Original Article

Protein Clustering and Interactome Analysis in Parkinson and Alzheimer's Diseases

Mostafa Rezaei-Tavirani PhD[1], Mona Zamanian-Azodi PhD·[2], Sepideh Rajabi MSc[1], Ali Masoudi-Nejad PhD[3], Mohammad Rostami-Nejad PhD[4], Sara Rahmatirad MSc[5]

Abstract

Background: Alzheimer and Parkinson diseases (AD and PD) are the two most important neurodegenerative disorders. This paper aims to determine the possible molecular linkage between these two common neurodegenerative diseases by a combination of computational investigations.

Methods: According to our aim, common sets of identified proteins from the KEGG database were further analyzed based on Gene Ontology (GO) annotation and sequence similarities by the agglomerative hierarchical clustering. Proteins possessing same characteristics were categorized based on biological features in distinct clusters using the R programming software. In addition to this, by the use of DAVID Program and PPI network analysis, more insight can be achieved.

Results: The results of this study indicated that 23 proteins are common between these two diseases. Their ontology evaluations by application of clustering methods showed that proteins belonging to a specific cluster indicate discrete properties that are different from other clusters. Furthermore, PPI network analysis confirms that the proteins with similarity ontology and sequence are also in close relationship.

Conclusion: In conclusion, assessment of protein features supported the idea that mitochondria are the main malfunction compartment in AD and PD. Some of these common properties are apoptosis and mitochondria oxidation pathways that can be used for drug targeting. Moreover, examination of other neurodegenerative diseases can be helpful for comprehensive understanding of the origin of these diseases.

Keywords: Alzheimer's disease, clustering, Cytoscape v: 3.2.1., DAVID bioinformatics resources, Parkinson's disease, R Programming Language Software

Cite this article as: Rezaei-Tavirani M, Zamanian-Azodi M, Rajabi S, Masoudi-Nejad A, Rostami-Nejad M, Rahmatirad S. Protein Clustering and Interactome Analysis in Parkinson and Alzheimer's Diseases. *Arch Iran Med*. 2016; 19(2): 101 – 109.

Introduction

Rapid progress in high-throughput (HT) biotechnologies datasets have led to the development of vast databases and their data-mining tools in recent years.[1] In fact, these large biological data are often overwhelming and challenging as they contain much more information than an investigator has generated. Therefore, advanced computational and statistical tools are required for big data analysis.[2] Large-scale studies have shown that a discrete biological function can rarely be attributed to an individual molecule. Indeed, complex interactions among numerous cellular constituents such as proteins, DNA, RNA, and small molecules manifest many biological characteristics.[3,4] One major piece of information pertains to proteins. Most of drugs act on proteins, as they can tie to pathogenesis of a disease by regular dysfunctions such as protein aggregation in Alzheimer and Parkinson diseases (AD and PD).[5,6] Consequently, increasingly proteins themselves are the main targets in medical studies.[7] Gene ontology as a part of a larger classification effort, the Open Biomedical Ontologies, is promising systematic unification of gene and gene products annotation. YPERLINK \o ", "HYPERLINK \o ", " It seeks to supply information based on three notable features of protein molecules: cellular component (CC), biological process (BP), and molecular function (MF).[8] Clustering is the method to analysis different sets of correlated proteins in a specific profile of a disease for underpinning the complexity. In fact, data objects are grouped into clusters where objects in the same cluster show stronger similarity than in other sets.[9,10] Clustering methods are usually categorized to hierarchical and nonhierarchical clustering techniques.[11] These methods have a wide range of clustering approaches. In addition, each protein in a profile exhibits high interaction with other elements of a complex.[12] Normally, most relevant proteins in a complex are fit in the same functional category.[13,14] The interaction between two or more proteins can result in distinct functional objectives that can be demonstrated in numerous different ways.[15,16] Therefore, for better deciphering the functional association of the disease proteome, protein map analysis is crucial to establish high sensitive algorithms to evaluate these high-throughput data.[2] In this study, two common neurodegenerative diseases are chosen for further biological evaluation. One is Alzheimer's disease (AD) as the most frequent type of progressive dementia in the world. Its onset is normally in the elderly population.[17] The symptoms consist of decline in memory,

Authors' affiliations: [1]Proteomics Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [2]Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [3]Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran. [4]Research Institute for Gastroenterology and Liver Diseases, Gastroenterology and Liver Diseases Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [5]Department of Cell and Molecular Biology, Faculty of Science, University of Tehran, Tehran, Iran.

·**Correspondence author and reprints:** Mona Zamanian-Azodi PhD, Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Tel: 00982122714248, Fax: 00982122714248, E-mail: mona.azodi@gmail.com

Accepted for publication: 4 November 2015

thinking and reasoning skills.[18] Another widespread neurodegenerative disease ranking second after AD, is Parkinson's disease (PD).[19] It is identified as a motor disease with no gender preference and unknown etiology.[20] Clinical symptoms related to this disease are resting tremors of extremities, muscular rigidity, shuffling gait, stoop posture, and bradykinesia.[21] In the present work, the possible relationship between AD and PD diseases is studied via cluster analysis. This can lead to a better understanding of the molecular basis of the two diseases. In addition, by identifying common proteins, it is possible to introduce common drug targets for AD and PD diseases.

Materials and Methods

Proteins and their related pathways to Alzheimer's disease and Parkinson's disease were identified and analyzed. Proteins can be analyzed based on different aspects such as structural, sequential, and gene ontology.[14] Raw data was acquired from the KEGG database and the protein map of the two diseases was compared. KEGG is a useful up-to-date database for deciphering various proteins information. These include map and other studies such as systematic analysis of gene functions, linking genomic information with higher order functional information. It is also provides protein and nucleotide sequence of each genes.[22] Common proteins were selected and then their human UniProt codes were retrieved for further analysis. Using a scoring system and the Gene Ontology database, these common proteins were examined for their ontology features. The Uniprot codes were applied for GO search. GO is the reference of controlled vocabulary of the terms of three biological annotations including cellular component, molecular function, and biological process.[23] Furthermore, the relevant information was downloaded from the database, and similarity between proteins was studied by SimUl graphical similarity comparison method.[24] The number of common nodes divided by the number of nodes in the combination of the two graphs are equalized by simUl, hence the similarity was calculated between 0 and 1.[25] The obtained scores were used to provide distance dissimilarity. Dissimilarity = 1-similarity as a dissimilarity matrix.[26] On the other hand, protein sequences were extracted from KEGG databases in FASTA format. Then, global alignment for sequence comparison was carried out using Needleman-Wunsch algorithm, available in EBI Database. This algorithm is a useful method for analyzing similarities between protein sequences.[27] Here, this algorithm is applied for pair-wise alignment between each set of two proteins. Similarity scores were obtained by comparing each of these 23 proteins and the relevant dissimilarity matrix was constructed. The constructed matrices were used for clustering analysis. Clustering can be useful for a thorough understanding of the annotation of protein molecules and their relationship to many diseases.[28] For clustering studies, resemblance is the key feature.[14] Here, agglomerative hierarchical clustering analysis (bottom up) was conducted. This method can provide good efficiency proven by many studies.[11] Interpretation and graphical illustrations of clusters of data were performed with the R software. [R software is a free statistical software available on (www.r-project.org)]. It offers various packages, one of which is the package cluster used to identify subgraphs with maximal density. As mentioned before, hierarchical algorithms were the applied clustering methods using AGNES function. In order to perform AGNES, the dissimilarity matrix is needed. Dendrogram and silhouette plots are the graphical representations of the clustering in this study. The data was analyzed by correlation distance mode. The distance between each cluster can be calculated as follows: D = 1-C, where D = distance and C = correlation between spot clusters. Average linkage (UPGMA) is the method used to define the distance between clusters as the average distance.[11] In AGNES method, dendrogram representation can be helpful to realize the hierarchical order of protein clusters. In fact, it is useful for retrieving information related to identifying which individual divisions are the most similar to each other by reading from bottom-up order. Based on agglomerative method, another graphical representation is the so-called 'silhouette' that illustrates a concise graphical representation of how well each object related to its cluster. This type of graphical display was first described by Peter J. Rousseeuw in 1986 to show which objects lie well within their cluster, and which ones are simply someplace in between clusters.[29] In addition, the strength and quality of clustering structures of the dataset were measured by agglomerative coefficient in AGNES function. The calculating range is between 0 and 1. The AC close to 1 implies well-structured cluster, while AC close to 0 indicates not a powerful clustering structure.[11] Furthermore, the annotation terms based on (GO) features was obtained using DAVID bioinformatics resources (http://david.abcc.ncifcrf.gov.). Annotation cluster analysis by DAVID program can be helpful in retrieving related information based on different annotation aspects such as (GO) function category.[30] For each protein, Uniprot Accession Number was used as cross-mapped to the DAVID knowledgebase and gene identifier. Count number indicates the number of annotated proteins for each mentioned term in DAVID output. Each cluster implies group of terms possessing similar biological meaning.[31]

