



# BioTechnology

An Indian Journal

FULL PAPER

BTALJ, 8(11), 2013 [1497-1500]

## The correlational research hotspots analysis of Hepatitis virology by PubMed

Li Jianguang<sup>1</sup>, Zhao Jinyuan<sup>2</sup>, He Jundong<sup>3</sup>, Zhang Bin<sup>3</sup>, Li Jungai<sup>1</sup>, Zhao Hongxia<sup>1</sup>, Zhang Gairu<sup>2</sup>, Gan Lanzhuang<sup>2</sup>, Jing Jianmei<sup>4</sup>, Zhang Lifang<sup>2</sup>, Liu Tao<sup>2</sup>, Guo Yicheng<sup>1</sup>, Ma Liangjun<sup>5</sup>, Hou Jinjie<sup>4\*</sup>

<sup>1</sup>Xingtai Medical College, 054000, Xingtai, Hebei, (CHINA)

<sup>2</sup>Hebei Civil Administration General Hospital, 054000, Xingtai, Hebei, (CHINA)

<sup>3</sup>Xingtai University, 054000, Xingtai, Hebei, (CHINA)

<sup>4</sup>Second Affiliated Hospital, Xingtai Medical College, 054000, Xingtai, Hebei, (CHINA)

<sup>5</sup>Liver Disease Branch, Traditional Chinese Medicine hospital, Qinghe County, 054800, Xingtai, Hebei, (CHINA)

E-mail: wein871@sohu.com

### ABSTRACT

Objective: To understand the research hotspots of Hepatitis virology by PubMed. Methods: With MS Excel, SPSS, Cytoscape software, we took MeSH (Medical Subject Headings) word frequency analysis, clustering analysis, co-word network graph of PubMed papers. Results: It shows that the current Hepatitis virology research hotspots had focus on Hepatitis C and B, Antiviral Agents, etc, also the most importance of which was the Hepatitis C. Conclusion: It is helpful and timesaving for researcher or doctor to understand the research hotspots in Hepatitis virology.

© 2013 Trade Science Inc. - INDIA

### KEYWORDS

Hepatitis;  
Virology;  
Word frequency analysis;  
Clustering analysis;  
Co-word network graph;  
Hepatitis C.

### INTRODUCTION

There are about 2 billion people around the world who have been infected with the hepatitis b virus, including more than 350 million people suffering from chronic infection, 500000 to 700000 people each year die from hepatitis b virus infection. About 130 to 170 million people have been infected with chronic hepatitis c virus (HCV), an estimated 350000 people have died of liver disease associated with hepatitis c<sup>[4]</sup>.

Viral hepatitis is widely circulated in the world at present, one of the most infectious diseases, we hope that through this research the analysis of the MeSH can draw the outline of hepatitis virology correlational re-

search hotspot.

Therefore this research retrieved the hepatitis papers of PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>) within recent five years, got 21786 papers, and analyzed MeSH of above papers using Co-word Analysis<sup>[6]</sup>.

### MATERIALS AND METHODS

First, we retrieved PubMed papers with publication dates between 27 February 2008 and 27 February 2013. Second, search terms was "Hepatitis"[Mesh] AND ("2008/02/27"[PDat] : "2013/02/27"[PDat]). Third, using Microsoft Excel we recorded All MeSH

## FULL PAPER

terms of above papers, and sort and filter the terms, and looked for the high frequency terms (occurrences), and we also counted occurrences of two high frequency terms together in the same paper, setting up the original co-word matrix.

Fourth, the statistical analysis: we made MeSH term's clustering analysis using SPSS13.0 statistical software, draw the co-word network graph of the high frequency terms using Cytoscape software<sup>[3]</sup>.

### THE MESH TERMS ANALYSIS OF PAPERS ABOUT HEPATITIS VIROLOGY

#### The MeSH terms word frequency analysis

We retrieved 21786 papers, every paper has MeSH terms, we extracted MeSH terms and established the MeSH terms database. We got 25 MeSH terms of virology which occurrences frequency was over 677 (including 677). From TABLE 1, we can infered some ideas:the relevant research of hepatitis virology hotspots mainly concentrated in the Hepatitis C and B, Antiviral Agents, etc, it also suggests that Hepatitis C has become hepatitis virology most major research hotspots.

#### Clustering analysis of the high frequency MeSH terms

This research used hierarchical clustering analysis which is one of the most commonly used Classify analysis to analyze the above 25 MeSH terms, drew a dendrogram, and the results were shown in Figure 1.

