BIOINFORMATICS DISCOVERY NOTE

Vol. 28 no. 24 2012, pages 3191–3194
doi:10.1093/bioinformatics/bts624

Systems biology

Advance Access publication October 25, 2012

Distinction immune genes of hepatitis-induced heptatocellular carcinoma

Jinyu Hu[1] and David Zhiwei Gao[2,*]

[1]Department of Automation, School of Electrical Engineering and Automation, Tianjin University, Tianjin 300072, China and [2]Division of Electric Engineering and Physics, School of Computing, Engineering and Information Sciences, Northumbria University, Newcastle Upon Tyne, NE1 8ST, UK.

Associate Editor: Martin Bishop

ABSTRACT

Motivation: Hepatitis B virus and hepatitis C virus are the two leading causes resulting in hepatocellular carcinoma (HCC). It is observed that hepatitis C virus (HCV) is relatively difficult to induce HCC compared with hepatitis B virus (HBV). This motivates us to reveal the reasons behind this from the viewpoint of immune genes.

Results: To distinguish the immune genes with low-level expression in HBV-induced HCC, but high-level expression in HCV-induced HCC, the concept of distinction immune gene is proposed. A filter is then designed to screen these genes. By using gene positive network with strong correlations between genes, the genes are further filtered to form the set of key distinction immune genes. The 23 key distinction immune genes are screened, which are divided into four clusters, T cells, B cells, immune signalling and major histocompatibility complex. It is evident that the screened genes are important immune genes, which are activated in HCV-induced HCC, but inactivated in HBV-induced HCC. In HCV-induced HCC, the structures of HCV adaptively update, so that they are difficult to be identified by antigens. Therefore, the clinic advice is either to increase the update speed of antigens or reduce the update speed of the viruses during the treatment of HCV-induced HCC. Moreover, it is also advised to add T cells or add the expression levels of T cells to strengthen the ability to kill cancer cells. In contrast, HBV updates slowly, but the immunity system in HBV-induced HCC has been damaged seriously. As a result, the clinic advice is to improve the immune ability of patients subjected to HBV-induced HCC, such as increasing immunoglobulin, T cells and B cells and so forth.

Contact: zhiwei.gao@northumbria.ac.uk

Received on August 7, 2012; revised on September 19, 2012; accepted on October 14, 2012

1 INTRODUCTION

Hepatocellular carcinoma (HCC) accounts for 70–85% of primary liver cancers, ranking it at the fifth position among the most common malignancies worldwide and in the third position as a cause of cancer death (Baffy, 2012). HCC is a devastating tumour, with a mean survival of much less than one year if left untreated HCC. More than 500 million people worldwide are persistently infected with the hepatitis B virus (HBV) and/or hepatitis C virus (HCV), who are at a risk of developing chronic liver diseases, cirrhosis and even HCC (Rehermann and Nascimbeni, 2005). Increasingly, experimental evidences suggest that viruses, such as HBV and HCV, contribute to HCC by directly modulating pathways that promote the malignant transformation of hepatocytes (Levrero *et al.,* 2012).

HCV infection is one of the major risk factors for HCC. However, cancer arises in only a subset of patients, which takes several years to develop. Therefore, the role of HCV in the aetiology of cancer has remained elusive (Lindenbach, 2012). HCV may activate the immune system, for instance, natural killer cells are activated by the HCV (Amadei and Urbani, 2010). It is also noticed that HCC may escape adaptive immune responses (Burke and Cox, 2010). In addition, cirrhosis is a transition state from normal liver to tumorigenesis in HCV-induced HCC (Huang *et al.,* 2010).

There is a strong incentive to seek key immune genes and identify the relationship between these genes and hepatitis-induced HCC. Motivated by this, a filter is designed to screen these genes, and experiments are implemented by using the data from the Stanford database (Chen *et al.,* 2002). According to the experiment's results and analyses, clinical recommendations are finally given for the treatments of HCV-induced HCC and HBV-induced HCC. Equation-based dynamic network model (Wang *et al.*, 2008) and graph-based gene community network (Hu and Gao, 2012) are usually used to model gene networks. In this study, gene community network will be used to describe gene networks.

2 DATA AND METHODS

2.1 Data

The data come from the Stanford database at http://genome-www.stanford.edu/hcc/Figures/SupTable3.xls (Chen *et al.,* 2002). There are 156 samples, and 3180 genes (represented by 3964 cDNAs). Only 136 samples related to hepatitis B or hepatitis C are chosen for our research, including four types of samples, 72 HBV-induced HCC samples, 4 HCV-induced HCC samples, 54 HBV-infected non-tumour liver samples and 6 HCV-infected non-tumour liver samples.