Cytoscape v:3.2.1 also was used for the illustration of predicted interactions of the common proteins. This tool is an open source that provides a resourceful environment for data visualization in the form of complex networks available on (www.cytoscape.org).[32]

Results

The KEGG database is the source for basic examination in this study. The raw data was extracted from this database. Each protein related to Alzheimer's disease and Parkinson's disease was identified and common proteins between them were extracted. Twenty-three proteins were found as common proteins between these two diseases (Table 1).

To realize how gene products behave in cellular context, the application of GO database is helpful. Gene annotation of 23 proteins were extracted from this database and compared via SimUl scoring methods. Therefore, the distance between pairs of proteins was calculated as similarity graph based on Gene Ontology terms. These measured distances were then applied for protein clustering as annotation similarity relation. As mentioned before, in this method, common nodes of two GO graphs divided by summation of nodes of two graphs is the similarity numbers, which is between 0 and 1. The associated dissimilarity is 1 – similarity. Therefore, the lower the scores, the better for sharing within the same cluster (See supplementary Tables 1, 2, and 3).

The degree of sequence similarity between proteins is presented as a constructed dissimilarity matrix. Numbers are between 0 and 1 (See supplementary Table 4).

In this study, clustering method was performed based on gene ontology graphs and sequence comparison. Dendrogram cluster-

Table 1. Common proteins between Alzheimer's disease and Parkinson's disease retrieved from KEGG database.

Row	Protein Name
1	Casp3: Caspase 3, apoptosis-related cysteine peptidase
2	Casp9 :Caspase 9, apoptosis-related cysteine peptidase
3	Cox7c: Cytochrome c oxidase subunit VIIc
4	Cox1:Cytochrome c oxidase subunit I
5	Cox8A : Cytochrome c oxidase subunit 8A (ubiquitous)
6	COX7A1: Cytochrome c oxidase subunit VIIa polypeptide 1
7	COX6B1: Cytochrome c oxidase subunit Vib polypeptide 1
8	COX6A1: Cytochrome c oxidase subunit VIa polypeptide 1
9	SDHA : Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial
10	ATP8 :ATP synthase protein 8
11	CYCS : Cytochrome c
12	APAF1 : Apoptotic protease-activating factor 1
13	ATP5H : ATP synthase subunit d, mitochondrial
14	CYC1 : Cytochrome c1, heme protein, mitochondrial
15	UQCR : Cytochrome b-c1 complex subunit 10
16	ATP5C1 : ATP synthase subunit gamma, mitochondrial
17	ATP5F1 : ATP synthase lipid-binding protein, mitochondrial
18	ATP5G1 : ATP synthase lipid-binding protein, mitochondrial
19	ATP5J : ATP synthase-coupling factor 6, mitochondrial
20	ATP5G2 : ATP synthase lipid-binding protein, mitochondrial
21	CYTB : Cytochrome b
22	ATP5A1: ATP synthase subunit alpha, mitochondrial
23	ATP6: ATP synthase F0 subunit 6

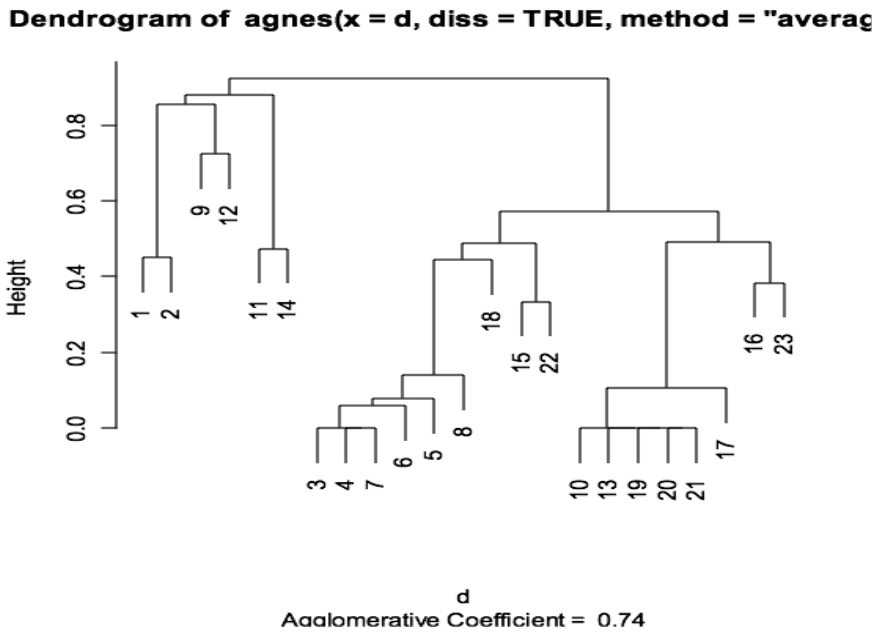

Figure 1. Dendrogram clustering plot based on biological process, the agglomerative coefficient (AC) is calculated 0.74.

ing plot of biological process features are presented in Figure 1. Based on the results from hierarchical agglomerative clustering, proteins are categorized from the most specific features to the most common. In fact, this type of hierarchical clustering starts from small clusters composed by single objects to greater ones, in such a way that, the height implies on distance relation between clusters. Moreover, the height of a node refers to the distance between sub-branch clusters. The strength of clustering agglomerative coefficients structured natural as it is estimated 0.74.

Silhouette plot is useful for clustering data illustration. It is presented based on biological processes. Agglomerative coefficient is 0.74 and indicates good clustering structure (See Figure 2).

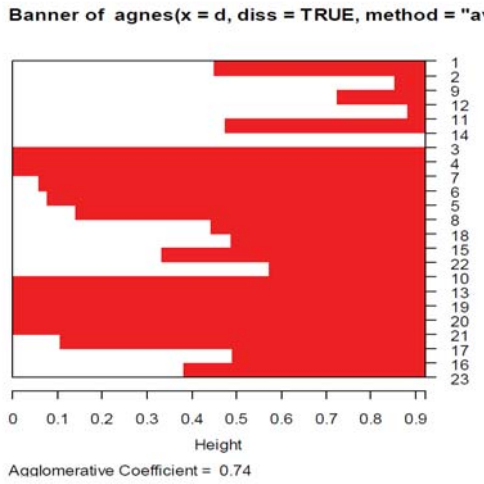

Figure 2. Silhouette plot based on biological process, the agglomerative coefficient (AC) is 0.74 for this analysis.

