Identification of modules of hepatic encephalopathy based on protein-protein network and gene expression data

HAO WU¹, MIAO LIU¹ and JIAJUN ZHUANG²

¹Department of Internal Medicine, Jining Psychiatric Hospital, Jining, Shandong 272051; ²No. 1 Department of Neurology, Weifang People's Hospital, Weifang, Shandong 261000, P.R. China

Received August 22, 2017; Accepted January 30, 2018

DOI: 10.3892/etm.2018.5924

Abstract. Hepatic encephalopathy (HE) is regarded as a complication of liver cirrhosis, and 50-75% of patients who have been diagnosed with cirrhosis have HE syndrome. The aim of this study was to identify genes and pathways associated with HE alcoholics. Human protein-protein interactions were downloaded from the STRING database. Gene expression data were downloaded from EMBL-EBI. Combined score and Pearson's correlation coefficient were calculated to construct differential co-expression networks. Graph-theoretical measure was used to calculate the module connectivity dynamic score of multiple differential modules. In total, 11,134 genes were obtained after mapping between probes and genes. Then, 501,736 pairs and 16,496 genes were obtained to form background protein-protein interaction networks, 1,435 edges and 460 nodes were obtained constituting differential co-expression networks. Twenty-three seed genes and 10 significantly differential modules were identified. Four significantly differential modules which had larger connectivity alternation were observed. The identified seed genes and significantly differential modules offer novel understanding and molecular targets for the treatment of HE alcoholics.

Introduction

Hepatic encephalopathy (HE) is a complication of liver cirrhosis (1), and 50-75% of patients who have been diagnosed with cirrhosis have the HE syndrome (2,3). The HE also has other types classified according to the standard constituted at the World Congress of Gastroenterology in 1998, including encephalopathy associated with acute liver disease and encephalopathy without liver disease (4). It is reported that the damage of astrocytes triggers disturbance of neurotransmission and induces development of HE (5). Astrocytes play an important role in the central nervous system (CNS) and constitute about a third of brain cortical volume (6). Astrocytes act as the main regulators in neurotransmitters, such as transition of amino acids (7). Glutamine synthase, which maintains an ammonia balance in the human body, converts glutamate and ammonia into glutamine. Glutamine synthase is mainly contained in astrocytes (8).

Ammonia is known as a main cause of HE. It is reported that most of HE is associated with high a concentration of ammonia (9). Ammonia is mainly produced in the gut and synthesized by bacteria. High levels of ammonia disturb the neural system function and neurotransmission. Up to now, the main treatment for HE focuses on reducing the production and concentration of ammonia (10). It is reported that sodium benzoate and sodium phenylacetate are used as ammonia cleaners in patients who are diagnosed with hyperammonemia or urea cycle disorder (11). Glycerol phenylbutyrate is used to treat HE through the regulation of ammonia metabolism. Glycerol phenylbutyrate is a pro-drug of phenylacetate and decomposes nitrogen to urinary phenylacetylglutamine (12). Polyethylene glycol 3350 is used to clean the intestine and has proven to be a more effective therapy than lactulose (β-1, 4-galactosido-fructose) (13). Bass et al reported rifaximin significantly decreased the risk of HE compared with placebo and approximately 90% of patients were treated with lactulose (14). Lactulose, a disaccharide, cannot be digested by human intestinal disaccharidases (15). Lactulose is digested into small molecular organic acids, such as acetic acid and lactic acid. The osmotic effect produced by these acids induces fermentative diarrhoea (16). Lactulose was first reported to be used as treatment for HE in 1966 (17). Rahimi et al reported that lactulose caused more acute electrolyte imbalance and loss than PEG (18). Besides that, hepatic encephalopathy occurs as a complication of alcoholic liver disease mainly found in the most advanced stage (19). And while hepatic encephalopathy occurs in earlier stages, it is usually a consequence of excessive alcohol consumption before the onset of encephalopathy.

