

2012 AASRI Conference on Computational Intelligence and Bioinformatics

The Development and Research of Bioinformatics in Neuroscience

Qing Xu^{a,c}, Wuke Zhang^{a,c}, Linfeng Hu^{a,c}, Jiajiong Wang^{b,c}, Jia Jin^{a,c,*}

^a*School of Management, Zhejiang University, Zijingang Campus, Hangzhou 310058, China*

^b*College of Public Administration, Zhejiang University, Zijingang Campus, Hangzhou 310058, China*

^c*Neuromanagement Lab, Zhejiang University, Yuquan Campus, Hangzhou 310027, China*

Abstract

Recent years, the inter-discipline of bioinformatics and neuroscience has been developed with more and more researchers studying in molecular and cellular level, which resulting huge number of publications and findings. A tool of scientometrics, Citespace II was used to identify the hot topic and evolution map of this inter-discipline with the data download from Web of Science. Five research clusters were found with the method of co-citation analysis. The evolution map of knowledge base with 1991-2012 was described.

2012 Published by Elsevier B.V. Selection and/or peer review under responsibility of American Applied Science Research Institute Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: bioinformatics; neuroscience; CiteSpace II;

1. Introduction

Bioinformatics is an interdisciplinary science, which contains the acquisition, processing, storage, distribution, analysis, interpretation and other aspects of biological information. Various tools of mathematics, computer and biology are used in bioinformatics. It has been spanned from the original gene sequence

* Corresponding author. Tel.: +86-13675899085; fax: +86-0571-87952790.
E-mail address: jinjia.163@163.com.

extended to gene function analysis and explained a large number of biological notions. Bioinformatics has become one of the important frontiers of today's life sciences and natural sciences.

Neuroscience appeared in the 1960s and mainly studied the structure and function of nervous system. Research at the molecular and cellular level is an important part of it. With the development of information technology and molecular biotechnology, more and more bioinformatics knowledge, such as gene, are used by neuroscience researchers. The interdisciplinary bring new development for bioinformatics and neuroscience.

Co-citation analysis* of the papers in the cross field of bioinformatics and neuroscience are used with the tool of scientometrics, CiteSpace II, which was developed by Professor Chen from Drexel University. The evolution processes of hot topic and frontier were mapped.

2. Data Source

The data was downloaded from the Web of Science (WOS) database provided by American Institute for Scientific Information (ISI) in March 12, 2012. "Bioinformatics" was used to search the English articles containing this keyword in title, abstract and keywords. The publication time range is 1899 to 2012. In total, 16031 articles were found. Setting "Neuroscience Neurology" as subject area, we get a total of 235 articles and 2852 citation articles. Because we find that the earliest article of 235 articles was published in 1991, the co-citation analysis of the papers starts from 1991 to 2012.

3. Data Analysis

3.1. The main research of bioinformatics in neuroscience

We use CiteSpace II to carry out the visualized analysis on the data. Various CiteSpace II options are selected. These include (a) the time interval of analysis (1991-2012); (b) the unit of analysis (4 years); (c) the threshold value (2, 1, 10), (2, 1, 10), (2, 1, 10). Through the co-citation analysis of literatures, we obtained the network knowledge mapping of co-citation literatures (Fig. 1.), which included 312 nodes and 14552 connecting lines. Each node represents a piece of literature. The color in the center of the nodes represents the time of the literature to be cited for the first time whereas the exterior of the central part represent other time of being cited, corresponding to the color of the time line above the pattern. The radius of the nodes indicates the frequency of the literature to be cited by others. The bigger the radius is, the more frequent the literature is cited. The key nodes are those which surrounded by purple rings, connecting more than two different clusters in most cases. They are the nodes with a high centrality† degree and frequency of being cited. These key nodes may not only transit one time period to another, but also one field to another(Chen 2005).

* Co-citation analysis: if two literatures are simultaneously cited by another literature, these two build the relationship of co-citation. The more the two literatures are co-cited, the tighter their relationship is. Co-citation analysis aims to pick out the different fields of researches by the way of classifying the typical literatures in that subject according to the relationship mentioned above.

† Centrality is an index quantizing the importance of a dot in the network. A high centrality dot is usually located on the path between two different clusters.