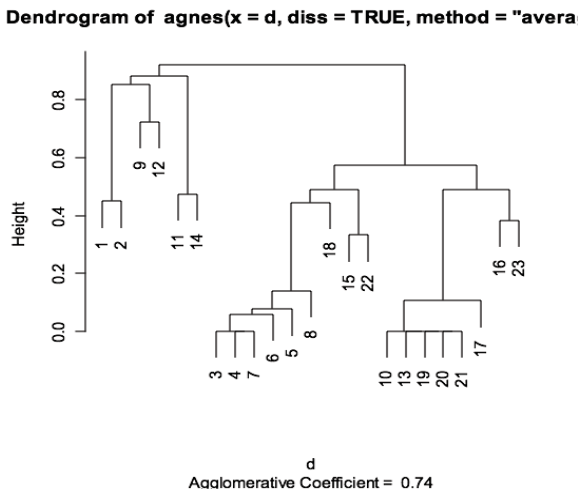

Figure 3. Dendrogram clustering plot molecular function the agglomerative coefficient is 0.74.

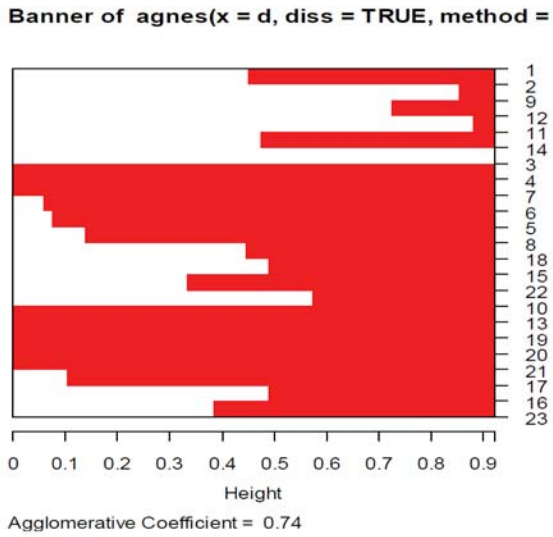

Figure 4. Silhouette plot molecular function, the agglomerative coefficient (AC) is 0.74 for this analysis.

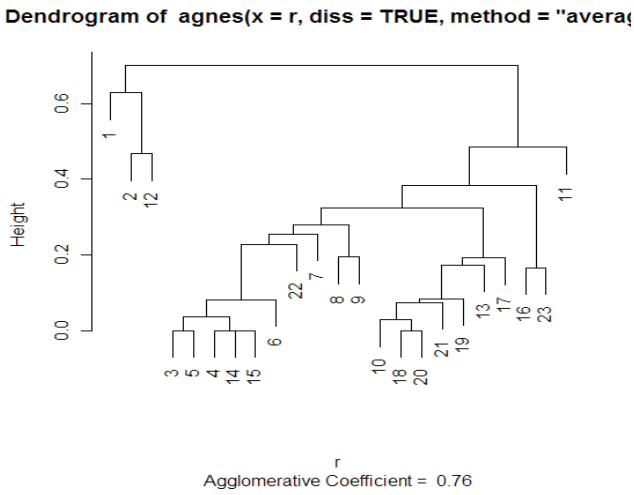

Figure 5. Dendrogram clustering plot presentation of cell component, the agglomerative coefficient (AC) is 0.76.

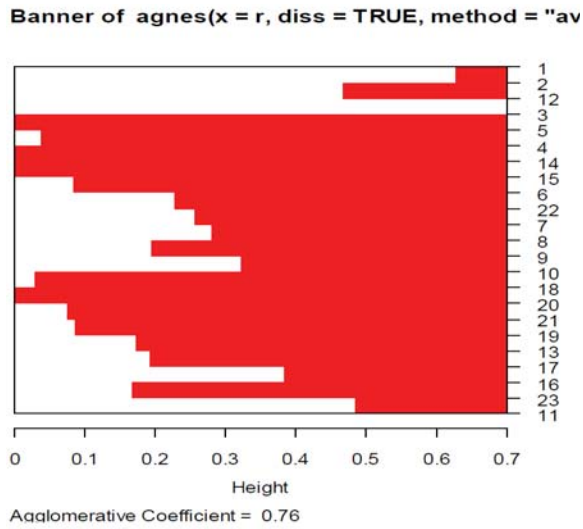

Figure 6. Silhouette plot presentation of cell compartment, the agglomerative coefficient (AC) is 0.76

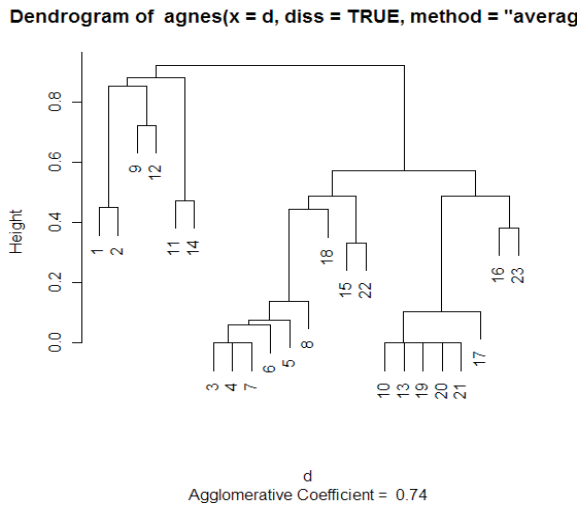

Figure 7. Dendrogram clustering plot of global pair-wise alignment between each pair sets of investigated proteins of the diseases. The agglomerative coefficient (AC) is 0.74 for this analysis.

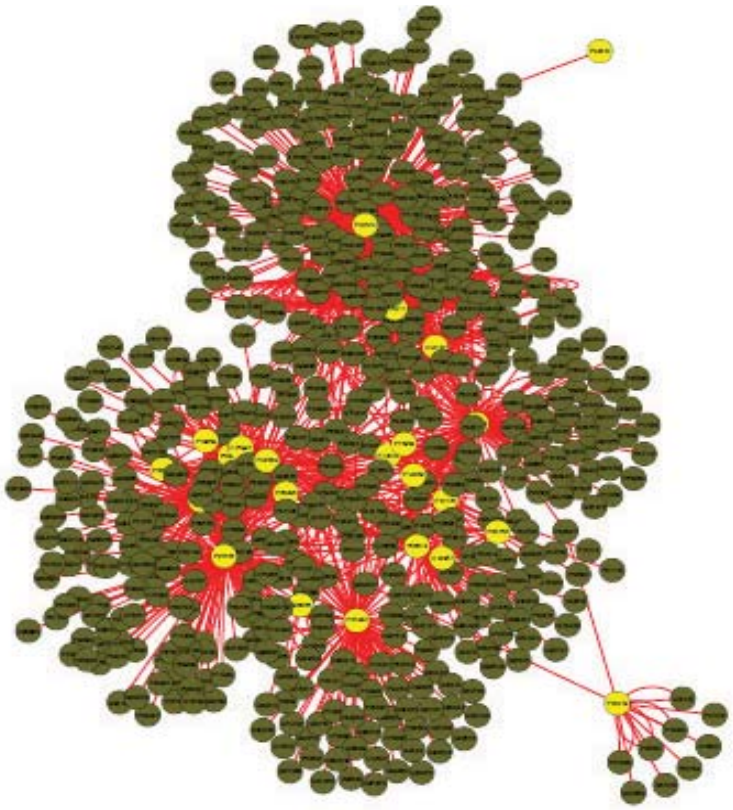

Figure 8. Protein-protein interaction network analysis of 23 proteins using Cytoscape v.3.2.1 software derived from Mentha, Reactome-Fls public databases. This PPI network consists of 600 nodes and 1469 edges. The designate proteins are shown as highlighted nodes. The important hub proteins are CASP3 with 218 and 559 edges, ATP51 with 123 nodes and 317 edges, CYCS with 97 nodes and 420 edges, CASP9 with 57 nodes and 350 edges and APAF1 with 40 nodes and 282 edges.

