Transcribedlocus; Genes with the same name are isomers.

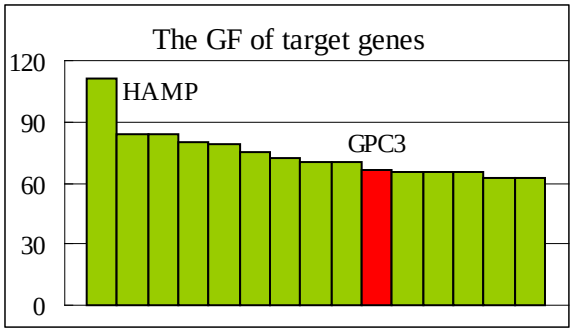
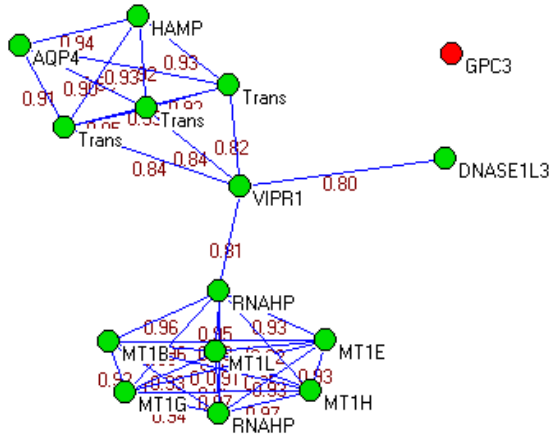


Figure 3. The GF of target genes. The HAMP gene has the biggest GF values, that impacts the gene is the very important in HCC. And GPC3 is the only one oncogen of the target genes.

4.3 Seek core genes

On the basis of the 15 target genes, we can construct the gene community network (GCN) in order to find the most important core genes, which is useful for providing treatment strategy. When the threshold value of the correlation is selected as 0.8, one can establish a POS network in Figure 4, which describes the correlation of the 15 target genes.



ISSN 2348 – 7968

Figure 4. Strongly related POS network with the threshold value at T=0.8. The weights of the edges are correlation coefficients. The tumor suppressor genes denoted by blue nodes in the connected graph have proved that they are strongly related.

Setting the threshold value at T=-0.55, one can an updated NEG network exhibited by Figure 5. In this case, all the genes in the MT family are related to the GPC3, indicating a relatively strong inhibition interaction between the GPC3 and the genes in the MT family. In addition, VIPR1 is the only gene within the HAMP family, which has a relatively strong inhibition correlation with the GPC3.

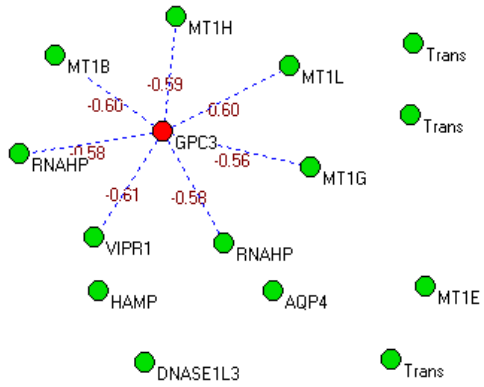


Figure 5. Strong moderate correlated NEG network at T=-0.55, which shows a relatively strong inhibition between gene GPC3 and MTs cluster.

According the analysis above, the fifteen target genes are obtained, which can be divided into three clustering sets including HAMP Cluster ={ HAMP, Trans, AQP4, VIPR1}, MT Cluster ={MT1H, MT1B, MT1G, MT1E, MTIL, RNAHP, DNASE1L3} and GPC3 Cluster ={GPC3}. The core genes of each clusters are then derived, that is, HAMP, MTs and GPC3 respectively, where MTs is a general name for a group of metallothionein genes including MT1H, MT1B, MT1G, MT1E, and MTIL.

5. HCC Treatments

Before presenting the treatment strategy for HCC, we would like to revisit the properties of the core genes such as HMAP, MTs, and GP3.

The main purpose of the HAMP gene is to maintain iron homeostasis. Specifically, when the iron in the blood serum is found insufficient, the iron in liver would be released. While the iron in the blood serum is beyond the normal level, it is necessary for the regulation of iron storage in macrophages, and for intestinal iron absorption [20]. Iron overload is associated with HCC in people [21]. It is noted that HAMP and AQP4 are extremely strongly related. Therefore, AQP4 also plays an important role in hepatocellular carcinoma [22].

Metallothioneins are transcriptionally regulated by both heavy metals and glucocorticoids, which are highly related to liver cancers [23]. For HCC patients, there is relatively higher copper content and significantly lower content of zinc, leading to a significantly higher value of Cu–Zn ratio in the blood serum [24]. It is also noticed that the MTs and RNAHP are extremely strongly related. In addition, RNAPH (aliases: DDX42) has been proven to relate to cell apoptosis [25].