2.2 Methods

To distinguish the immune genes with low-level expression in HBV-induced HCC, but with high-level expression in HCV-induced HCC, the concept of distinction immune gene is proposed.

*To whom correspondence should be addressed.

Definition 1. *Distinction Immune Gene (DIG)*
A gene is called distinction immune gene (DIG) of the hepatitis-induced HCC provided that the following criteria are satisfied:

(i) The average expression level of the gene in the HCV-induced HCC tissues, simply denoted by HC, is higher than the average expression level of the gene in the HCV-infected non-tumour liver tissues, simply called LC.
(ii) The average expression level of the gene in the HBV-induced HCC samples, simply denoted by HB, is lower than the average expression level of the gene in HBV-infected non-tumour liver samples, simply dented by LB.
(iii) The average expression level of the gene in HCV-infected non-tumour liver tissues is greater than the average expression level of the gene in HBV-infected non-tumour liver tissues.

Definition 1 implies that a DIG should meet the condition HB<LB<LC<HC. As a result, a three-layer filter is proposed as follows:

$$F = F1 \cdot F2 \cdot F3$$

where

$$F1 = \begin{cases} 1, & \mathrm{HC} - \mathrm{LC} > 0, \\ 0, & \text{Otherwise} \end{cases} \quad (1)$$

$$F2 = \begin{cases} 1, & \mathrm{LC} - \mathrm{LB} > 0 \\ 0, & \text{Otherwise} \end{cases} \quad (2)$$

$$F3 = \begin{cases} 1, & \mathrm{LB} - \mathrm{HB} > 0 \\ 0, & \text{Otherwise} \end{cases} \quad (3)$$

By using the three-layer filter, one can obtain the DIGs. To look at the relation among these genes, gene positive network is defined as follows.

Definition 2. *Gene positive network*
Gene positive network (GPN) is used to describe the positive relationship between genes. A connection matrix C is used to store gene community networks, whose element is c_{ij}, defined by:

$$c_{ij} = \begin{cases} p_{ij}, & i > j \text{ and } 0 \le \mathrm{p_{ij}} < T \le 1 \\ 0, & i \le j \end{cases} \quad (4)$$

where p_{ij} is the Pearson correlation coefficient between the i-th gene and j-th gene, and T is the pre-set threshold value (Gao and Hu, 2012). For instance, when $T=0.6$ or $T=0.8$, the obtained GPN represents a strongly positive network or an extremely strong positive network.

The higher is the threshold value, the stronger is the obtained gene positive network, describing the correlations between the genes. In this research, $T=0.8$ is used to screen key DIGs.

To address the importance of the obtained DIGs, the following definition is given.

Definition 3. *Distinction immune gene force*
Distinction immune gene force (DIGF) is defined by

$$\mathrm{DIGF} = \mathrm{SIS} \times \mathrm{GIF} \quad (5)$$

where

$$\mathrm{GIF} = \sum \mathrm{p_{ij}}, 0 \le \mathrm{p_{ij}} < T \le 1 \quad (6)$$

$$\mathrm{SIS} = \mathrm{HC} - \mathrm{HB} \quad (7)$$

Actually, SIS represents the strength of an individual distinction immune gene (DIG), whereas GIF stands for the major impacts of the gene on others genes in the network. As a result, the product of the SIS and GIF, namely DIGF, can describe the degree of importance of this gene in the immune system of the HCC. Greater is the DIGF, more important is the DIG.

3 RESULT

3.1 DIGs

By using the proposed three-layer filter, we obtain 128 DIGs. Deleting the duplicate genes, there are 120 DIGs left, which are shown in Table 1.

3.2 Key DIGs and biological functions

In this subsection, we will look at the correlations between the obtained 128 DIGs. By setting the threshold value as $T=0.8$, an extremely positive-related network is established, shown in Figure 1. It is noted that there are eight sets of genes forming extremely strong correlated positive networks, and the remaining genes have no big impacts on each other. In addition, the gene network with two nodes is generally composed of two same genes with different structures. Therefore, we only emphasize the gene networks with at least three nodes, leading to the four sets of key DIGs, denoted by A, B, C and D in Figure 1.

The biological functions of the four clusters composed of 23 genes are shown by Table 2.