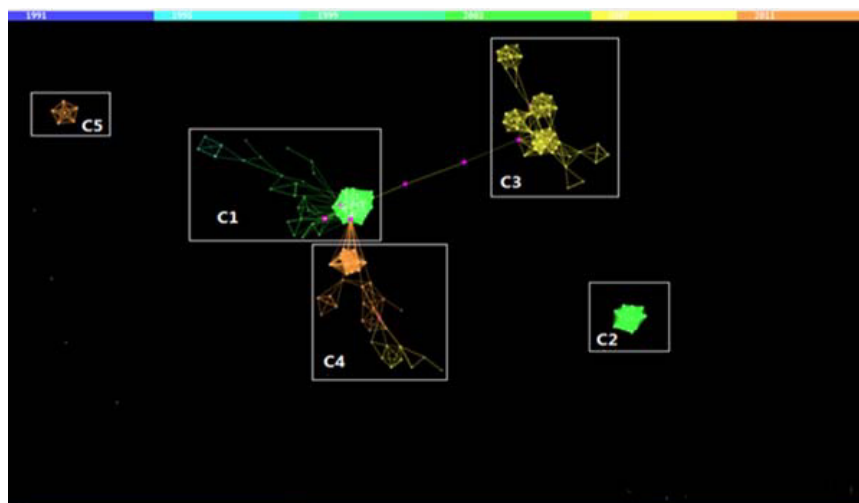


Fig. 1. Network Knowledge Mapping of Co-citation Literature of Bioinformatics in Neuroscience

The network knowledge mapping of co-citation literature composed by these nodes are shown in Fig. 1. According to the property of the key nodes and their space distribution, we divided the network knowledge mapping into 5 clusters and we marked them by C1 to C5 in time sequence, shown in Fig. 1. These five clusters represent the main research direction of Bioinformatics in Neuroscience. After reading and analyzing the typical literature (frequently cited literature) of the five clusters, we summarized the themes of the clusters from C1 to C2 (shown in table 3).

Table 3. Themes of the clusters

Cluster number	Theme of research	Time period of being cited
C1	The research tools and methods of gene expression and sequence	2003~2006
C2	Research on GABAB receptor	2003~2006
C3	The application of the micro array technology in diseases study	2007~2010
C4	The relationship between microRNA and diseases	2011~2012
C5	The research on the codon and transcriptional process	2011~2012

Note: the cited literature refers to the original version of the literature represented by the nodes in Fig. 1.

C1 cluster is mainly about the tools and methods on gene expression and sequencing and the study period focused on the 2003-2006. With the rapid development of Human Genome Project, the data of the nucleic acid, protein sequence and structure is exponential growing which leads to a higher requirement to the tools and methods. As early as 1990, Altschul proposed basic local alignment sequence search tool (BLAST), which optimized the local similarity measures and became a fast sequence comparison methods (Altschul, Gish et al. 1990). Then in 1998, Eisen raised a system of cluster analysis and visualization presents the results. He

analyzed the genome-wide expression data from DNA microarray hybridization and the output was displayed graphically, covering the clustering and the underlying expression data simultaneously in a form intuitive for biologists. In the result, he found that genes of known similar function grouped together and a similar tendency in human data. Thus he proposed that co-expression of genes of known function with poorly characterized or novel genes may provide a simple means of gaining led to the functions of many genes for which information was available currently (Eisen, Spellman et al. 1998). At the same time, with the development of IT, network database arises which, to a certain extent, solves the problem of the dispersion of biological knowledge. For example, Hosack introduced how to get intuitive graphical abstracts from DAVID and use them to make annotations and summaries of identifiers for gene and protein (Dennis Jr, Sherman et al. 2003).

The topic of researches in C2 cluster is GABAB receptors and citing articles are mainly published in 2003-2006. GABAB receptors are widely present in the central and peripheral nervous system and participate in many important physiological activities and pathological changes working with Gi/o protein. And it has great significance to study the function and regulation mechanism of GABAB receptor and reveal the molecular mechanism of the GABAB receptor in pathogenic. Margeta-Mitrovic in 2001 analyzed the function of GB1 and GB2 subunits in G protein coupling of GABAB receptors and found that only with GB2, G protein coupling of GABAB receptors was also successfully completed. This result supported an important model: a single GPCR monomer with a specific G-protein contact is sufficient (to achieve its function) (Margeta-Mitrovic, Jan et al. 2001). As of 2001, there is no direct evidence that GABAB receptors are present in the arthropod. Mezler proposed the assumption of *Drosophila* has the GABAB three subclasses, and based it he found the combined effect of two subclasses is similar to the effect of GABAB subclasses in mammals by cloning and functional expression of the three subclasses of *Drosophila* (Mezler, Müller et al. 2001). It is new indirect evidence for the presentation of GABAB receptors in arthropod body.