GPC3 gene may regulate the growth and the tumor predisposition. GPC3 is highly expressed in fetal but not in normal adult liver [26]. GPC3 proteins are distributed widely in the tissues of liver cancer, which may be a novel tumor marker for liver cancer [27].

Now we are ready to give a treatment strategy for HCC. The relationship between the three core genes, composed of GPC3, HAMP and MTs, with iron, copper and zinc in the serum of HCC patients can be summarized as follows:

1) GPC3 may promote cancer cell proliferation and irons are need in this process.

2) HAMP with Low-level expression may provide irons. However, the provided irons are not enough for cancer cells proliferation. As a result, HAMP is always in the low-level expression.

3) The main component of the ceruloplasmin in the liver is copper. The metabolism of iron in the serum can not be separated form ceruloplasmin [28].

4) The metallothionein genes (MTs) with high-level expression will prevent coppers from entering into the serum. As copper has always been in demand, the metallothionein genes (MTs) are in the low-level expression in order to release cuppers.

5) The low-level expression of the metallothionein genes (MTs) may have the adverse effects on the absorption of zinc. Therefore, the content of zinc in the serum will reduce.

On the basis of the above analyses, we propose the solution for the HCC treatment is: Supplement Iron and Zinc after surgery.

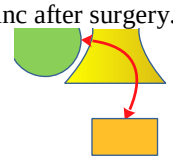


Figure 6. A simple schematic of "Supplement Iron and Zinc after surgery". The solid line represents the mutual promotion and the dashed line indicates mutual inhibition

ISSN 2348 – 7968

The mechanism of "Supplement Iron and Zinc after surgery" depicted by Figure 13 can be addressed as follows:

i) The gene GPC3 in cancer cell is at high-level expression, which promotes cell proliferation in tumours. The goal of surgery is to reduce the total content of the GPC3 genes.

ii) The reason of iron supplementation is to increase the iron content in the serum, which may active the expression of HAMP.

iii) Zinc deficiency associates with many symptoms of liver cancer patients [29]. Therefore, adding zinc can activate the expression of MTs, and produce large amounts of metallothionein to bond copper ions [30]. It is noted that the MTs are highly positive related to HAMP genes. The HAMP gene expression will be up-regulated when the MT genes are up-regulated so that the iron content in the serum may reduce.

iv) As GPC3 are negatively related to MTs and HAMP, the expression level of GPC3 will decline as the increase of expression of MTs and HAMP.

References

[1] G. Baffy. Editorial: Hepatocellular Carcinoma in Type 2 Diabetes: More Than Meets the Eye. Am J Gastroenterol, 2012, 107:53–55.

[2] C. Leong, V. Bexiga, J. P. Teixeira, R. Bugalho, M. Ferreira, P. Rodrigues, J. C. Silva, P. Lousã, J. Varela and I. C. Teixeira. Automatic adjustment of a medical imaging data acquisition system to unknown delays in the input communication channels. Analog Integrated Circuits and Signal Processing, 2012, 70(2):213-227.

[3] S. Nizami, J. Green and C. McGregor. Service oriented architecture to support real-time implementation of artifact detection in critical care monitoring, Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, 2011, 4925 -4928.

[4] Y. Maruyama, Y. Kawamura, T. Nishikawa, T. Isogai, N. Nomura and N. Goshimal. HGPD: Human Gene and Protein Database, Oxford Journals, 2012, 40(D1): 924-929.

[5] C. Lee, and Y. Leu. A novel hybrid feature selection method for microarray data analysis.Applied Soft Computing,2011, 11(1):208–213.

[6] W. Sung, H. Zheng, S. Li, R. Chen et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nature Genetics,2012.

[7]Weisen Pan, Shizhan Chen, Zhiyong Feng. Automatic Clustering of Social Tag using Community Detection. Applied Mathematics & Information Sciences, 2013, 7(2): 675-681.

[8]Weisen Pan, Shizhan Chen, Zhiyong Feng. Service-Oriented Ontology and Its Evolution. Proceedings of 7th International Conference on Grid and Pervasive Computing, Lecture Notes in Computer Science, Vol.7296, Springer, 2012, pp. 109-121.

[9]Weisen Pan, Shizhan Chen, Zhiyong Feng. Investigating the Collaborative Intention and Semantic Structure among Co-occurring Tags using Graph Theory. Proceedings of 16th International Enterprise Distributed Object Computing Conference Workshops, IEEE, 2012, pp. 190-195.