In the clustering set A, there are two major genes, named *COPB2* and *MARVELD2*. It is known that the gene *COPB2* is related to the signal transduction of G protein (Yan *et al.,* 2009), and *MARVELD2* plays an important role in immune system signalling (Mariano *et al.,* 2011). Therefore, the cluster A is about immune signals and immune regulation. In the clustering set B, there are four genes, including *TRA@*, *TRB@*, *TRD@* and *EPC1*, which are all related to the T cells (Nakahata *et al.,* 2009) with immune functions. As a result, the cluster B may reflect the relationship between the abnormalities of T cells and liver cancers. In the cluster C, there are genes, such as *ID4*, *PSMF1*, *IGKC*, *CSF2RA*, *PHC2*, *SULT1E1*, *IGHV*, *IGL@*, *TNFSF10* and so on, which form B cell with immune functions. In the module D, there are three genes, such as *MB*, *HLA-C* and *MSTP9*, to form the major histocompatibility complex (MHC). The proteins encoded by the MHC genes present the antigens to T cells (Yan *et al*., 2009).

3.3 Degree to importance to key DIGs

In this subsection, the degrees of the importance of the four gene clusters are discussed. Setting the threshold at $T=0.6$, the networks of the four clusters are shown by Figure 2. It can be seen

Table 1. One hundred twenty DIGs

One hundred twenty DIGs of hepatitis-included HCC
IGL@, IGHV, TNFSF10, IGHV5-78, SULT1E1, IGKC, PHC2, NKG7, MIG, CD8B1, HNRPU, TUFM, FAM76A, EPC1, Trans, TRD@, ID4, SCYA5, PIM2, TRB@, GBP1, CD79A, CSF2RA, CD3D, LCK, GZMK, CORO1A, PLEK, HLA-DQA1, FLRT3, HABP2, MEOX1, IL2RG, HLA-DQB1, VNN3, TRA@, EMP3, ISL1, ARHGAP30, CTSS, ICAM1, ARMCX3, ISG20, CECR1, SPTLC2, TFF3, PIK3CD, HLA-DQB2, SLA, MSTP9, TBXAS1, BIN2, HSPC022, D6S49E, SLC12A2, CXorf30, COPB2, ENO2, ACP5, OLAH, CFL1, CD53, SEMA5A, TLR2, MARVELD2, LITAF, EVI2A, CDK6, RAB8B, NCF4, G2692X, GLIS3, GABBR1, LACTB2, SPIB, CD3G, PSMF1, CNN1, C4orf19, LOC100506965, GSDMB, RALGAPA1, H6PD, LIPG, QPP, C3AR1, ADAM8, RRBP1, GBP3, AMD1, GPSM3, ASB1, TMEM205, SEMA4D, SLC7A7, ABHD15, HLA-C, MB, RAB31, KCNAB2, ITM2C, MAFB, CMTM7, BACE2, SGK, ZDHHC8, RHOJ, MED11, FAM46A, THRB, C12orf23, G277X, SYK, SLC11A3, LIMK2, CDKN2A, SLC3A2, ZNF218, NQO2, P4HA1

Trans = transcribed locus; G2692X = no gene information of GENE2692X; G277X = no gene information of GENE227X.

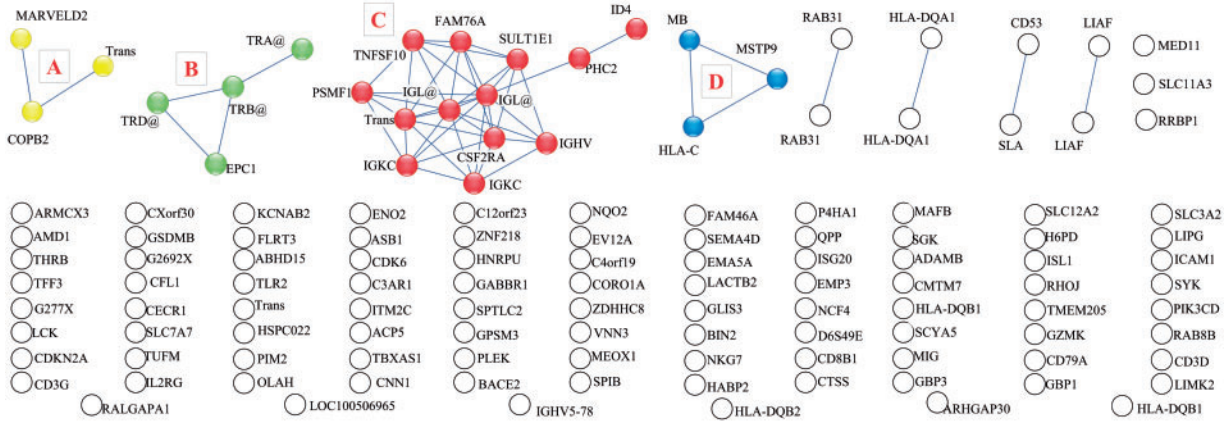

Fig. 1. The extremely strong positive network of 128 DIGs (GPN, $T = 0.8$). A: immune signalling; B: T cell; C: B cell; and D: MHC