Above clustering method is also performed for (MF) and (CC) with good clustering structure (AC= 0.74, AC= 0.76, respectively) (See Figures 3 – 6)

Protein alignment shows novel data that can increase related annotations. Similarity is the main concept for sequence evaluation. Pair-wise sequence alignment was performed among pairs of these 23 common proteins, and significant result is attained as a dendrogram cluster plot (Figure 7). Agglomerative coefficient is 0.74 and indicates good clustering structure. As the agglomerative coefficient is high for each analysis, AGNES provided high quality clustered dataset in this study. In other words, the clustering algorithm is valid for this study. This can be seen from dendrogram.

As indicated in classification Tables 2, 3, and 4, for each annotation category of interest, the smaller the *P* value, the more significantly proteins belong to that related enriched term. As it is clear, mitochondria and their main functions and processes are highly related to the studied common genes.

In Figure 8, the interactions of 23 proteins with their neighbors are presented. The integrated network was obtained from Mentha, Reactome-Fls Databases by the application of PSICQUIC source in Cytoscape software. Furthermore, topological parameter analysis (degree centrality) of the network by the use of Network Analyzer, presents the important key proteins in the network.

Discussion

In large-scale approaches, there is an exclusive requirement for analyzing vast amount of data.[6] Clustering is one of the useful methods for this purpose. Proteins play a fundamental role in different diseases. As several functions are linked to proteins,[6] bioinformatics analysis is needed for clarification. Studies showed that most of the neurodegenerative diseases have malfunction in mitochondria compartment, as a key role in regulation of energy production of a cell.[7] For this reason, further evaluation is needed for deciphering Gene Ontology of these diseases. It is proven that energy production pathway has fundamental functions as the key molecular process in the cell and acts as a strong link between Alzheimer's disease and Parkinson's disease.[33] Cluster analysis, as an appropriate statistical procedure, can classify proteins based on distinct similarities.[34] Here related proteins are subdivided in groups by the aid of this method. In the first step, common proteins in Alzheimer's disease and Parkinson's disease pathways were analyzed. As shown in Table 1, a number of 23 proteins were detected as common molecules of these two diseases based on protein profiles comparison. By analyzing these 23 proteins, it is possible to identify common pathways and ontology properties. It seems that these findings can be used as a relatively common procedure for the management of these two diseases. Examining each protein one by one can be a taught assignment, while by application of agglomerative clustering analysis of the whole protein, considerable results may be achieved. Considering Figures 1 to 4 based on (MF) and (BP), clustering results are similar, but in comparison with (CC) (figures 5-6), there are some differences. However, protein numbers 1, 2, 12 including Casp3, Casp9, and APAF1 and protein numbers 3, 4, 5, 6 (Cox7c, Cox1, Cox8A, Cox7A1) are classified in the distinct clusters in all the GO analy-

Table 2. Constructed matrix of dissimilarity between 23 proteins of table 1 based on biological process (gene ontology).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0.00	0.88	0.98	0.96	0.97	0.97	0.98	0.97	0.94	0.95	0.82	0.77	0.95	0.97	0.96	0.95	0.95	0.95	0.95	0.95	0.90	0.90	0.93
2	0.88	0.00	0.91	0.94	0.91	0.91	0.94	0.91	0.93	0.95	0.46	0.64	0.95	0.92	0.94	0.95	0.95	0.94	0.95	0.94	0.97	0.93	0.95
3	0.98	0.91	0.00	0.80	0.20	0.20	0.60	0.20	0.88	0.92	0.92	0.96	0.92	0.60	0.80	0.92	0.92	0.91	0.92	0.91	0.94	0.95	0.92
4	0.96	0.94	0.80	0.00	0.76	0.76	0.85	0.76	0.66	0.80	0.81	0.96	0.80	0.52	0.13	0.80	0.80	0.78	0.82	0.78	0.75	0.87	0.81
5	0.97	0.91	0.20	0.76	0.00	0.01	0.66	0.01	0.85	0.90	0.91	0.95	0.90	0.50	0.75	0.90	0.90	0.89	0.90	0.89	0.93	0.94	0.91
6	0.97	0.91	0.20	0.76	0.01	0.00	0.66	0.01	0.85	0.90	0.91	0.95	0.90	0.50	0.75	0.90	0.90	0.89	0.90	0.89	0.93	0.94	0.91
7	0.98	0.94	0.60	0.85	0.66	0.66	0.00	0.66	0.91	0.96	0.94	0.97	0.96	0.70	0.85	0.96	0.96	0.95	0.96	0.95	0.96	0.97	0.96
8	0.97	0.91	0.20	0.76	0.01	0.01	0.66	0.00	0.85	0.90	0.91	0.95	0.90	0.50	0.75	0.90	0.90	0.89	0.90	0.89	0.93	0.94	0.91
9	0.94	0.93	0.88	0.66	0.85	0.85	0.91	0.85	0.00	0.89	0.82	0.89	0.89	0.71	0.69	0.89	0.89	0.89	0.91	0.89	0.80	0.89	0.90
10	0.95	0.95	0.92	0.80	0.90	0.90	0.96	0.90	0.89	0.00	0.88	0.94	0.01	0.85	0.80	0.01	0.01	0.09	0.07	0.09	0.84	0.45	0.10
11	0.82	0.46	0.92	0.81	0.91	0.91	0.94	0.91	0.82	0.88	0.00	0.48	0.88	0.82	0.81	0.88	0.88	0.88	0.89	0.88	0.90	0.88	0.89
12	0.77	0.64	0.96	0.96	0.95	0.95	0.97	0.95	0.89	0.94	0.48	0.00	0.94	0.95	0.96	0.94	0.94	0.94	0.94	0.94	0.91	0.90	0.93
13	0.95	0.95	0.92	0.80	0.90	0.90	0.96	0.90	0.89	0.01	0.88	0.94	0.00	0.85	0.80	0.01	0.01	0.09	0.07	0.09	0.84	0.45	0.10
14	0.97	0.92	0.60	0.52	0.50	0.50	0.70	0.50	0.71	0.85	0.82	0.95	0.85	0.00	0.50	0.84	0.84	0.83	0.87	0.83	0.86	0.91	0.86
15	0.96	0.94	0.80	0.13	0.75	0.75	0.85	0.75	0.69	0.80	0.81	0.96	0.80	0.50	0.00	0.79	0.79	0.78	0.81	0.78	0.73	0.87	0.81
16	0.95	0.95	0.92	0.80	0.90	0.90	0.96	0.90	0.89	0.01	0.88	0.94	0.01	0.84	0.79	0.00	0.01	0.11	0.09	0.11	0.84	0.46	0.08
17	0.95	0.95	0.92	0.80	0.90	0.90	0.96	0.90	0.89	0.01	0.88	0.94	0.01	0.84	0.79	0.01	0.00	0.11	0.09	0.11	0.84	0.46	0.08
18	0.95	0.94	0.91	0.78	0.89	0.89	0.95	0.89	0.89	0.96	0.88	0.94	0.09	0.83	0.78	0.11	0.11	0.00	0.16	0.01	0.83	0.51	0.19
19	0.95	0.95	0.92	0.82	0.90	0.90	0.96	0.90	0.91	0.07	0.89	0.94	0.07	0.87	0.81	0.09	0.09	0.16	0.00	0.16	0.86	0.48	0.16
20	0.95	0.94	0.91	0.78	0.89	0.89	0.95	0.89	0.89	0.09	0.88	0.94	0.09	0.83	0.78	0.11	0.11	0.01	0.16	0.00	0.83	0.51	0.19
21	0.90	0.97	0.94	0.75	0.93	0.93	0.96	0.93	0.80	0.84	0.90	0.91	0.84	0.86	0.73	0.84	0.84	0.83	0.86	0.83	0.00	0.87	0.80
22	0.90	0.93	0.95	0.87	0.94	0.94	0.97	0.94	0.89	0.45	0.88	0.90	0.45	0.91	0.87	0.46	0.46	0.51	0.48	0.51	0.87	0.00	0.49
23	0.93	0.95	0.92	0.81	0.91	0.91	0.96	0.91	0.90	0.10	0.89	0.93	0.10	0.86	0.81	0.08	0.08	0.19	0.16	0.19	0.80	0.49	0.00