The C3 cluster is focusing on the relationship between gene and diseases with the method of microarray technology, and the cited literature was published in the year from 2007 to 2010. In the post-gene era, there is a rapid expansion in the field of bioinformatics, making it an efficient mean to explore the cause of some kinds of disease using the biological computation in which studying the complicated disease from a genetic perspective has become a hot spot. The data mining methods in discovering morbigenous gene, such as the method based on biological network, sequence analysis, gene expression data, are extremely plentiful. Gene expression microarray technology that rising in the post-gene era is the most used. Just as mentioned in the cluster 3, it is applied to diagnose and cure autoimmune encephalomyelitis (Robinson, Fontoura et al. 2003) and gliomas (Rickman, Bobek et al. 2001).

The correlation between genetic expression and disease is discussed in C4 cluster from a microRNA angle, and the cited literature was published in the year from 2011 to 2012. In this area, explore the link between the disease and microRNA is an important direction. In 2005, Lu et al. proposed to classify the cancer type by the microRNA expression profiles. They analyzed the expression of 217 microRNAs of the mammalian from 334 samples comprising some human cancer samples. It is present that microRNA universally declines in tumor tissue compared to normal tissue (Lu, Getz et al. 2005). This finding has potential value for cancer diagnose. In 2009, the quantity and stability of the microRNA in the temporal neocortex of the Alzheimer's disease patients were measured by Sethi et al (Sethi and Lukiw 2009). In the same year, Liu et al tested microRNAs in the spinal cord injury patients. The abnormal expression of the microRNA may contribute to the study of the cause of spinal cord injury and the treatment after the injury (Liu, Wang et al. 2009).

C5 cluster focuses on codon and transcription related field and the cited literature was published in the year from 2011 to 2012. Grantham put forward codon catalog and genome hypothesis that the genome instead of the individual gene is the basic unit in codon selection in 1980 (Grantham, Gautier et al. 1980). Then in 2000, Nakaamura Yusuke find the P53AIP1. This gene has an important impact on cell apoptosis and its

transcriptional activation can be regulated by the phosphorylation ser-46(Oda, Arakawa et al. 2000). Castrochavez compared the common codons of men and other animals(drosophila, squid and so on) and revealed that the third distribution of the most common codons in humans is 12-8-2,CGA,nucleosides ending while 13-6-3,UAG,nucleotide ending in squid(Castro-Chavez 2011).

From the analysis of the clustering, the intersection of bioinformatics and neuroscience appears abundantly in the past 10 years, mainly focusing on five fields: the research tools and methods in gene expression and sequencing study, GABAB receptor research, microarray technology applied in diseases, the relationship of microRNAs and diseases and the codon and transcription ,with a mainly object of gene. The evolutionary relationships between each specifics cluster will be described in detail according to the timeline.

3.2. The development process of Bioinformatics in Neuroscience

Using CiteSpace II to demonstrate the development of all the clusters dynamically, we can find that in the period from 1991 to 2002, the number of Crossover study of bioinformatics and neuroscience is small; From 2003 to 2006, the research about the tools and methods of gene expression and sequence appeared in large number and the literatures form the C1 cluster; Like C1 cluster, the literatures about GABAB receptor form the C2 cluster during the same period and the literatures about the application of the micro array technology in diseases form the C3 cluster in the period from 2007 to 2010. C4 cluster is about the relationship between microRNA and diseases and C5 cluster is about the codon and transcriptional process. Both of these two clusters appeared in the last two years. C1 connects C3 through two key literatures (key nodes). One of the key literatures studies the synchronization in complex networks(Arenas, Díaz-Guilera et al. 2008) and the other studies the structure of material produced by Relativistic Heavy Ion Collider (RHIC) (Adcox, Adler et al. 2005).C1 connects C4 through one key literature which is about how to use the network database DAVID to note and summarize gene or protein identifiers(Dennis Jr, Sherman et al. 2003). Theme of C3 and C4 is the study of principle of the disease from genetic point, and C3 focuses on microarray technology, while the C4 focuses on microRNA. It is easy to see that the researches in C3 and C4 cluster are based on the researches in C1 cluster and the three key literatures provide C3 and C4's researches with theoretical and technical guidance. C2 and C5 can be said to develop independently because they both have no contact with the other clusters(seen in Fig. 1.). However, because the C5 is appearing in the last two years, the researches in C5 are hot topics.

