We predicted potential genes and identified pathways associated with hepatitis C. The gene expression profiles of GSE40184 from blood samples and GSE38597 from liver biopsy samples were downloaded from the GEO database. Differentially expressed genes (DEGs) were recognized using the Limma Package. The Pearson correlation test was used to construct the co-expression network of DEGs. Gene set enrichment analysis was used to define significant functions and pathways for DEGs. A total of 165 DEGs in blood samples and 523 DEGs in liver biopsy samples were identified. Eight DEGs were common between these samples. Gene Ontology enrichment analysis showed that 165 DEGs in blood samples were significantly enriched regarding the response to protein binding, receptor binding, G-protein coupled receptor binding, cytokine receptor binding, and cytokine activity. The most significant term of the Kyoto Encyclopedia of Genes and Genomes pathway was the cytokine-cytokine receptor
interaction. Protein-protein interaction network analysis indicated that three subnetworks with more nodes and edges were involved in these interactions. We used robust biomarkers that were differentially expressed in hepatitis C and determined their relevance in the biological function, signal pathways, protein-protein interaction network, and co-expression network of hepatitis C.

**Key words:** Co-expression network; Differentially expressed genes; Hepatitis C; Pathway

**INTRODUCTION**

Hepatitis C is a common infection with a variable course that can result in chronic hepatitis, cirrhosis, and hepatocellular carcinoma (Kensington, 1997). Approximately 60-70% of the various forms of chronic viral hepatitis are caused by hepatitis C virus (HCV). The virus is acquired inconspicuously (Siebert et al., 2003). HCV is a hepatotropic, positive-stranded RNA virus that belongs to the Flaviviridae family (Hoofnagle, 2002) and is related to high morbidity and mortality rates (Shepard et al., 2005).

Histologic examination of the liver is quite important for evaluating patients with chronic hepatitis C (Gebo et al., 2002). It is essential to identify the stage of liver fibrosis when predicting and deciding antiviral treatment (Dienstag, 2002). It has been reported that epidemiological, viral, and host factors are associated with the differences in HCV clearance or persistence and that a strong host immune response against HCV contributes to viral clearance (Cooper et al., 1999). Variation in genes involved in the immune response may contribute to viral clearing (Thomas et al., 2009). The Egyptian HCV epidemic is composed of multiple lineages of genotypes 1 and 4 (Ray et al., 2000). miR-122, a mammalian liver-specific microRNA, is specifically expressed and highly enriched in the human liver, and can regulate HCV gene expression (Jopling et al., 2005). Several studies have analyzed gene expression of hepatitis C. The genome-wide RNA expression profile of genes may be helpful in exposing gene functions in the nosetiology of hepatitis C. In particular, pathway and network analysis would provide useful information for treatment and for understanding the pathogenesis of hepatitis C (Ma et al., 2014). Gene expression analyses on liver biopsy samples from chimpanzees with acute-resolving HCV infections have been performed previously (Su et al., 2002). Changes in liver gene expression in 10 chimpanzees chronically infected with HCV were characterized using DNA microarray analysis (Bigger et al., 2004). Studies also reported that chronic HCV infection has an obvious effect on gene expression in peripheral blood mononuclear cells in infected individuals, and significantly elevates the expression of a subset of interferon-stimulated genes (Bolen et al., 2013) and interferon-stimulated genes (Dill et al., 2012).

The pathogenesis of hepatitis C is complex and not well understood. Therefore, the aim of this study was to survey changes in genes during the occurrence and development of hepatitis C. We investigated differences in gene expression profiles in blood samples and liver biopsy samples from patients with hepatitis C, and identified DEGs that were the same in the 2 groups. Function and pathway enrichment analysis was conducted to determine the pathogenesis of hepatitis C and provide support for gene therapy of hepatitis C.
MATERIAL AND METHODS

Samples

Microarray expression profile data of GSE40184 and GSE38597 were downloaded from the Gene Expression Omnibus (GEO) database. In GSE40184, 16 blood samples from 10 HCV patients and 6 healthy volunteers were analyzed. In GSE38597, 6 liver biopsy samples from 6 patients with acute HCV infection were analyzed. The platform for GSE40184 was GPL570 Affymetrix Human Genome U133 Plus 2.0 Array (Santa Clara, CA, USA). The platform for GSE38597 was GPL96 Affymetrix Human Genome U133A Array.