Table 3. Constructed distance matrix based on dissimilarity of molecular function (gene ontology) of 23 proteins.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0.00	0.45	0.93	0.93	0.96	0.94	0.93	0.93	0.87	0.90	0.84	0.88	0.90	0.92	0.93	0.93	0.84	0.93	0.90	0.90	0.90	0.89	0.91
2	0.45	0.00	0.93	0.92	0.96	0.92	0.92	0.92	0.84	0.88	0.80	0.80	0.88	0.90	0.92	0.91	0.80	0.90	0.88	0.88	0.88	0.87	0.90
3	0.93	0.93	0.00	0.01	0.06	0.05	0.01	0.11	0.90	0.50	0.93	0.97	0.50	0.96	0.40	0.66	0.54	0.44	0.50	0.50	0.50	0.57	0.76
4	0.93	0.92	0.01	0.00	0.06	0.05	0.01	0.11	0.90	0.50	0.93	0.97	0.50	0.96	0.40	0.66	0.54	0.44	0.50	0.50	0.50	0.57	0.76
5	0.96	0.96	0.06	0.06	0.00	0.11	0.06	0.17	0.93	0.54	0.96	0.96	0.54	0.95	0.45	0.69	0.58	0.41	0.54	0.54	0.54	0.60	0.78
6	0.94	0.92	0.05	0.05	0.11	0.00	0.05	0.16	0.87	0.52	0.90	0.97	0.52	0.92	0.35	0.67	0.56	0.47	0.52	0.52	0.52	0.53	0.77
7	0.93	0.92	0.01	0.01	0.06	0.05	0.00	0.11	0.90	0.50	0.93	0.97	0.50	0.96	0.40	0.66	0.54	0.44	0.50	0.50	0.50	0.57	0.76
8	0.93	0.92	0.11	0.11	0.17	0.16	0.11	0.00	0.90	0.50	0.93	0.97	0.50	0.96	0.40	0.66	0.54	0.44	0.50	0.50	0.50	0.57	0.76
9	0.87	0.84	0.90	0.90	0.93	0.87	0.90	0.90	0.00	0.93	0.84	0.72	0.93	0.88	0.86	0.95	0.87	0.92	0.93	0.93	0.93	0.83	0.84
10	0.90	0.88	0.50	0.50	0.54	0.52	0.50	0.50	0.93	0.00	0.90	0.97	0.00	0.96	0.50	0.39	0.10	0.47	0.01	0.01	0.01	0.63	0.58
11	0.84	0.80	0.93	0.93	0.96	0.90	0.93	0.93	0.84	0.90	0.00	0.91	0.90	0.47	0.90	0.93	0.84	0.93	0.90	0.90	0.90	0.69	0.91
12	0.88	0.80	0.97	0.97	0.96	0.97	0.97	0.97	0.72	0.97	0.91	0.00	0.97	0.92	0.97	0.97	0.91	0.93	0.97	0.97	0.97	0.92	0.78
13	0.90	0.88	0.50	0.50	0.54	0.52	0.50	0.50	0.93	0.00	0.90	0.97	0.00	0.96	0.50	0.39	0.10	0.47	0.01	0.01	0.01	0.63	0.58
14	0.92	0.90	0.96	0.96	0.95	0.92	0.96	0.96	0.88	0.96	0.47	0.92	0.96	0.00	0.91	0.97	0.92	0.90	0.96	0.96	0.96	0.74	0.96
15	0.93	0.92	0.40	0.40	0.45	0.35	0.40	0.40	0.86	0.50	0.90	0.97	0.50	0.91	0.00	0.66	0.54	0.44	0.50	0.50	0.50	0.33	0.76
16	0.93	0.91	0.66	0.66	0.69	0.67	0.66	0.66	0.95	0.39	0.93	0.97	0.39	0.97	0.66	0.00	0.43	0.66	0.39	0.39	0.39	0.73	0.38
17	0.84	0.80	0.54	0.54	0.58	0.56	0.54	0.54	0.87	0.10	0.84	0.91	0.10	0.92	0.54	0.43	0.00	0.45	0.10	0.10	0.10	0.56	0.54
18	0.93	0.90	0.44	0.44	0.41	0.47	0.44	0.44	0.92	0.47	0.93	0.93	0.47	0.90	0.44	0.66	0.45	0.00	0.47	0.47	0.47	0.56	0.75
19	0.90	0.88	0.50	0.50	0.54	0.52	0.50	0.50	0.93	0.01	0.90	0.97	0.01	0.96	0.50	0.39	0.10	0.47	0.00	0.01	0.01	0.63	0.58
20	0.90	0.88	0.50	0.50	0.54	0.52	0.50	0.50	0.93	0.01	0.90	0.97	0.01	0.96	0.50	0.39	0.10	0.47	0.01	0.00	0.01	0.63	0.58
21	0.90	0.88	0.50	0.50	0.54	0.52	0.50	0.50	0.93	0.01	0.90	0.97	0.01	0.96	0.50	0.39	0.10	0.47	0.01	0.01	0.00	0.63	0.58
22	0.89	0.87	0.57	0.57	0.60	0.53	0.57	0.57	0.83	0.63	0.69	0.92	0.63	0.74	0.33	0.73	0.56	0.56	0.63	0.63	0.63	0.00	0.75
23	0.91	0.90	0.76	0.76	0.78	0.77	0.76	0.76	0.84	0.58	0.91	0.78	0.58	0.96	0.76	0.38	0.54	0.75	0.58	0.58	0.58	0.75	0.00

sis. In addition, as shown in Figures 7 and 8 sequence comparison and PPI network analysis of these 23 proteins confirmed the (GO) findings; that is, proteins with similar functions demonstrate similar sequence similarities, as they are categorized in distinct clusters. The first group (1, 2, 12) is involved in the apoptosis pathway, whereas the second group proteins (3, 4, 5, 6) pertain to the mitochondria oxidation pathway (electron transport chain). It has been reported that mitochondria function and apoptosis are involved in dementia diseases.[35] As the function of mitochondria is associated with energy production, it can play a significant role in cellular regulation and any changes can lead to vast abnormal phenotypes. Therefore, in this clustering algorithm, the individual genes are correctly corresponded to their particular features including BP, CC, and MF. By considering these valid findings, it is

Table 4. Constructed distance dissimilarity matrix of 23 proteins based on cellular compartment data via gene ontology.