In addition, substantial molecular research is associated with HE. Protein-protein interactions (PPIs) processed by bioinformatic algorithms have been used to search for biomarkers and biological pathways in various types of cancer, such as breast cancer (20), lung cancer (21,22), colon cancer (23), ovarian cancer (24) and glioma (25).
In the present study, we downloaded all human PPI networks and gene expression data associated with HE alcoholics. Differential co-expression networks (DCNs), which comprise 1435 edges and 460 nodes were constructed based on PPI networks and gene expression data. Twenty-three seed genes and 13 multiple differential modules (M-DMs) were identified. Ten differential modules were found when P-values were <0.05. Four differential modules had major connectivity alternation using the graph-theoretical measure method.

Materials and methods

Gene expression data. Gene expression data associated with HE alcoholics E-GEOD-53808 were downloaded from the European Molecular Biology Laboratory at European Bioinformatics Institute (EMBL-EBI). The data contained 9,608 genes, 15 non-HE alcoholic samples and 8 HE alcoholic samples.

PPI networks. All human PPI networks were downloaded from the STRING database. There were 787,896 pairs of PPI networks and 16,730 genes. PPI network pairs were selected when the combined scores were >0.2. Then, 501,736 pairs and 16,496 genes were obtained, and these genes and networks formed background PPI networks. Expression profiling was chosen if they contained genes belonging to background PPI networks. The new expression profiling data contained 9,608 genes.

Construction of DCNs. The PPI networks were selected from background PPI networks if they contained genes belonging to the new expression profiling. The Pearson's correlation coefficients of HE and non-HE alcoholics were calculated respectively. Edges were observed when absolute values of Pearson's correlation coefficient were greater than δ (δ=0.9). P-values of genes in the DCNs of two groups were calculated using one-side t-test algorithm. Weight value between gene i and gene j was calculated as:

\[
\begin{align*}
\omega_{ij} = \begin{cases} 
\frac{(\log p_i + \log p_j)^{1/2}}{(2 \times \max_{\sigma \in \sigma} \log p_{\sigma})^{1/2}}, & \text{if } \text{cor}(i, j) \geq \delta, \\
0, & \text{if } \text{cor}(i, j) < \delta,
\end{cases}
\end{align*}
\]

V denotes a node set of DCN.

Identification of M-DMs. M-DMs were identified from DCNs. The process comprises of three steps: i) seed prioritization, ii) module search and iii) refinement of M-modules. Importance value (score value) of each gene in DCNs was calculated as:

\[
g(i) = \sum_{j \in N_k(i)} A_{ij} g(j)
\]

Each network had an adjacent matrix. N_k(i) represents adjacent nodes of gene i in G_k network, A_k represents adjacent matrix which was weighted by normalized degree; g(i) equals z-score. The z-score of each gene in the DCNs was averaged and ranked. One seed gene was regarded as differential module C. Then, gene u, which was adjacent to gene v, was joined into module C to form module C'. The entropy of the two modules was calculated as:

\[
H_k(C) = -p_k^L \log p_k^L - (1 - p_k^L) \log(1 - p_k^L)
\]

\[
p_k^L = \frac{L_k(i)}{L_k(i) + L_k(-i)}
\]

\[
L_k(i) = \sum_{j \in \delta_k} a_{ij}
\]

L_k(i) denotes total weight between gene i and other nodes in modules C. L_k(i) represents weight between gene i and other nodes in modules C.

The candidate modules which had <4 nodes were eliminated. Two modules were merged into one module if the overlapped degree of two modules was >0.05. In total, 13 modules were identified.

Significant statistical test of candidate modules. One thousand, four hundred and thirty-five edges were selected from 178,888 edges and regarded as the random network. Module search processing was done following the above methods. Construction of random networks was repeated 100 times and 3,696 modules were generated. The P-value of the candidate module was calculated as the probability of the module, which has the observed score or smaller by chance. The Benjamini-Hochberg method (26) was used to correct the P-value. Ten modules were identified as significantly differential modules as P-values were ≤0.05.