Table 2. Function of four clusters

Sets	No.	Function	Symbol
A	3	Immune signalling	MARVELD2, COPB2, Trans
B	4	T cell	TRA@, TRB@, TRD@, EPC1
C	13	B cell	ID4, PSMF1, NO, Trans, IGKC(2), CSF2RA, PHC2, SULT1E1, IGHV, IGL@(2), TNFSF10
D	3	MHC	MB, HLA-C, MSTP9

that the clusters A, B and C are strongly correlated, and the cluster D is isolated. It is clear that the cluster B is the centre of the cluster A–B–C, and the clusters A and C work as the main forces in this cluster.

By using Definition 3, we can calculate the average DIGF for groups A, B, C and D as shown by Table 3. The results are generally consistent for $T = 0.4$, 0.6 and 0.8, which indicates the main force A is much weaker than the main force C. As a result, the clusters B and C have higher degrees of importance.

In consequence, we should pay more attention on the immune genes in the clusters B and C, that is, T cell and B cell genes. The T cell genes play a central role in the immune system, which is agreed with the conclusion by (Bowen and Walker, 2005). Moreover, the B cell genes are a major force in the immune system.

4 SUGGESTION FOR CLINIC TREATMENT

If a human body was regarded as a computer, the HCV and HBV would be the viruses, and the immune system would be the anti-virus software. It is noted that the HCV-induced HCC and HBV-induced HCC are different, thus we need to classify them and treat them in different ways.

4.1 Treatment advices to HCV-induced HCC

The immune genes are activated in HCV-induced HCC, which seem to be in a good position to kill cancer cells. However, HCV is an RNA virus, whose structures are variable. When an antigen identifies one of the structures, the HCV may mutate to form another structure (Rehermann, 2009). Thus, it is challenging to identify the HCV with varying structures. The human body is just like a computer installed with anti-virus software, where the

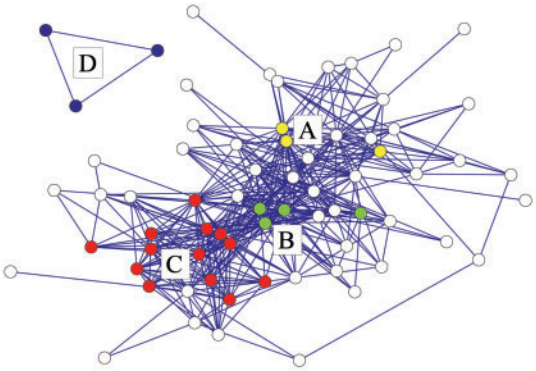

Fig. 2. GPN for clusters with $T=0.6$. Clusters **A**, **B** and **C** are strongly correlated, and the cluster **D** is an isolated module. A: immune signalling; B: T cell; C: B cell; and D: MHC

Table 3. The averaged DIGF values of each cluster

Average	A	B	C	D
DIGF ($T=0.4$)	36.33	67.96	113.70	10.14
DIGF ($T=0.6$)	14.66	35.38	54.87	2.73
DIGF ($T=0.8$)	2.15	4.64	22.52	2.73

update of viruses is always faster than the update of the software, eventually leading to the collapse of the computer. As a result, it is essential either to increase the update speed of antigens or reduce the update speed of the viruses during the treatment of HCV-induced HCC.

Moreover, we have found that the T cells, a centre of DIGs, which is agreed with the conclusion by the reference (Bowen and Walker, 2005), are relatively weaker compared with B cells, the major force of DIGs. The weak T cells may be a reason that cancer cells cannot be killed efficiently. Therefore, it is also advised to add T cells or add the expression levels of T cells for the treatment of HCV-induced HCC.