Data preprocessing and identification of differentially expressed genes (DEGs)

A robust multi array average algorithm was used to perform background correction and quartile data normalization after probe-level data in CEL files were converted into expression measures.

The relationship between the probes and gene symbols was mapped using the files GPL570 and GPL96 in the platform annotation files provided by Affymetrix. A gene was filtered if the probe did not have the corresponding gene symbol. The average value of the gene symbol with multiple probes was further analyzed.

The primary comparison of blood samples to liver biopsy samples was conducted using the Limma Package method. DEGs were identified by assimilating a set of the Student $t$-test results with the threshold of false discovery rate $\leq 0.05$.

Co-expression network construction for DEGs

Biological functions can be conducted when many genes work together; highly co-expressed genes participate in similar biological processes and pathways. The expression values of DEGs were acquired, and the Pearson correlation test was then used to construct a co-expression network.

Protein-protein interaction (PPI) module mining

Biological modules can be approximately reflected using gene sets. The modules in the PPI network were tested using the MCODE Plugin of Cytoscape software. Modules with specific functions could be identified with MCODE by selecting the clusters of densely connected nodes from the network. Degree cutoff = 2, K-core = 2, and maximum depth = 100 were set as parameters in MCODE to detect modules in the PPI network of DEGs.

Topological coefficients

The topological coefficient (Stelzl et al., 2005) $T_n$ of a node $n$ with $k_n$ neighbors was computed using equation (1) as follows:
Here, \( J(n,m) \) was defined for all nodes \( m \) sharing at least one neighbor with \( n \). The value \( J(n,m) \) was the number of neighbors shared between the nodes \( n \) and \( m \), plus 1 if there was a direct link between \( n \) and \( m \).

The topological coefficient is often used to measure the extent to which a node shares neighbors with other nodes. The topological coefficients for all nodes with more than 1 neighbor were computed using Network Analyzer in the network. Nodes with 1 or no neighbors were assigned a topological coefficient of 0 (zero).

**Functional enrichment and pathway enrichment analysis**

Gene Ontology (GO) functional enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for modules were performed using the online tool Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang et al., 2008). Biological meaning could be systematically extracted from a large number of genes or proteins using DAVID bioinformatic resources containing an integrated biological knowledgebase and analytic tools. GO terms and KEGG pathways with a P value less than 0.05 were selected based on the Gene Set Enrichment Analysis (GSEA) test implemented in DAVID. The principle of GSEA is as follows:

\[
Z_k = \frac{1}{\sqrt{|K|}} \sum_{k \in K} t_k
\]

(Equation 2)

where \( K \) denotes the gene set, \( t_k \) is the \( t \)-statistic of each gene, and \( |K| \) is the number of genes in the gene set; \( Z_k \) showed an approximately standard normal distribution.

**RESULTS**

**Identification of DEGs**

After normalization and preprocessing of the expression profile data of blood samples using the Limma package, we obtained 165 DEGs with a false discovery rate \( \leq 0.05 \), including 39 up-regulated genes and 126 down-regulated genes. For liver biopsy samples, 523 DEGs with a false discovery rate \( \leq 0.05 \) were obtained, including 266 up-regulated genes and 257 down-regulated genes. There were 8 common genes in the 2 groups of DEGs, including \( IKZF1, TFPI2, FCGR3B, IFI44L, CD86, MEF2C, ASAP1, \) and \( SPP1 \).

**Co-expression network of DEGs**

A previous study reported that genes with similar functions typically have similar expression patterns in the co-expression network (Firestein, 2003). The expression values of the top 20 up-regulated and top 20 down-regulated DEGs were compared using the Pearson cor-
relation test to construct the co-expression network. The co-expression network of the top 20 up-regulated genes and top 20 down-regulated genes is shown in Figures 1 and 2, respectively. In the co-expression network of the top 20 up-regulated genes, genes **N1AP**, **LGMN**, **SLC7A8**, **CD19**, **MEF2C**, and **CD180** showed more edges than others. For down-regulated genes, genes **IL6**, **INHBA**, **CXCL3**, **MREG**, **IL1RN**, **MAFF**, **QPCT**, and **BHLHE41** showed more edges.