	p1	p2	p3	P4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	p15	p16	p17	p18	p19	p20	p21	p22	p23
p1	0.00	0.70	0.60	0.61	0.60	0.63	0.56	0.65	0.67	0.68	0.51	0.55	0.66	0.61	0.61	0.62	0.61	0.67	0.66	0.67	0.68	0.68	0.58
p2	0.70	0.00	0.74	0.75	0.74	0.76	0.74	0.78	0.80	0.80	0.76	0.46	0.79	0.75	0.75	0.82	0.82	0.80	0.79	0.80	0.80	0.80	0.82
p3	0.60	0.74	0.00	0.03	0.01	0.10	0.20	0.27	0.33	0.25	0.50	0.63	0.36	0.03	0.03	0.41	0.40	0.23	0.31	0.23	0.25	0.25	0.40
p4	0.61	0.75	0.03	0.00	0.03	0.06	0.23	0.24	0.30	0.27	0.47	0.64	0.37	0.01	0.01	0.42	0.41	0.25	0.33	0.25	0.27	0.22	0.41
p5	0.60	0.74	0.01	0.03	0.00	0.10	0.20	0.27	0.33	0.25	0.50	0.63	0.36	0.03	0.03	0.41	0.40	0.23	0.31	0.23	0.25	0.25	0.40
p6	0.63	0.76	0.10	0.06	0.10	0.00	0.28	0.18	0.25	0.27	0.50	0.66	0.41	0.06	0.06	0.45	0.40	0.25	0.32	0.25	0.27	0.17	0.40
p7	0.56	0.74	0.20	0.23	0.20	0.28	0.00	0.32	0.37	0.39	0.42	0.63	0.40	0.23	0.23	0.41	0.40	0.37	0.36	0.37	0.39	0.39	0.40
p8	0.65	0.78	0.27	0.24	0.27	0.18	0.32	0.00	0.19	0.30	0.45	0.68	0.35	0.24	0.24	0.40	0.35	0.28	0.27	0.28	0.30	0.21	0.39
p9	0.67	0.80	0.33	0.30	0.33	0.25	0.37	0.19	0.00	0.35	0.48	0.71	0.40	0.30	0.30	0.44	0.40	0.34	0.32	0.34	0.35	0.27	0.43
p10	0.68	0.80	0.25	0.27	0.25	0.27	0.39	0.30	0.35	0.00	0.53	0.71	0.16	0.27	0.27	0.35	0.17	0.02	0.05	0.02	0.05	0.29	0.34
p11	0.51	0.76	0.50	0.47	0.50	0.50	0.42	0.45	0.48	0.53	0.00	0.63	0.54	0.47	0.47	0.41	0.40	0.52	0.51	0.52	0.53	0.50	0.43
p12	0.55	0.46	0.63	0.64	0.63	0.66	0.63	0.68	0.71	0.71	0.63	0.00	0.70	0.64	0.64	0.74	0.73	0.71	0.70	0.71	0.71	0.71	0.73
p13	0.66	0.79	0.36	0.37	0.36	0.41	0.40	0.35	0.40	0.16	0.54	0.70	0.00	0.37	0.37	0.36	0.22	0.18	0.11	0.18	0.21	0.41	0.38
p14	0.61	0.75	0.03	0.01	0.03	0.06	0.23	0.24	0.30	0.27	0.47	0.64	0.37	0.00	0.01	0.42	0.41	0.25	0.33	0.25	0.27	0.22	0.41
p15	0.61	0.75	0.03	0.01	0.03	0.06	0.23	0.24	0.30	0.27	0.47	0.64	0.37	0.01	0.00	0.42	0.41	0.25	0.33	0.25	0.27	0.22	0.41
p16	0.62	0.82	0.41	0.42	0.41	0.45	0.41	0.40	0.44	0.35	0.41	0.74	0.36	0.42	0.42	0.00	0.20	0.34	0.32	0.34	0.39	0.45	0.16
p17	0.61	0.82	0.40	0.41	0.40	0.40	0.40	0.35	0.40	0.17	0.40	0.73	0.22	0.41	0.41	0.20	0.00	0.20	0.13	0.20	0.21	0.41	0.19
p18	0.67	0.80	0.23	0.25	0.23	0.25	0.37	0.28	0.34	0.02	0.52	0.71	0.18	0.25	0.25	0.34	0.20	0.00	0.08	0.01	0.83	0.27	0.32
p19	0.66	0.79	0.31	0.33	0.31	0.32	0.36	0.27	0.32	0.05	0.51	0.70	0.11	0.33	0.33	0.32	0.13	0.08	0.00	0.08	0.11	0.34	0.30
p20	0.67	0.80	0.23	0.25	0.23	0.25	0.37	0.28	0.34	0.02	0.52	0.71	0.18	0.25	0.25	0.34	0.20	0.01	0.08	0.00	0.83	0.27	0.32
p21	0.68	0.80	0.25	0.27	0.25	0.27	0.39	0.30	0.35	0.05	0.53	0.71	0.21	0.27	0.27	0.39	0.21	0.08	0.11	0.08	0.00	0.29	0.37
p22	0.68	0.80	0.25	0.22	0.25	0.17	0.39	0.21	0.27	0.29	0.50	0.71	0.41	0.22	0.22	0.45	0.41	0.27	0.34	0.27	0.29	0.00	0.44
p23	0.58	0.82	0.40	0.41	0.40	0.40	0.40	0.39	0.43	0.34	0.43	0.73	0.38	0.41	0.41	0.16	0.19	0.32	0.30	0.32	0.37	0.44	0.00