[10]Weisen Pan, Shizhan Chen, Zhiyong Feng. Domain Concept Extraction Model Based on Folksonomy. The 2011 IEEE Asia-Pacific Services Computing Conference, IEEE, 2011, pp. 161-165.

[11]C. Giovannini, L. Gramantieri, M. Minguzzi, F. Fornari et al. CDKN1C/P57 Is Regulated by the Notch Target Gene Hes1 and Induces Senescence in Human Hepatocellular Carcinoma. The American Journal of Pathology, 2012.

[12]B. Damania.Viral-Encoded Genes and Cancer Associated Viruses, Current Cancer Research, 2012, 81-99.

[13]C. Guichard, G. Amaddeo, S. Imbeaud,Y. Ladeiro et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nature Genetics, 2012, 1-7.

[14]O. Humbert, L. Davis, and N. Maizels. Targeted gene therapies: tools, applications, optimization. Critical Reviews in Biochemistry and Molecular Biology, 2012, 47(3):264-281.

[15]Y. Sawada, T. Yoshikawa, D. Nobua, H. Shirakawa et al. Phase I trial of glypican-3-derived peptide vaccine for advanced hepatocellular carcinoma showed immunological evidence and potential for improving overall survival. Clin Cancer Res clincanres, 2011.

[16]Jinyu Hu and Zhiwei Gao. Modules identification in gene positive networks of hepatocellular carcinoma using Pearson agglomerative method and Pearson cohesion coupling modularity[J]. Journal of Applied Mathematics, 2012 (2012).

[17]Jinyu Hu and Zhiwei Gao. Distinction immune genes of hepatitis-induced heptatocellular carcinoma. Bioinformatics, 2012, 28(24): 3191-3194.

[18]Jinyu Hu, Zhiwei Gao and Weisen Pan. Multiangle Social Network Recommendation Algorithms and Similarity Network Evaluation. Journal of Applied Mathematics, 2013 (2013).

[19]Jinyu Hu and Zhiwei Gao. Data-based core genes screening for Hepatocellular Carcinoma[]//Industrial Electronics and Applications (ICIEA), 2014 IEEE 9th Conference on. IEEE, 2014: 1102-1107.

[20]D. Barisani, S. Pelucchi, R. Mariani, S. Galimberti et al. Hepcidin and iron-related gene expression in subjects with Dysmetabolic Hepatic Iron Overload. Journal of Hepatology, 2008, 49(1): 23-133

[21]S. Nekhai, and V. Gordeuk. Iron Metabolism in Cancer and Infection. Iron Physiology and Pathophysiology in Humans, 2012, (4):477-495.

[22]A. Verkman, H. Mariko, and M. Papadopoulos. Aquaporins—new players in cancer biology. Journal of Molecular Medicine, 2008, 86(5):523-529.

[23]M. Pedersena, A. Larsenb, M. Stoltenbergb, and M. Penkowaa. The role of metallothionein in oncogenesis and cancer prognosis. Progress in Histochemistry and Cytochemistry, 2009, 44 (1): 29–64.

[24]E. Abou, E. Abd, G. Magdy, and E. Noha. Copper and zinc levels in serum and tissue in Egyptian patients with hepatocellular carcinoma and cirrhosis. Egyptian Liver Journal,2012, 2(1):7–11.

[25]H. Uhlmann-Schiffler, S. Kiermayer and H. Stahl. The DEAD box protein Ddx42p modulates the function of ASPP2, a stimulator of apoptosisDdx42p counteracts ASPP2. Oncogene, 2009, 28: 2065-2073.

ISSN 2348 – 7968

[26]Y. Wang, Z. Zhu, D. Teng, Z. Yao, W. Gao et al. Glypican-3 expression and its relationship with recurrence of HCC after liver transplantation. World J Gastroenterol, 2012, 18(19): 2408-2414.

[27]A. Mark, and F. Jorge.Therapeutic Potential of Targeting Glypican-3 in Hepatocellular Carcinoma, Anti-Cancer Agents in Medicinal Chemistry, 2011, 11(6): 543-548.

[28]J. Harned, J. Ferrell, S. Nagar, M. Goralska et al. Ceruloplasmin alters intracellular iron regulated proteins and pathways: Ferritin, transferrin receptor, glutamate and hypoxia-inducible factor-1α. Experimental Eye Research ,2012, 97(1):90-7.

[29]M. Pankhurst, D. Gell, C. Butler, M. Kirkcaldie et al. Metallothionein (MT) -I and MT-II Expression Are Induced and Cause Zinc Sequestration in the Liver after Brain Injury. PloS one, 2012, 7(2):1-8.

[30]M. Mohammad, Z. Zhou, M. Cave, A. Barve et al. Zinc and Liver Disease. Nutr Clin Pract February, 2012, 27(1): 8-20.