Engineering, **2013, 5, 530-533**
http://dx.doi.org/10.4236/eng.2013.510B109 Published Online October 2013 (http://www.scirp.org/journal/eng)

Fuzzy Cluster Analysis of Alzheimer's Disease-Related Gene Sequences*

Jing Yang[1#], Jiarui Si[2], Xiaoxuan Gu[1], Ouyan Shi[2#]
[1]School of Public Health, Tianjin Medical University, Tianjin, China
[2]School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China
Email: #yangjing@tijmu.edu.cn, #shiouy@tijmu.edu.cn

Received 2013

ABSTRACT

The objective of this paper is to analyze the relationship among the interrelated gene sequences of Alzheimer's disease (AD). Further this paper will provide a study on genetic factor of the occurrence about Alzheimer's disease, so as to provide more information on the prevention of Alzheimer's disease, the clinical diagnosis and gene therapy for Alzheimer's disease. The respective alignment of the Alzheimer's disease interrelated gene sequences with those in The National Center for Biotechnology Information (NCBI) database was studied, and the measurement relationship of these sequences was identified and analyzed by the method of fuzzy cluster. The result of fuzzy cluster analysis indicates that the gene sequences interrelated within one group is consistently having closer relationship within the group other than in another group.

Keywords: Alzheimer's Disease; Gene; mRNA; Sequence; Alignment; Fuzzy Cluster

1. Introduction

Alzheimer's disease is a degenerative neurological disorder characterized by neural loss and brain lesions associated with plaques containing large amounts of the beta/A4 amyloid peptide. It has been identified in an emerging multigene family. The more old people aged above 65 there are in China, the more old people have Alzheimer's disease. The cause of this disease is still not fully clear.

Mutations in presenilins are responsible for approximately 40% of all early-onset familial Alzheimer disease (FAD) cases in which a genetic cause has been identified [1]. Missense mutations in the genes encoding amyloid precursor protein (APP), presenilin-1 and presenilin-2 have been found to cause some forms of autosomal dominant early-onset Alzheimer disease. Autosomal dominant point mutations in the APP gene are associated with beta-amyloid peptide-related cerebral amyloid angiopathy and Alzheimer's disease [2]. In general, there are more and more the results to drawn studier's attention to genetic factor of the Alzheimer's disease occurrence.

This paper is to study more genetic factor of the Alzheimer's disease occurrence and provide information about prevention, diagnosis and treatment in genetic level. The method is that fuzzy cluster analysis divides the data of Alzheimer's disease-related gene sequences into groups such that the similar data objects belong to the same cluster and the dissimilar data objects to different clusters. The fuzzy clustering method is based on the measure of distance. This measure is the score that respective pair wise sequence alignment of interrelated gene sequences with Alzheimer's disease in NCBI database. Fuzzy cluster analysis can yield useful information on the intrinsic characters or property of this data. The measurement relationship of these sequences was identified and analyzed by the fuzzy cluster.

In this paper, we aim to provide the results of clusters (or groups) that are the Alzheimer's disease associated gene sequences by different number of α-cuts (thresholds). From fuzzy cluster point of view, the gene sequences in one group have been consisted to have close relationship and similar functions and characters. This may possibly be made use of information reference in clinical diagnosis and treatment of the Alzheimer's disease by the results of fuzzy cluster.

2. Materials and Methods

2.1. The Dataset

Through searching for keyword "Alzheimers" in Nucleotide database of The National Center for Biotechnology Information (NCBI), there were 4 sequences to be found

*This work was supported by grant (30870791) from the National Natural Science Foundation of China and grant (2011KZ87) from the Scientific Research Foundation of Tianjin Bureau of Public Health.
#Corresponding authors.

Copyright © 2013 SciRes.

on March 2007 and 8 sequences to be found on May 2011. The 12 mRNA sequences were selected which interrelated with Alzheimer's disease and applied into sequences alignment analysis. The sequences numbers of identify (accession numbers in Nucleotide database) are nm_005166; nm_001642; nm_001024807; nm_002704; nm_023959;nm_080478; nm_017522; nm_033300; nm_001018054; nm_004631; nm_001013018; nm_178003. The set of all mRNA sequences' numbers is set M, *i.e.* M = {nm_005166; nm_001642; nm_001024807; nm_002704; nm_023959; nm_080478; nm_017522; nm_033300; nm_001018054; nm_004631; nm_001013018; nm_178003 }.

2.2. The Data Score Matrix

Let the data set M be defined as "X". The set X consists of 12 data points: X={x1, x2 ⋯ x12}, x1= nm_005166; x2 = nm_001642; ⋯; x12 = nm_178003.

We performed the operation of the Smith-Waterman algorithm is a member of the class of algorithms that can calculate the best score on this sequence set X, by using "EMBOSS Pairwise Alignment Algorithms" software (from: http://www.ebi.ac.uk/emboss/align/index.html/).