**Figure 1.** Co-expression network of the top 20 up-regulated genes.

**Figure 2.** Co-expression network of the top 20 down-regulated genes.

**PPI module mining**

The MCODE plugin of the Cytoscape software was used to mine modules for PPI sub-networks of DEGs. When we set degree cutoff = 2, K-core = 2, maximum depth = 100, we obtained 3 sub-networks. The PPI network and sub-networks are shown in Figure 3. Sub-network 1 included 11 nodes and 31 edges, sub-network 2 included 7 nodes and 16 edges, and sub-network 3 included 10 nodes and 18 edges. The following rank in importance was observed: 1 > 2 > 3. Nodes represent biomolecules and edges between nodes indicate physical or functional interactions in a molecular network (Sharan et al., 2007).
The topological coefficient was used to study the characteristics of the interaction network (Goldberg and Roth 2003). The topological coefficient is shown in Figure 4. The topological coefficient slowly decreased with the number of neighbors, closely following a power law of $y = ax^b$. The correlation = 0.858, $a = 0.707$, $b = -0.463$, and the $R^2$ was 0.759. The low number of neighbors and high topological coefficient indicated that proteins may lie in the boundary between the 2 clusters. Proteins with a large number of neighbors showed a low topological coefficient. Among these, proteins with higher topological coefficients may be involved in the network. The results showed that $IL1A$ had a topological coefficient of 0.26229508 and a degree of 13, $CXCL1$ had a topological coefficient of 0.30555556 and a degree of 12, $CXCL5$ had a topological coefficient of 0.28282828 and a degree of 11, and $CSF1$ had a topological coefficient of 0.25373134 and a degree of 13.

**Figure 3.** PPI sub-networks of DEGs. A. PPI network; B. sub-network 1; C. sub-network 2; D. sub-network 3.

**Topological coefficients**

The topological coefficient was used to study the characteristics of the interaction network (Goldberg and Roth 2003). The topological coefficient is shown in Figure 4. The topological coefficient slowly decreased with the number of neighbors, closely following a power law of $y = ax^b$. The correlation = 0.858, $a = 0.707$, $b = -0.463$, and the $R^2$ was 0.759. The low number of neighbors and high topological coefficient indicated that proteins may lie in the boundary between the 2 clusters. Proteins with a large number of neighbors showed a low topological coefficient. Among these, proteins with higher topological coefficients may be involved in the network. The results showed that $IL1A$ had a topological coefficient of 0.26229508 and a degree of 13, $CXCL1$ had a topological coefficient of 0.30555556 and a degree of 12, $CXCL5$ had a topological coefficient of 0.28282828 and a degree of 11, and $CSF1$ had a topological coefficient of 0.25373134 and a degree of 13.
Differentially expressed genes related to hepatitis C

Go enrichment analysis

The GO term “molecular function” was used as an over-represented term to gain insight into the biological processes associated with regulated genes. The “molecular function” GO term identified factors related to protein binding (P = 0.007), receptor binding (0.0002), G-protein coupled receptor binding (0.0005), cytokine receptor binding (1.31E-08), and cytokine activity (2.09E-11).

Pathway enrichment analysis

In this study, we conducted GSEA analysis of various factors, including the statistical significance of the set of DEGs in the pathway, the topology of the signaling pathway, and their interactions, among others. The impact analysis method yielded many significant pathways, including the cytokine-cytokine receptor interaction pathway (5.87E-06), hematopoietic cell lineage pathway (1.38E-09), and rheumatoid arthritis pathway (3.64E-05). The rheumatoid arthritis pathway included the Toll-like receptor signaling pathway. The 3 pathways are shown in Figure 5. DEGs such as CXCL3, IL8, IL3RA, IL1B, and LTBR were involved in the cytokine-cytokine receptor interaction pathway. ITGAM and IL3RA were involved in the hematopoietic cell lineage pathway. TNF and IL6 were involved in all 3 pathways.
DISCUSSION

In this study, we used bioinformatic analysis to compare the gene expression profiles of samples of hepatitis C patients and control subjects. A total of 165 DEGs were identified, including 39 up-regulated genes and 126 down-regulated genes. We identified 3 statistically significant KEGG pathways and identified some DEGs involved in these pathways, indicating that these genes play important roles in hepatitis C development.