Table 5. Blosum 62 score similarity matrix obtained based on Needleman–Wunsch algorithm protein global alignment for 23 proteins.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	1	0.33	0.07	0.01	0.1	0.05	0.1	0.01	0.05	0.03	0.15	0.08	0.26	0.28	0.02	0.28	0.11	0.01	0.16	0	0.18	0.09	0.01
2	0.33	1	0.03	0.03	0.07	0.08	0.05	0.08	0.2	0.06	0.07	0.09	0.07	0.24	0.06	0.04	0.05	0.12	0.01	0.11	0.07	0.22	0.01
3	0.07	0.03	1	0.05	0.37	0.22	0.12	0.26	0.04	0.17	0.15	0.01	0.02	0	0.09	0.09	0	0.12	0.2	0.07	0.05	0.06	0.09
4	0.01	0.03	0.05	1	0.04	0.05	0.05	0.1	0.01	0.06	0.03	0.04	0.03	0.17	0.06	0	0.17	0.06	0.08	0.06	0.26	0.14	0.16
5	0.1	0.07	0.37	0.04	1	0.37	0.04	0.04	0.04	0.2	0.03	0.02	0.1	0.05	0.07	0.03	0.12	0.1	0	0.15	0.09	0.05	0.09
6	0.05	0.08	0.22	0.05	0.37	1	0.28	0.03	0.05	0.01	0.16	0.02	0.05	0.06	0.13	0.11	0.13	0.22	0.2	0.17	0.08	0.06	0.05
7	0.1	0.05	0.12	0.05	0.04	0.28	1	0.05	0.05	0.1	0.01	0.02	0.06	0.09	0.14	0.08	0.11	0.02	0.03	0	0.01	0.02	0.01
8	0.01	0.08	0.26	0.1	0.04	0.03	0.05	1	0.07	0.26	0.24	0.04	0.17	139	0.2	0.14	0.08	0.12	0.2	0.11	0.13	0.05	0.08
9	0.05	0.2	0.04	0.01	0.04	0.05	0.05	0.07	1	0.05	0.07	0.13	0.07	0.16	0.02	0.17	0.15	0.07	0.06	0.07	0.01	0.25	0.01
10	0.03	0.06	0.17	0.06	0.02	0.01	0.1	0.26	0.05	1	0.07	0.02	0.13	0.04	0.28	0.04	0.01	0.06	0.01	0.02	0.04	0.01	0.12
11	0.15	0.07	0.15	0.03	0.03	0.16	0.01	0.24	0.07	0.07	1	0.03	0.07	0.16	0.22	0.12	0.14	0	0.19	0.03	0.02	0.07	0.01
12	0.08	0.09	0.01	0.04	0.02	0.02	0.02	0.04	0.13	0.02	0.03	1	0.06	0.09	0.02	0.1	0.09	0	0.02	0.02	0.04	0.08	0.08
13	0.03	0.07	0.02	0.03	0.1	0.05	0.06	0.17	0.07	0.13	0.07	0.06	1	0.16	0.05	0.05	0.05	0.06	0.29	0.1	0.03	0.08	0.07
14	0.28	0.24	0.03	0.17	0.05	0.06	0.09	0.14	0.16	0.04	0.16	0.09	0.16	1	0.03	0.19	0.13	0.08	0.09	0.07	0.24	0.22	0.07
15	0.02	0.06	0.09	0.06	0.07	0.13	0.14	0.2	0.02	0.28	0.22	0.02	0.05	0.03	1	0.04	0.09	0.14	0.06	0.13	0.07	0.01	0.09
16	0.28	0.04	0.09	0	0.03	0.11	0.08	0.14	0.17	0.04	0.12	0.1	0.05	0.19	0.04	1	0.17	0.01	0.16	0.05	0.15	0.2	0.08
17	0.11	0.05	0.04	0.17	0.12	0.13	0.11	0.08	0.15	0.1	0.14	0.09	0.05	0.13	0.09	0.17	1	0.04	0.17	0.12	0.17	0.19	0.2
18	0.01	0.12	0.12	0.06	0.1	0.22	0.02	0.12	0.07	0.06	0	0	0.06	0.08	0.14	0.01	0.04	1	0.02	0.78	0.14	0.1	0.05
19	0.16	0.01	0.2	0.08	0.03	0.2	0.03	0.2	0.06	0.01	0.19	0.02	0.29	0.09	0.06	0.16	0.17	0.02	1	0.04	0.02	0.09	0.04
20	0	0.11	0.07	0.06	0.15	0.17	0	0.11	0.07	0.02	0.03	0.02	0.1	0.07	0.01	0.05	0.01	0.08	0.04	1	0.13	0.1	0.24
21	0.18	0.07	0.05	0.0.258	0.09	0.08	0.01	0.13	0.01	0.04	0.02	0.04	0.03	0.24	0.07	0.15	0.17	0.14	0.02	0.13	1	0.07	0.25
22	0.09	0.22	0.06	0.14	0.05	0.06	0.02	0.05	0.25	0.01	0.07	0.08	0.08	0.22	0.01	0.2	0.19	0.1	0.09	0.1	0.07	1	0.04
23	0.1	0.01	0.09	0.16	0.09	0.05	0.01	0.08	0.01	0.12	0.01	0.08	0.07	0.07	0.09	0.08	0.2	0.05	0.04	0.24	0.25	0.04	1

Table 6. Highly integrated annotation based on molecular function by the use of DAVID program.

Annotation Cluster 1	Enrichment Score: 20.35		Count	P_Value	Benjamini
GOTERM_MF_FAT	hydrogen ion transmembrane transporter activity	RT	13	2.8E-22	2.2E-20
GOTERM_MF_FAT	monovalent inorganic cation transmembrane transporter activity	RT	13	1.7E-21	6.8E-20
GOTERM_MF_FAT	inorganic cation transmembrane transporter activity	RT	13	1.8E-19	4.8E-18
Annotation Cluster 2	**Enrichment Score: 7.18**		**Count**	**P_Value**	**Benjamini**
GOTERM_MF_FAT	cytochrome-c oxidase activity	RT	5	6.6E-8	1.3E-6
GOTERM_MF_FAT	heme-copper terminal oxidase activity	RT	5	6.6E-8	1.3E-6
GOTERM_MF_FAT	oxidoreductase activity, acting on heme group of donors	RT	5	6.6E-8	1.3E-6
GOTERM_MF_FAT	oxidoreductase activity, acting on heme group of donors, oxygen as acceptor	RT	5	6.6E-8	1.3E-6
Annotation Cluster 3	**Enrichment Score: 0.04**		**Count**	**P_Value**	**Benjamini**
GOTERM_MF_FAT	metal ion binding	RT	5	9.0E-1	1.0E0
GOTERM_MF_FAT	cation binding	RT	5	9.1E-1	1.0E0
GOTERM_MF_FAT	ion binding	RT	5	9.1E-1	1.0E0

Table 7. Highly integrated annotation based on cell component by the use of DAVID program.

Annotation Cluster 1	Enrichment Score: 16.84		Count	P_Value	Benjamini
GOTERM_CC_FAT	mitochondrial envelope	RT	16	1.4E-19	8.2E-18
GOTERM_CC_FAT	mitochondrial inner membrane	RT	15	1.6E-19	4.5E-18
GOTERM_CC_FAT	organelle inner membrane	RT	15	4.4E-19	8.3E-18
GOTERM_CC_FAT	organelle envelope	RT	17	6.5E-19	9.3E-18
GOTERM_CC_FAT	envelope	RT	17	6.9E-19	7.8E-18
GOTERM_CC_FAT	mitochondrial membrane	RT	15	5.6E-18	5.3E-17
GOTERM_CC_FAT	mitochondrial part	RT	16	2.8E-17	2.3E-16
GOTERM_CC_FAT	mitochondrion	RT	17	5.1E-15	3.6E-14
GOTERM_CC_FAT	organelle membrane	RT	15	8.3E-12	3.9E-11
Annotation Cluster 2	**Enrichment Score: 10.84**		**Count**	**P_Value**	**Benjamini**
GOTERM_CC_FAT	mitochondrial membrane part	RT	10	5.2E-14	3.3E-13
GOTERM_CC_FAT	mitochondrial proton-transporting ATP synthase complex	RT	7	1.2E-13	6.9E-13
GOTERM_CC_FAT	proton-transporting ATP synthase complex	RT	7	2.4E-13	1.2E-12
GOTERM_CC_FAT	proton-transporting two-sector ATPase complex	RT	7	3.5E-11	1.5E-10
GOTERM_CC_FAT	proton-transporting ATP synthase complex, coupling factor F(o)	RT	5	4.7E-9	1.9E-8
GOTERM_CC_FAT	proton-transporting two-sector ATPase complex, proton-transporting domain	RT	5	3.6E-8	1.4E-7
Annotation Cluster 3	**Enrichment Score: 0.67**		**Count**	**P_Value**	**Benjamini**
GOTERM_CC_FAT	membrane-enclosed lumen	RT	6	1.3E-1	3.7E-1
GOTERM_CC_FAT	intracellular organelle lumen	RT	5	2.7E-1	6.3E-1
GOTERM_CC_FAT	organelle lumen	RT	5	2.8E-1	6.3E-1

Table 8. Highly integrated annotation based on biological process by the use of DAVID program.

Annotation Cluster 1	Enrichment Score: 6.7		Count	P_Value	Benjamini
GOTERM_BP_FAT	oxidative phosphorylation	RT	9	4.0E-13	5.3E-11
GOTERM_BP_FAT	phosphorylation	RT	9	6.1E-6	5.3E-4
GOTERM_BP_FAT	phosphorus metabolic process	RT	9	2.6E-5	1.3E-3
GOTERM_BP_FAT	phosphate metabolic process	RT	9	2.6E-5	1.3E-3

feasible to extract more information related to other clusters.[18] It can be concluded that, clustering methods can provide novel description about molecular aspects of diseases. Furthermore, DAVID program annotation analysis from Tables 2 to 4, confirmed the clustering results significantly. That is, the mitochondria are the main malfunctioning organelle with the highlighted role of oxidative processes in these two diseases. On the other hand, PPI network analysis provided a presentation of interactions between these common proteins.[4] These finding can lead to determining new proteins that are involved with these known proteins. It is possible to analyze the important roles of each protein in the network for finding suitable candidates for drug targeting and treatment purposes. In PPI network, Casp3,Casp9, and APAF1 are in close relationship and are the most significant hub proteins; on the other hand, Cox7c, Cox1, Cox8A, Cox7A1 proteins are also in direct connection, but with no significant centrality. It seems that despite their different phenotypes, AD and PD diseases can have a common molecular basis. If proven, it can be concluded that neurodegenerative diseases have a common origin in which alternation in one specific biological process can trigger the onset of different types of these diseases.