We are trying to find the best region of similarity between two sequences, use the "Water" program. "Water" program uses the Smith-Waterman algorithm (modified for speed enhancements) to calculate the local alignment. "Water" program finds an alignment with the maximum possible score where the score of an alignment is equal to the sum of the matches taken from the scoring matrix.

There is a scoring matrix X of the sequences' pair wise alignment to be shown in the **Figure 1**.

2.3. The Fuzzy Clustering

In the fuzzy cluster the degree of belonging to a cluster is quantified by means of membership function r(x, y). The value of r(x, y) is belonged in the interval [0, 1].

The fuzzy relation matrix R (to be shown in **Figure 2**) was obtained by method of "Maxim and Minim" [3] for above the scoring matrix X, as in (1).

$$R = (r_{ij}), \quad (i, j = 1,2,\cdots,n),$$

$$r_{ij} = r(x_i, x_j) = \frac{\sum_{k=1}^{n} \wedge (x_{ik}, x_{jk})}{\sum_{k=1}^{n} \vee (x_{ik}, x_{jk})}.$$

The symbol "$\wedge$" is defined by (2) and The symbol "$\vee$" is defined by (3).

$$\wedge (x_i, x_j) = \min(x_i, x_j)$$

$$\vee (x_i, x_j) = \max(x_i, x_j)$$

From the fuzzy relation matrix R (**Figure 2**), one can see that:

The relation matrix R is a reflexive matrix.

The relation matrix R is a symmetric matrix.

The relation matrix R is not a max-min transitive matrix.

A fuzzy binary relation matrix that is reflexive, symmetric and transitive is known as a fuzzy equivalence relation matrix. A binary relation matrix that is reflexive and symmetric is called a compatibility relation matrix. The fuzzy relation matrix R (**Figure 2**) is a fuzzy compatibility relation matrix and not an equivalence relation matrix. So that a transitive closure of the relation matrix R is necessary for the fuzzy cluster. Transitive closure of the relation matrix R can be obtained by (4) and (5).

$$R^2 = R \circ R = (\bigvee_{k=1}^{n} [x_{ik} \wedge x_{jk}])$$

$$R^4 = R^2 \circ R^2 \xrightarrow{till} R^{2k} = R^k$$

One can have fuzzy cluster analysis for a fuzzy equivalent matrix. For the fuzzy relation matrix R (**Figure 2**), there is a fuzzy equivalent matrix R^6 (to be shown in **Figure 3**) by the aid of above step (4) and step (5). The fuzzy equivalent matrix R^6 (**Figure 3**) can be used in fuzzy clustering.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
x_1	12235											
x_2	3223	19135										
x_3	12224	3221	1225									
x_4	885	1257	884	6535								
x_5	2479	2895	2487	1062	16695							
x_6	2259	2552	2261	1062	2597	12770						
x_7	2590	3812	2591	1430	3162	2500	34970					
x_8	2608	3791	2614	1426	3162	2519	34064	36365				
x_9	2698	3780	2696	1426	3263	2527	34645	35117	38030			
x_{10}	2752	3772	2749	1426	3332	2531	34547	36100	37932	38915		
x_{11}	2588	4290	2592	1135	2710	2103	2553	2365	2395	2379	11656	
x_{12}	2301	2515	2310	908	2603	2454	2522	2533	2540	2582	1983	13190

Figure 1. The score matrix X of the sequences' pairwise alignment.

Copyright © 2013 SciRes.

$R = (r_{ij})_{12*12}$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
x_1	1											
x_2	0.43	1										
x_3	0.78	0.46	1									
x_4	0.25	0.46	0.32	1								
x_5	0.42	0.47	0.47	0.39	1							
x_6	0.48	0.43	0.52	0.51	0.47	1						
x_7	0.17	0.20	0.17	0.09	0.21	0.16	1					
x_8	0.17	0.20	0.16	0.09	0.20	0.16	0.97	1				
x_9	0.16	0.19	0.16	0.08	0.20	0.16	0.95	0.96	1			
x_{10}	0.16	0.19	0.16	0.11	0.20	0.16	0.94	0.96	0.99	1		
x_{11}	0.47	0.51	0.54	0.32	0.47	0.53	0.17	0.17	0.16	0.16	1	
x_{12}	0.15	0.15	0.15	0.08	0.16	0.16	0.10	0.10	0.09	0.09	0.16	1

Figure 2. The fuzzy relation matrix R.