4. Conclusion

Bioinformatics and neuroscience are both important subjects in biological sciences and play a very important role in the development of life sciences and natural sciences. Through the statistical and visual analysis of the research literatures downloaded from the Web of Science database of which the theme is bioinformatics and subject area is "Neuroscience Neurology" with Citespace II, the study has the following conclusions:

First, five research fields are found in the cross-cutting areas of Bioinformatics and Neuroscience. C1's theme is the research tools and methods of gene expression and sequence, C2's theme is research on GABAB receptor, C3's theme is the application of the micro array technology in diseases study, C4's theme is the relationship between microRNA and diseases and C5's theme is the research on the codon and transcriptional process. From the content of the research, we find that bioinformatics research in the field of neuroscience is mainly about the use of gene technology in disease study.

Second, from the development process of the five clusters, we can see that researches of bioinformatics in neuroscience started after 2003. Between 1991 and 2003, there were almost no researches appearing. This

happened because from 1991 to 2003 was the genomic era when the human genome project just began and bioinformatics had been formed, but still in its infancy and the subject knowledge and technology of bioinformatics was not mature. Therefore, few researches in the field of neuroscience appeared. From 2003 is the post-genomic era because the job of sequencing of the human genome project was basically completed. At this time, academia had a higher demand for gene sequencing and gene expression studies and the tools and methods of gene expression and sequencing was improved which not only promoted the development of bioinformatics, but also promoted the bioinformatics researches' development in the field of neuroscience.

References

- [1] Chen C. The centrality of pivotal points in the evolution of scientific networks: ACM; 2005.
- [2] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology* 1990;215:403-10.
- [3] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences* 1998;95:14863.
- [4] Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 2003;4:P3.
- [5] Margeta-Mitrovic M, Jan YN, Jan LY. Function of GB1 and GB2 subunits in G protein coupling of GABAB receptors. *Science's STKE* 2001;98:14649.
- [6] Mezler M, Müller T, Raming K. Cloning and functional expression of GABAB receptors from *Drosophila*. *European Journal of Neuroscience* 2001;13:477-86.
- [7] Robinson WH, Fontoura P, Lee BJ, de Vegvar HEN, Tom J, Pedotti R, DiGennaro CD, Mitchell DJ, Fong D, Ho PPK. Protein microarrays guide tolerizing DNA vaccine treatment of autoimmune encephalomyelitis. *Nature biotechnology* 2003;21:1033-9.
- [8] Rickman DS, Bobek MP, Misek DE, Kuick R, Blaivas M, Kurnit DM, Taylor J, Hanash SM. Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Cancer research* 2001;61:6885.
- [9] Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA. MicroRNA expression profiles classify human cancers. *nature* 2005;435:834-8.
- [10] Sethi P, Lukiw WJ. Micro-RNA abundance and stability in human brain: specific alterations in Alzheimer's disease temporal lobe neocortex. *Neuroscience letters* 2009;459:100-4.
- [11] Liu NK, Wang XF, Lu QB, Xu XM. Altered microRNA expression following traumatic spinal cord injury. *Experimental neurology* 2009;219:424-9.
- [12] Grantham R, Gautier C, Gouy M, Mercier R, Pavé A. Codon catalog usage and the genome hypothesis. *Nucleic acids research* 1980;8:197-.
- [13] Oda K, Arakawa H, Tanaka T, Matsuda K, Tanikawa C, Mori T, Nishimori H, Tamai K, Tokino T, Nakamura Y. p53AIP1, a potential mediator of p53-dependent apoptosis, and its regulation by Ser-46-phosphorylated p53. *Cell* 2000;102:849.
- [14] Castro-Chavez F. Most used codons per amino acid and per genome in the code of man compared to other organisms according to the rotating circular genetic code. *NeuroQuantology* 2011;9.
- [15] Arenas A, Díaz-Guilera A, Kurths J, Moreno Y, Zhou C. Synchronization in complex networks. *Physics Reports* 2008;469:93-153.
- [16] Adcox K, Adler S, Afanasiev S, Aidala C, Ajitanand N, Akiba Y, Al-Jamel A, Alexander J, Amirkas R, Aoki K. Formation of dense partonic matter in relativistic nucleus–nucleus collisions at RHIC: Experimental evaluation by the PHENIX Collaboration. *Nuclear Physics A* 2005;757:184-283.