Quantification of M-DMs connectivity and significance analysis. Graph-theoretical measure method was used to calculate module connectivity dynamic score (MCDS). MCDS between two adjacent modules was calculated as:

\[
MCDS = \Delta A_{ij} C = \left| A_{ij} - A_{ij} C \right| / |C|
\]

C denotes one differential module, A_{ij} presents adjacent matrix of C. The total MCDS of differential module C equals the average of MCDS of all adjacent modules.

\[
\tau(C) = \sum_{A_{ij}} \Delta A_{ij} C / (M - 1)
\]

Differential modules were regarded as significantly differential modules of connectivity alternation at P-value <0.05.

Results

Gene expression data and PPI network processing. Gene expression data E-GEOD-53808 were downloaded from the EMBL-EBI database, and 11,134 genes were obtained following mapping between probes and genes. The human PPI networks were downloaded from the STRING database. Networks comprised of 787,896 pairs of PPI networks and 16,730 genes.

PPI network pairs were selected if their combined scores were >0.2. Pairs (50,1736) and 16,496 genes were obtained to form background PPI networks. The new expression profiling data identified contained 9,608 genes.

Identification of DCNs. Interaction pairs (178,888) which contained genes of new expression profiling were identified. These pairs were selected from background PPIs. The edges were selected if absolute values of Pearson's correlation coefficient were >0.9. One thousand, four hundred and thirty-five edges and 460 nodes were obtained and constituted DCNs. Two DCNs, constituting the non-HE alcoholics and HE alcoholics groups, were identified. Two DCNs had the same node sets and different edge sets.
Identification of M-DMs. The important value (z-score) of each gene was calculated. The z-score of each gene in all the DCNs were averaged and ranked. Top 5% genes were selected as seed genes and 23 seed genes were identified (Table I).

Modules were searched based on each gene. Entropy of two modules and change of entropy were calculated. Candidate modules were eliminated if nodes were <5. Two modules were merged into one module if the overlap degree was ≥0.5. Finally, 13 modules were identified.

Significant test of candidate modules. One thousand, four hundred and thirty-five edges were chosen from 178,888 edges randomly to constitute one random network. The random networks were constructed 100 times, and gene rated 3,696 modules. After P-values were corrected by the Benjamini-Hochberg algorithm, 10 significant differential modules were identified.

Quantification of connectivity dynamics of M-DMs and statistical significant test. Graph-theoretical measure algorithm was used to calculate MCDS. MCDS between two modules and total MCDS were calculated. Four significant differential modules, which had larger connectivity alternation were found (P<0.05) (Fig. 1). Module 1 contains 80 nodes and 478 edges; module 2 comprises 36 nodes and 119 edges; module 3 contains 42 nodes and 145 edges; module 4 comprises of 25 nodes and 64 edges.

Discussion

Alcohol damage on hepatic cells impairs enzyme activities and detoxification of liver. In addition, impairment of hepatic cells disturb brain functions (27). Hepatic encephalopathy (HE) is a serious complication of alcoholic-associated hepatic disease (19). HE has various clinical syndromes, such as nausea, malaise, asterixis and coma (28,29).

In the present study, we obtained all the PPI networks and gene expression data from international databases. We identified 23 seed genes and 10 significant differential modules associated with HE alcoholics. Four differential modules, which had lager connectivity alternations, were obtained by calculating the module connectivity dynamic score of M-DMs.

The 23 genes identified in this study included PSMA3, RPS13, RPL8, PSMA2, AHS1, RPS5, EEF1B2, PSMC2, ATP5B and NDUFV2. PSMA3 is one subunit of 20S proteasome and interacted with ROA1, PCBP2 and IREB2. These
proteins function in RNA processing, such as splicing (30). PSMA3 (rs2348071) GG homozygote is related to children with asthma, and children with GG homozygotes are susceptible to disease in Taiwan (31). The potential function of PSMA3 was analyzed in HE alcoholics.