$R^6 =$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
x_1	1											
x_2	0.51	1										
x_3	0.78	0.51	1									
x_4	0.51	0.51	0.51	1								
x_5	0.47	0.47	0.47	0.47	1							
x_6	0.53	0.51	0.53	0.51	0.47	1						
x_7	0.21	0.21	0.21	0.21	0.21	0.21	1					
x_8	0.21	0.21	0.21	0.21	0.21	0.21	0.97	1				
x_9	0.21	0.21	0.21	0.21	0.21	0.21	0.96	0.96	1			
x_{10}	0.21	0.21	0.21	0.21	0.21	0.21	0.96	0.96	0.99	1		
x_{11}	0.54	0.51	0.54	0.51	0.47	0.53	0.21	0.21	0.21	0.21	1	
x_{12}	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	1

Figure 3. The fuzzy equivalent matrix R^6.

3. Conclusions

We can have fuzzy cluster analysis for the fuzzy equivalent matrix R^6 (**Figure 3**) and have the fuzzy cluster graph (to be shown in **Figure 4**) by different number of alpha-cuts (thresholds).

By the fuzzy cluster analysis, some sequences are interrelated with one group and the sequences in one group have been consisted to have close relationship and functions. To choose the appropriate alpha-cuts, the fuzzy cluster graph can be applied to analyze the problem. For instance, in the case alpha = 0.50, the sequence set X = (x1, x2 ⋯ x12) is divided into 5 groups: {x1, x3, x6, x11}; {x2, x4}; {x7, x8, x9, x10}; {x5}; {x12}.

In the group {x1, x3, x6, x11}, the x1 = nm_005166 is a sequence in connection with the APLP1 gene. Amyloid-precursor-like protein 1 (APLP1) is a membrane-associated glycoprotein, whose gene is homologous to the APP gene, which has been shown to be involved in the pathogenesis of Alzheimer's disease [4]. That has been implicated on genetic factor of the Alzheimer's disease occurrence. The other correlated sequences in the same group may have the same function or characteristic.

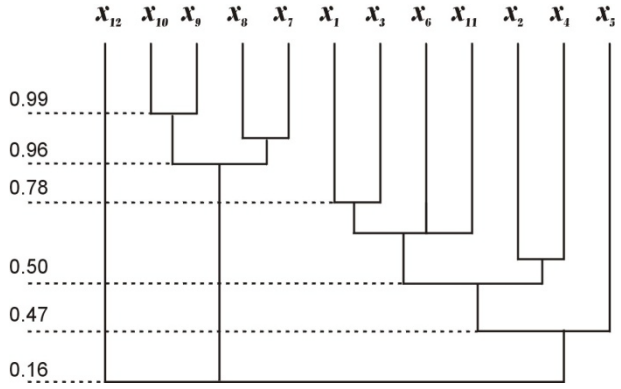

Figure 4. The fuzzy cluster graph by different number of *α*-cuts.

For the point of view of fuzzy cluster, the gene sequences interrelated within one group may be consisted

Copyright © 2013 SciRes.

to have closer relationship and similar functions and characters. The results should be applicable referenced information in the study of Alzheimer's disease. Some results of analysis by fuzzy cluster about gene sequences of the Alzheimer's disease should be confirmed by medicine before it was used to Alzheimer's disease prevention, clinical diagnosis and treatment.

A large amount of gene sequences has already been collected and deposited in public databases and these are important resources not only for use as markers to identify disease-relates genes, but also to provide useful information to understand the disease mechanisms in genome level. Study on interrelated gene sequences of the Alzheimer's disease by fuzzy cluster is one of the topics on gene research. By way of the disease interrelated gene sequences exploration, we hope our explanation will prove more or less helpful in the Alzheimer's disease prevention, clinical diagnosis and treatment.

REFERENCES

[1] A. Rovelet-Lecrux, T. Frebourg, H. Tuominen, K. Majamaa, D. Campion and A. M. Remes, "APP Locus Duplication in a Finnish Family with Dementia and Intracerebral Haemorrhage," *Journal of Neurology, Neurosurgery & Psychiatry*, Vol. 78, 2007, p. 1158.

[2] O. Nelson, H. Tu, T. Lei, M. Bentahir, B. Strooper and I. Bezprozvanny, "Familial Alzheimer Disease-Linked Mutations Specifically Disrupt Ca^{2+} Leak Function of Presenilin 1," *The Journal of Clinical Investigation*, Vol. 117, 2007, pp. 1230-1239.

[3] B. Q. Hu, "The Elements Theory of Muzzy Mathematics," Wu Han University Press, 2004, pp. 153-158.

[4] B. Weber, C. Schaper, J. Scholz, B. Bein, C. Rodde and P. H. Tonner, "Interaction of the Amyloid Precursor like Protein 1 with the Alpha2a-Adrenergic Receptor Increases Agonist-Mediated Inhibition of Adenylate Cyclase," *Cell Signal*, Vol. 18, 2006, pp. 1748-1757.

Copyright © 2013 SciRes.