In conclusion, it was clear that the clustering methods reveal novel annotation patterns within the dataset that would not have been identified otherwise. Furthermore, significant molecular relationships with these diseases suggest examining them as two highly-related diseases for diagnosis and therapeutic purposes. Therefore, by applying comprehensive statistical analysis, it is hoped that clinical approaches can be feasible. In addition to this, these proteins can be assigned as potential biomarkers in drug targeting and other clinical approaches. It is also suggested that other neurodegenerative diseases with similar neuropathological features should be evaluated to elicit possible common molecular

origin. This theory may not be proven, so before whole analysis, studying pairs of diseases can be helpful.

Acknowledgment

This research has been supported by Proteomics Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran. It is derived from Miss Rajabi's MSc thesis and Miss Mona Zamanian Azodi's Ph.D. thesis.

References

1. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature.* 2003; 422(6938): 198 – 207.
2. King AD, Pržulj N, Jurisica I. Protein complex prediction via cost-based clustering. *Bioinformatics.* 2004; 20(17): 3013 – 3020.
3. Zali H, Rezaei Tavirani M. Meningioma protein-protein interaction network. *Arch Iran Med.* 2014; 17(4): 262.
4. Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, Goliaei B, Peyvandi A. Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterol Hepatol bed bench.* 2014; 7(1): 17.
5. Renaud J, Nabavi Sm, Daglia M, Nabavi SM, Martinoli MG. Epigallocatechin-3-gallate, a promising molecule for Parkinson's disease? *Rejuvenation Res.* 2015; 18(3): 257 – 269.
6. Zali H, Zamanian-Azodi M, Rezaei Tavirani M,Akbar-zadeh Baghban A. Protein Drug Targets of Lavandula angustifolia on treatment of Rat Alzheimer's Disease. *Iran J Pharm Res*. 2015; 14(1): 291.
7. Ghamari E, Zali H, Rezaei Tavirani M, Hesami Takalu S, Goshadrou F, Ahmadi N, Zamaheni S, et al . Proteomic study in the rat hippocampus as a measure of human Alzheimer's disease. *Koomesh.* 2015; 16(4): 611 – 620.
8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM. Gene ontology: tool for the unification of biology. *Nature Genetics.* 2000; 25(1): 25 – 29.
9. Consortium GO. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 2004; 32(suppl 1): 258 – 261.
10. Consortium GO. Creating the gene ontology resource: design and implementation. *Genome Res.* 2001; 11(8): 1425 – 1433.
11. Nia YZ, Majd HA, Azodi M,Khayyer N. Using partitioning and non-partitioning clustering algorithms for included proteins sequences in esophagus, stomach and colon cancer. *J Paramed Sci.* 2011; 2(2): 9 – 16.
12. Brun C, Herrmann C, Guénoche A. Clustering proteins from interaction networks for the prediction of cellular functions. *BMC Bioinformatics.* 2004; 5(1): 95.
13. Gil J, Esteban M. Induction of apoptosis by the dsRNA-dependent protein kinase (PKR): mechanism of action. *Apoptosis.* 2000; 5(2): 107 – 114.
14. Zamanian Azodi M, Rezaei Tavirani M, Rahmatirad S, Hadi Hasanzadeh,Majid Rezaei Tavirani, Samaneh Sadat Seyyedi. Protein-Protein Interaction Network could reveal the relationship between the breast and colon cancer. *Gastroenterol Hepatol Bed Bench.* 2015; 8(3): 215 – 224.
15. Phizicky EM, Fields S. Protein-protein interactions: methods for detection and analysis. *Microbiol Mol Biol Rev*. 1995; 59(1): 94 – 123.
16. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck F, Goehler H, et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell.* 2005; 122(6): 957 – 968.
17. Soheili M, Rezaei Tavirani M, Salami M. Clearance of amyloid beta plaques from brain of Alzheimeric rats by lavandula angustifolia. *Neurosci Med.* 2012; 3(4): 362 – 367.
18. Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, et al. Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol.* 2010; 9(11): 1118 – 1127.
19. Habibi E, Masoudi-Nejad A, Abdolmaleky HM, Haggarty S. Emerging roles of epigenetic mechanisms in Parkinson's disease. *Funct Integr Genomics*. 2011; 11(4): 523 – 537.
20. Deep-Brain Stimulation for Parkinson's Disease Study Group. Deep-brain stimulation of the subthalamic nucleus or the pars interna of the globus pallidus in Parkinson's disease. *N Engl J Med,* 2001; 345(13): 956 – 963.
21. Hoang QQ. Pathway for Parkinson disease. *Proc Natl Acad Sci.* 2014; 111(7): 2402 – 2403.
22. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids Res.* 2000; 28(1): 27 – 30.
23. Camon E, Magrane M, Barrell D, Barrell D, Lee V, Dimmer E, Maslen J. The gene ontology annotation (GOA) database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res.* 2004; 32: 262 – 266.
24. Qiu X, Brooks AI, Klebanov L, Yakovlev A. The effects of normalization on the correlation structure of microarray data. *BMC Bioinformatics.* 2005; 6: 120.
25. Wolting C, McGlade CJ, Tritchler D. Cluster analysis of protein array results via similarity of gene ontology annotation. *BMC Bioinformatics.* 2006; 7(1): 1 – 13
26. Zarnegarnia Y, Majd HA, Rezaei-Tavirani R, Khaier N, Maboodi AAK. Application of fuzzy clustering in analysis of included proteins in esophagus, stomach and colon cancers based on similarity of gene ontology annotation. *Koomesh.* 2010; 12(1): 14 – 21.
27. Nordström KJ, Almén MS, Edstam MM, Fredriksson R, Schiöth H. Independent HHsearch, Needleman--Wunsch-based, and motif analyses reveal the overall hierarchy for most of the G protein-coupled receptor families. *Mol Biol Evol.* 2011; 28(9): 2471 – 2480.
28. Owen DM, Williamson D, Magenau A. Optical techniques for imaging membrane domains in live cells (live-cell palm of protein clustering). *Methods Enzymol.* 2012; 504: 221 – 235.
29. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* 1987; 20: 53 – 65
30. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2008; 4(1): 44 – 57.
31. Da Wei Huang BTS, Stephens R, Baseler MW,Baseler M, Lane C, Lempicki R. DAVID gene ID conversion tool. *Bioinformation.* 2008; 2(10): 428 – 430.
32. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009; 37(1): 1 – 13.
33. Rojo DR, Prough DS, Falduto MT, Boone DR, Micci MA, Kahrig KM, et al. Influence of stochastic gene expression on the cell survival rheostat after traumatic brain injury. *PLoS One*. 2011; 6(8): e23111.
34. Chattopadhyay AK, Nasiev D, Flower DR. A statistical physics perspective on alignment-independent protein sequence comparison. *Bioinformatics*. 2015; 31(15): 2469 – 2474.
35. Chen X, Yan SD. Mitochondrial abeta: a potential cause of metabolic dysfunction in alzheimer's disease. *IUBMB Life*. 2006; 58: 686 – 694