Proteasome 26S subunit ATPase 2 (PSMC2) is an important component of 26S proteasome. The complex processes substrates into 20S pivotal component (32). The complex, which plays an important role in recognition of initi-

| Name    | z-score 1 | z-score 2 | Average |
|---------|-----------|-----------|---------|
| PSMA3   | 86.535583 | 122.6082221 | 104.57089 |
| RPS13   | 34.8732349 | 120.1981589 | 77.535697 |
| RPL8    | 25.1907276 | 97.03482872 | 61.112778 |
| PSMA2   | 57.812847 | 64.10648625 | 60.959667 |
| AHS1A   | 81.0358135 | 37.6419885 | 59.338901 |
| RPS5    | 22.1380737 | 96.04627227 | 59.092173 |
| EEFF1B  | 29.0671882 | 88.5651997 | 58.816194 |
| RPS12   | 20.7253557 | 95.23469314 | 57.980024 |
| RPL35   | 26.199604 | 89.41832806 | 57.808966 |
| PSMC2   | 58.4824325 | 50.02097225 | 54.251702 |
| ATP5B   | 31.4457881 | 76.47658876 | 53.961188 |
| SPCS1   | 37.2801529 | 69.5430352 | 53.411594 |
| NDUFAB1 | 26.0584344 | 79.74001391 | 52.899224 |
| EIF1B   | 43.2034053 | 62.06634957 | 52.634877 |
| NDUFB3  | 25.9831661 | 78.69291867 | 52.338042 |
| NDUFV2  | 33.9425327 | 69.17172639 | 51.55713 |
| RPL15A  | 20.9168676 | 81.97652628 | 51.446697 |
| RPL19   | 25.2334709 | 76.47090293 | 50.852187 |
| GAPDH   | 23.1407005 | 76.70369699 | 49.9222 |
| PSMA5   | 52.2804348 | 47.28292135 | 49.781678 |
| RPL18   | 23.4177997 | 73.9560672 | 48.686933 |
| SYP     | 25.0252229 | 71.74102033 | 48.383122 |
| RPS16   | 23.0946453 | 73.32578699 | 48.210216 |

Table I. Twenty-three seed genes identified by z-score.

It is reported that EEFF1B has four loci, including EEFF1B1, EEFF1B2, EEFF1B3 and EEFF1B4. Three of them are functional (39). EEFF1B2 is mapped to chromosome 2 by PCR method (40). EEFF1B2 is related with retrotransposition function and capable of transcriptional activity (41).

NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2) encodes one 24 kDa component of NADH-ubiquinone oxidoreductase complex and is involved in electron transportion (42). NDUFV2 is involved in neuronal mobility and psychiatric dysfunction (43). The 3542G>A polymorphism, which is located at the promoter region of NDUFV2, is associated with bipolar disorder (44). Therefore, we suggested that NDUFV2 may play an important role in HE progression.

In conclusion, we identified 23 seed genes and 10 significantly differential modules associated with HE alcoholics. Four modules were identified as they had relatively larger connectivity alternations. The newly identified seed genes and modules offer understanding of the potential mechanisms and biomarkers for the therapeutic target of HE alcoholics.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author on reasonable request.

Authors' contributions

HW conceived and designed the study and wrote the manuscript; ML performed the data analyses; JZ contributed to the conception of the study. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Ferenci P, Lockwood A, Mullen K, Tarter R, Weissenborn K and Blei AT: Hepatic encephalopathy - definition, nomenclature, diagnosis, and quantification: Final report of the working party at the 11th World Congresses of Gastroenterology, Vienna, 1998. Hepatology 35: 716-721, 2002.
2. Cichoz-Lach H and Michalak A: Current pathogenetic aspects of hepatic encephalopathy and noncirrhotic hyperammonemic encephalopathy. World J Gastroenterol 19: 26-34, 2013.