From the Figure 1, in addition to individual MeSH term as "RNA, Viral; Ribavirin; Hepatitis C, Chronic;", we could seen the other high frequency MeSH terms could be divided into the following five groups. Group 1 contains MeSH terms (Interferon-alpha; Recombinant Proteins; Antiviral Agents; Hepacivirus; Viral Load), it suggests that antiviral drugs such as Interferon alpha are usually made by recombinant protein production technology<sup>[2]</sup>. Group 2 contains MeSH terms (Hepatitis C; Hepatitis C Antibodies; HIV Infections), it suggests that HIV and hepatitis C mostly meanwhile infected people as abusing drugs, so its high proportion of merge infection is very high<sup>[5]</sup>. Group 3 contains MeSH terms (Hepatitis B; Hepatitis B Vaccines; Hepatitis B Antibodies), it suggests that the hepatitis b vac-

TABLE 1 : The top 25 MeSH terms about hepatitis virology

Ranking	MeSH terms	Occurrences
		Frequency (times)
1	virology	7858
2	Hepatitis C	6225
3	Antiviral Agents	5579
4	Hepatitis C, Chronic	5386
5	Hepatitis B	4685
6	Hepacivirus	4638
7	Hepatitis B, Chronic	3508
8	Hepatitis B virus	3344
9	Interferon-alpha	2729
10	Ribavirin	2322
11	HIV Infections (human immunodeficiency virus)	2282
12	Recombinant Proteins	2267
13	RNA, Viral	2102
14	Viral Load	1920
15	DNA, Viral	1857
16	Hepatitis B Surface Antigens	1618
17	Hepatitis B Vaccines	897
18	Hepatitis B e Antigens	824
19	Drug Resistance, Viral	800
20	Hepatitis C Antibodies	790
21	Virus Replication	786
22	Hepatitis B Antibodies	766
23	Lamivudine	762
24	Hepatitis A	743
25	Hepatitis E	677

cine and hepatitis b antibodies are the category of immunotherapy. Group 4 contains MeSH terms (Hepatitis B virus; DNA, Viral; Hepatitis B Surface Antigens; Drug Resistance, Viral; Lamivudine; Hepatitis B, Chronic; Hepatitis B e Antigens; virology; Virus Replication), it suggests that Hepatitis B virus is a DNA virus (including Hepatitis B surface antigen and Hepatitis B e antigen), is easy to delay into chronic hepatitis; Lamivudine as a nucleoside analogues can inhibit virus DNA synthesis and extension, which can terminated viral replication<sup>[1]</sup>. Group 5 contains MeSH terms (Hepatitis A; Hepatitis E), it suggests that infection source and route of transmission on Hepatitis A and hepatitis E

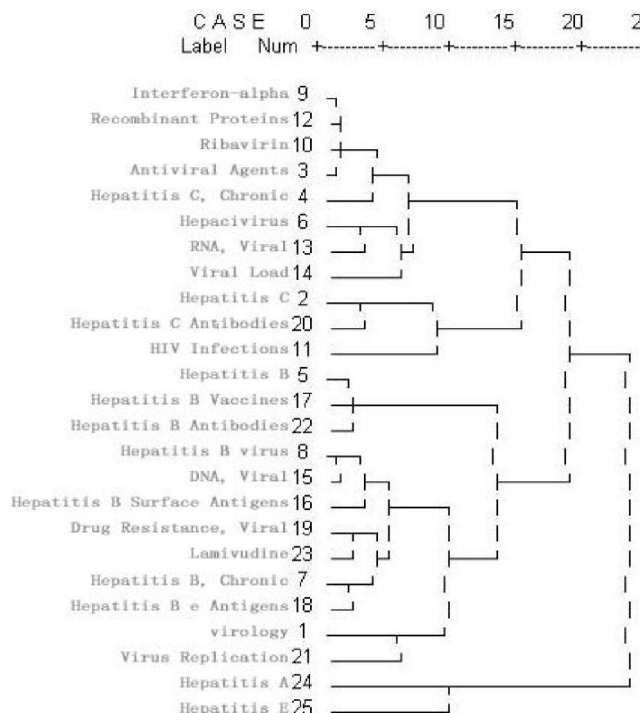


Figure 1 : Hierarchical clustering analysis dendrogram of MeSH terms

is the same, the clinical manifestations of acute hepatitis E is similar to hepatitis A<sup>[7]</sup>.