4.2 Treatment advices to HBV-induced HCC

HBV is a DNA virus, whose structures are rather stable (Liaw and Chu, 2009). Unfortunately, the immune system of HBV–HCC has been seriously damaged, which has been clearly reflected by the downregulated expressions of the DIGs. In this case, the human body is just like a computer without protection from anti-virus software. Although the viruses update slowly, the virus can be infinitely replicated because of the absence of the protection of anti-virus software. Consequently, the computer or the human body will eventually collapse.

To treat the HBV-induced HCC, the human body should be installed with anti-virus software, or called defence system. In other words, we need to strengthen the immune system of the patients with HBV-induced HCC, such as increasing immunoglobulin and T cells, B cells and so on.

5 CONCLUSION

The main contribution is summarized as follows.

5.1 Theory and method

The concept of the DIG has been proposed, and a three-layer filter is designed to screen these immune genes. By using the GPN and DIGF, the degrees of the importance of these genes have been characterized.

5.2 Experimental results

The 23 key DIGs are screened, which are divided into four clusters, T cells, B cells, immune signalling and MHC. It is evident that immune signalling, T cells and B cells form a strong correlated positive network, where T cells are in the centre, and B cells and immune signalling are the two main forces in this community; B cells are the strongest from the viewpoint of distinction immune gene forces in the community.

5.3 Clinical suggestion

For the HCV-induced HCC, the clinic advice is either to increase the update speed of antigens or reduce the update speed of the viruses. Moreover, it is also advised to add T cells or add the expression levels of T cells to strengthen the ability to kill cancer cells. For the HBV-induced HCC, the clinic advice is to improve the immune ability, such as increasing immunoglobulin, T cells and B cells and so forth.

Conflict of Interest: none declared.

REFERENCES

Amadei,B. and Urbani,S. (2010) Activation of natural killer cells during acute infection with hepatitis C virus. *Gastroenterology*, **138**, 1536–1545.

Baffy,G. (2012) Editorial: hepatocellular carcinoma in type 2 diabetes: more than meets the eye. *Am. J. Gastroenterol.*, **107**, 53–55.

Bowen,D. and Walker,C. (2005) Adaptive immune responses in acute and chronic hepatitis C virus infection. *Nature*, **436**, 946–952.

Burke,K. and Cox,A. (2010) Hepatitis C virus evasion of adaptive immune responses: a model for viral persistence. *Immunol. Res.*, **47**, 216–227.

Chen,X. *et al.* (2002) Gene expression patterns in human liver cancers. *Mol. Biol. Cell*, **13**, 1929–1939.

Hu,J. and Gao,Z. (2012) Modules identification in gene positive networks of hepatocellular carcinoma using Pearson agglomerative method and Pearson cohesion coupling modularity. *J. Applied Math.*, doi:10.1155/2012/248658.

Huang,T. *et al.* (2010) Dysfunctional gene/protein networks in hepatitis C virus-induced hepatocellular cirrhosis and carcinoma. In *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology*. New York, NY, pp. 502–507.

Levrero,M. *et al.* (2012) Inflammation and Cancer: Hepatitis B, hepatitis C and HCC. In *Workshop on Inflammation and Cancer*. pp. 13–15.

Liaw,Y. and Chu,C. (2009) Hepatitis B virus infection. *Lancet*, **373**, 582–592.

Lindenbach,B. (2012) Hepatitis C virus and hepatocellular carcinoma. In *Cancer Associated Viruses*. Springer, New York, pp. 571–583.

Mariano,C. *et al.* (2011) Evidence of tricellulin expression by immune cells, particularly microglia. *Biochem. Biophys. Res. Commun.*, **409**, 799–802.

Nakahata,S. *et al.* (2009) Alteration of enhancer of polycomb 1 at 10p11.2 is one of the genetic events leading to development of adult T-cell leukemia/lymphoma. *Genes Chromosomes Cancer*, **48**, 768–776.

Rehermann,B. (2009) Hepatitis C virus versus innate and adaptive immune responses: a tale of coevolution and coexistence. *J. Clin. Invest.*, **119**, 1745–1754.

Rehermann,B. and Nascimbeni,M. (2005) Immunology of hepatitis B virus and hepatitis C virus infection. *Nat. Rev. Immunol.*, **5**, 215–229.

Wang,Z. *et al.* (2008) Stocastic dynamic modelling of short gene expression time series data. *IEEE Trans. NanoBioscience*, **7**, 44–55.

Yan,P. *et al.* (2009) Appliction of random matrix theory to microay data for discovering functional gene modules of hepatocellular carcinoma. *Acta Biophysica Sinica*, **25**, 192–202.