3. Patel D, McPhail MJ, Cobbold JF and Taylor-Robinson SD: Hepatic encephalopathy. Br J Hosp Med (Lond) 73: 79-85, 2012.

4. Atfuri DK, Prakash R and Mullen KD: Pathogenesis, diagnosis, and treatment of hepatic encephalopathy. J Clin Exp Hepatol 1: 77-86, 2011.

5. Hänselmann D, Kircheis G, Fischer R, Schliess F and vom Dahl S: Hepatic encephalopathy in chronic liver disease: A clinical manifestation of astrocyte swelling and low-grade cerebral edema? J Hepatol 32: 1035-1038, 2000.

6. Wilson JX: Antioxidant defense of the brain: A role for astrocytes. Can J Physiol Pharmacol 75: 1149-1163, 1997.

7. Henn FA and Hamberger A: Glial cell function: Uptake of transmitter substances. Proc Natl Acad Sci USA 68: 2686-2690, 1971.

8. Torgner J and Kvatme E: Synthesis of transmitter glutamate and the glial-neuron interrelationship. Mol Chem Neuropathol 12: 11-17, 1990.

9. Adams RD and Foley JM: The neurological disorder associated with liver disease. Rev Publ Assoc Res Nerv Ment Dis 32: 198-237, 1953.

10. Tapper EB, Jiang ZG and Patwardhan VR: Refining the ammonia hypothesis: A physiology-driven approach to the treatment of hepatic encephalopathy. Mayo Clin Proc 90: 646-658, 2015.

11. Ghabril M, Zupanets IA, Vierling J, Mantry P, Rockey D, Wolf D, O’Shea R, Dickinson G, Killsay H, Norris C, et al: Glycophosphatidylinositol (GPI)-linked patients with cirrhosis and episodic hepatic encephalopathy: A pilot study of safety and effect on venous ammonia concentration. Clin Pharmacol Drug Dev 2: 278-284, 2013.

12. Dobler J, Rheed W, Diaz GA, Scharschmidt BF, Mian A, Shchelochkov O, Marier JF; Beliveau M, Mauney J and Dickinson K: Phase 2 comparison of a novel ammonia scavenging agent with sodium phenylbutyrate in patients with urea cycle disorders: Safety, pharmacokinetics and ammonia control. Mol Genet Metab 100: 221-228, 2010.

13. Rahimi RS and Rockey DC: Hepatic encephalopathy: Pharmacological therapies targeting ammonia. Semin Liver Dis 36: 48-55, 2016.

14. Bass NM, Mullen KD, Sanyal A, Poorbad F, Neff G, Leeby CV, Sigal S, Stolovitch-Hunt Y, Slivers K, El-Mofty S, et al: Rifaximin treatment in hepatic encephalopathy. N Engl J Med 362: 1071-1081, 2010.

15. Mozdziak PE, Populai S, Borworrpinnoy S and Pettyte JN: Transgenic chickens expressing beta-galactosidase hydrolyze lactose in the intestine. J Nutr 137: 3076-3079, 2007.

16. Haemmerli UP, Kistler H, Ammann R, Marthaler T, Semenza G, Auricchio S and Prader A: Acquired milk intolerance in the adult intestine. J Nutr 133: 3076-3079, 2003.

17. Adams RD and Foley JM: The neurological disorder associated with liver disease. Res Publ Assoc Res Nerv Ment Dis 32: 177-133, 1953.

18. Pizzuti A, Gennarelli M, Novelli G, Colosimo A, Lo Cicero S, Caskey CT and Dallapiccola B: Human elongation factor-1 beta: Cloning and characterization of the EF1 beta 5a gene and assignment of EF-1 beta isoforms to chromosomes 2,1,5 and X. Biochem Biophys Res Commun 197: 154-162, 1993.