The above clustering results suggest that several MeSH terms within one group have certain inherent logic connection between each other; If there are no known correlation between the MeSH terms, it indicates we find a new research hotspot. Co-word network graph of the high frequency MeSH terms pair

By analyzing MeSH terms of the top 10 (word fre-

TABLE 2 : The top 9 MeSH terms pair

Ranking	MeSH terms A	MeSH terms B	Co-word occurrences Frequency (times)
1	virology	Hepacivirus	2783
2	Hepatitis C	Hepacivirus	2624
3	Antiviral Agents	Hepatitis C, Chronic	2519
4	Antiviral Agents	Interferon-alpha	2403
5	virology	Hepatitis C	2254
6	virology	Antiviral Agents	2180
7	Antiviral Agents	Ribavirin	2154
8	virology	Hepatitis B virus	2052
9	Hepatitis C, Chronic	Hepacivirus	2018

quency), we got the top 9 MeSH terms pair (A and B, see TABLE 2) and co-word network graph of the MeSH terms pair (see Figure 2). Especially the first MeSH terms pair of virology and Hepacivirus appeared 2783 times in the same paper, it was far higher than that of the second MeSH terms pair (2624 times, Hepatitis C and Hepacivirus).

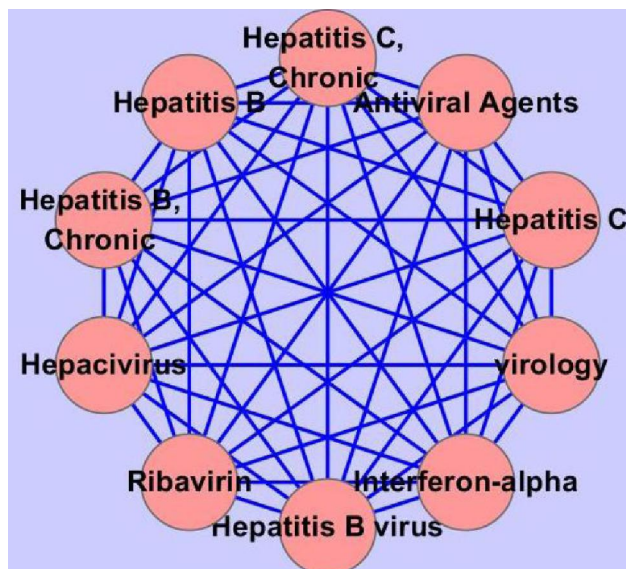


Figure 2 : Co-word network graph of the high frequency MeSH terms pair

In Figure 2 the edge represents the concurrence relationship between MeSH terms pair and if the edge between one MeSH term to other MeSH term, it suggests that the one MeSH term is more important, it is in the center of the research hotspots. So we could infer that Hepatitis C is the research hotspots now.

### CONCLUDING REMARKS

By analyzing MeSH terms (word frequency analysis, clustering analysis, co-word network graph) of PubMed papers about hepatitis, we could infer that the current Hepatitis virology research hotspots had focus on Hepatitis C and B, Antiviral Agents, etc, also the most importance of which was the Hepatitis C.

### REFERENCES

[1] M.S.Chang, S.K.Olsen, E.M.Pichardo et al.; Prevention of De Novo Hepatitis B in Recipients of Core Antibody-Positive Livers With Lamivudine and Other Nucleos (t)ides: A 12-Year Experience.

## FULL PAPER

- Transplantation, **95(7)**, 960-965 (**2013**).
- [2] E.M.Jung, B.S.An, Y.K.Kim et al.; Establishment of transgenic fibroblasts for producing recombinant human interferon- $\alpha$  and erythropoietin in bovine milk. *Mol Med Rep*, **7(2)**, 406-412 (**2013**).
- [3] C.T.Lopes, M.Franz, F.Kazi et al.; Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26(18)**: 2347-2348 (**2010**).
- [4] P.Marcellin; Hepatitis B and hepatitis C in 2009. *Liver Int* 2009; **29**, 1-8 (**2009**).
- [5] M.Sede, N.Laufer, D.Ojeda et al.; Analysis of sequences of hepatitis C virus NS5A genotype 1 in HIV-coinfected patients with a null response to nitazoxanide or peg-interferon plus ribavirin. *Arch Virol*. [Epub ahead of print], (**2013**).
- [6] M.I.Viedma-Del-Jesus, P.Perakakis, M.Á.Muñoz et al.; Sketching the first 45 years of the journal *Psychophysiology* (1964-2008): a co-word-based analysis. *Psychophysiology*, **48(8)**, 1029-1036 (**2011**).
- [7] K.Yano; Other viral food poisoning (hepatitis A and E). *Nihon Rinsho*, **70(8)**: 1386-1390 (**2012**).