Based on the co-expression networks of DEGs, we speculated that genes \textit{IL6}, \textit{NAZP}, \textit{LGMMN}, \textit{SLC7A8}, \textit{CD19}, \textit{MEF2C}, and \textit{CD180} of the top 20 down-regulated genes play important roles. Moreover, the effect of \textit{IL6 C174G} gene polymorphisms on the treatment of acute and chronic hepatitis C in human immunodeficiency virus (HIV)-infected patients has been reported previously (Nattermann et al., 2007), confirming our results. \textit{IL6}, a major proinflammatory cytokine, is produced in various tissues, including activated leukocytes, adipocytes, and endothelial cells (Pradhan et al., 2001). As previously reported, an \textit{IL6} promoter polymorphism is associated with a lifetime risk of developing Kaposi’s sarcoma in men infected...
with HIV (Foster et al., 2000). *IL6* is significantly associated with breast cancer (Heffler et al., 2005), lung cancer (Noponen-Hietala et al., 2005), intervertebral disc disease (Noponen-Hietala et al., 2005), and neuroblastoma (Egler et al., 2008). The expression of *IL6* receptor components in the myocardium of donor hearts before transplantation was investigated by Plenz et al. (2002).

As shown in Figure 5, *TNF* was found to be involved in the cytokine-cytokine receptor interaction pathway, hematopoietic cell lineage pathway, and rheumatoid arthritis pathway. Moreover, *TNF* showed the highest degree (39) in the PPI network. These results indicate that *TNF* is related to hepatitis C. *TNF* is a critical proinflammatory cytokine (Chu, 2013) that plays an important role in immunity and inflammation, as well as in controlling cell proliferation, differentiation, and apoptosis (Baud and Karin, 2001). *TNF* plays a central role in regulating HIV-1 replication (Lazdins et al., 1997), initiating toxic shock symptoms (Miettikke et al., 1992), and on acute systemic viral diseases (Clark, 2007). *TNF*-α initiates the cytokine cascade, and high levels have been associated with dementia and atherosclerosis in persons aged 100 years (Brunsungaard et al., 2003). A previous study revealed an inverse association between adipose tissue adiponectin and *TNF*-α expression, suggesting that adiponectin, similarly to *TNF*-α, plays an important role in obesity-associated insulin resistance (Kern et al., 2003). The expression of *TNF*-α protein in the subcutaneous and visceral adipose tissue was studied by Winkler et al. (2003). In conclusion, *TNF* is a key element in many diseases, including inflammation, and thus we speculate that *TNF* is related to hepatitis C; however, these conclusions must be confirmed.

Our results showed that *IL1A* with high topological coefficient and high degree involved in the PPI network may be used as a target gene for hepatitis C treatment. *IL1A* is a pro-inflammatory cytokine with many biological effects, including activation of many inflammatory processes (through the activation of T cells, for example), induction of expression of acute-phase proteins, an important function in neuroimmune responses, and direct effects on the brain itself (Allan et al., 2005). *IL1A* is a novel regulator of the blood-testis barrier in rats (Sarkar et al., 2008) and plays an important role in host defense mechanisms (Dinarello et al., 1987). Studies have shown that *IL1A* expression can affect the expression of other factors and itself. Bacterial cell wall polymers may stimulate *IL-1* and *IL-1A* expression *in vivo* and *in vitro* (McCull et al., 1994). *IL4* expression in human T cells is selectively inhibited by IL1-alpha and IL1-beta (Sandborg et al., 1995).

The investigation of topological properties of proteins in the PPI network can be used to discern basic protein functions and mechanisms of action (Wang et al., 2011). In this study, topological coefficients of the PPI network were used to analyze proteins in the network and showed that the degree of fit was quite good. This indicates that the results were reliable.

**CONCLUSIONS**

In this study, we identified 8 common genes that were differentially expressed in blood samples and liver biopsy samples. The pathogenic genes of hepatitis C were identified by analyzing the expression profiles of blood and liver biopsy samples. These genes, which are related to protein binding, G-protein coupled receptor binding, cytokine receptor binding, cytokine activity, cytokine-cytokine receptor interaction, hematopoietic cell lineage, and the Toll-like receptor signaling pathway, may be target genes for hepatitis C treatment. However, the results of our study should be further confirmed. Genes of *IL6*, *TNF* and *IL1A* and several other genes involved were identified, which may be related to the nosogenesis of hepatitis C.
Conflict of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research received no specific grants from any funding agency in public, commercial, or not-for-profit sectors.

REFERENCES

Clark IA (2007). How TNF was recognized as a key mechanism of disease. Cytokine Growth Factor Rev. 18: 335-343.
Differentially expressed genes related to hepatitis C


