

Potential hippocampal genes and pathways involved in Alzheimer's disease: a bioinformatic analysis

L. Zhang*, X.Q. Guo*, J.F. Chu, X. Zhang, Z.R. Yan and Y.Z. Li

Department of Neurology, Shandong Ji'ning No. 1 People' Hospital, Ji'ning, China

*These authors contributed equally to this study.

Corresponding author: Y.Z. Li

E-mail: yuzhi_li@yeah.net

Genet. Mol. Res. 14 (2): 7218-7232 (2015)

Received September 30, 2014

Accepted January 23, 2015

Published June 29, 2015

DOI <http://dx.doi.org/10.4238/2015.June.29.15>

ABSTRACT. Alzheimer's disease (AD) is a neurodegenerative disorder and the most common cause of dementia in elderly people. Numerous studies have focused on the dysregulated genes in AD, but the pathogenesis is still unknown. In this study, we explored critical hippocampal genes and pathways that might potentially be involved in the pathogenesis of AD. Four transcriptome datasets for the hippocampus of patients with AD were downloaded from ArrayExpress, and the gene signature was identified by integrated analysis of multiple transcriptomes using novel genome-wide relative significance and genome-wide global significance models. A protein-protein interaction network was constructed, and five clusters were selected. The biological functions and pathways were identified by Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. A total of 6994 genes were screened, and the top 300 genes were subjected to further analysis. Four significant KEGG pathways were identified, including oxidative phosphorylation and Parkinson's

disease, Huntington's disease, and Alzheimer's disease pathways. The hub network of cluster 1 with the highest average rank value was defined. The genes (*NDUFB3*, *NDUFA9*, *NDUFV1*, *NDUFV2*, *NDUFS3*, *NDUFA10*, *COX7B*, and *UQCRI*) were considered critical with high degree in cluster 1 as well as being shared by the four significant pathways. The oxidative phosphorylation process was also involved in the other three pathways and is considered to be relevant to energy-related AD pathology in the hippocampus. This research provides a perspective from which to explore critical genes and pathways for potential AD therapies.

Key words: Alzheimer's disease; Hippocampus; Critical genes; Pathways

INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disorder, and is the most prevalent cause of dementia with no reported cure to date (Desai et al., 2005). As AD progresses, patients usually develop the symptoms of short-term to long-term memory loss, accompanied by confusion, irritability, aggression, mood swings, language difficulties (Reisberg et al., 1987), and eventually death (Schoenberg et al., 1987). According to a recent report (Thies et al., 2013), approximately 76 million people in the world were living with AD in 2013. Brookmeyer et al. (2007) have estimated that AD could affect one in 85 people worldwide by 2050. Therefore, developing primary and secondary prevention strategies based on genes relevant to AD has become a popular field of research in recent decades (Boada et al., 2014). Liao et al. (2014) studied the effect of ATP-binding cassette subfamily A member 7 gene (*ABCA7*) on the risk of AD, and found that *ABCA7* was significantly associated with AD ($P = 0.0026$). A study of the influence of 10 recently identified AD risk genes on the age-at-onset phenotype was carried out by Thambisetty et al. (2013) to identify genes that could potentially delay the onset of AD.

Despite the numerous studies on risk genes for AD, the molecular pathogenesis is still unclear (Yin et al., 2014), which constitutes a major challenge for research on AD. So far, several hypotheses have attempted to explain the causes of AD, including genetics, and the cholinergic, amyloid, and tau hypotheses. Among them, the cholinergic hypothesis forms the basis for most currently available drug therapies (Babic et al., 1999). A relationship has been identified between the cognitive and behavioral symptoms of AD and the dysfunction of certain brain regions (Liang et al., 2008a). A comprehensive study carried out by Liang et al. (2012) based the identification of the most critical dysregulated genes in six different AD-relevant brain regions on a sub-graph algorithm. However, this study only analyzed one dataset for each brain region and ignored the large body of research that has focused on one particular brain region.

The hippocampus is located under the cerebral cortex (Freund and Buzsáki, 1996), and is a major component of the human brain. The hippocampus is considered to be the first affected region in the brains of AD patients, because it plays vital role in consolidating information from short-term to long-term memory (West et al., 2000), the dysregulation of which is an early and severe symptom in AD patients (Lopez et al., 2003). Thus, we have focused our attention on the hippocampus to identify vital genes and pathways that might potentially give insight into effective therapies for AD.

In recent years, a large amount of transcription data has been produced and deposited in publically available data repositories (Parkinson et al., 2007; Barrett et al., 2011). These repositories allow researchers to discover genetic and diagnostic signatures by data integration and bioinformatic analysis, which can provide insight into the biological mechanisms of AD. In order to integrate transcription data from different platforms and protocols, a new model has been employed, which measures the genome-wide relative significance (GWRS) and genome-wide global significance (GWGS) (Liu et al., 2013).

In the current study, we initially downloaded transcription data from the hippocampus of AD patients in ArrayExpress. Using the GWRS and GWGS method, robust gene signatures were identified. To interrogate the interrelationships within these genes, protein-protein interactions (PPIs) were used to discover hub genes and sub-networks. These gene signatures were also tested by functional and pathway enrichment analysis. This research might provide information on vital genes and pathways that could potentially be exploited by therapies for AD.

MATERIAL AND METHODS

Data recruitment and preprocessing

In this study, four hippocampal microarray expression profiles from patients with AD and age-matched normal controls were downloaded from ArrayExpress, including E-GEOD-1297 (Blalock et al., 2004), E-GEOD-5281 (Liang et al., 2007, 2008b), E-GEOD-28146 (Blalock et al., 2011), and E-GEOD-36980 (Hokama et al., 2014). From these four datasets, a total of 61 AD patients and 40 normal controls were included in the current study. The characteristics of these studies are shown in Table 1.

Table 1. Characteristics of the previous studies included in the current study.

Accession No.	Year	Sample size	Platform
		Total (cases/controls)	
E-GEOD-1297	2004	31 (22/9)	Affymetrix HG-U133A
E-GEOD-5281	2007	23 (10/13)	Affymetrix HG-U133Plus2
E-GEOD-28146	2011	30 (22/8)	Affymetrix HG-U133Plus2
E-GEOD-36980	2014	17 (7/10)	Affymetrix HuGene-1.0st

Prior to analysis, the original expression information from all conditions was subjected to data preprocessing. For each dataset, in order to eliminate the influence of nonspecific hybridization, background correction and normalization were carried out using the robust multichip average (RMA) method (Ma et al., 2006) and a quartile-based algorithm (Rifai et al., 2001), respectively. Perfect match and mismatch values were revised using the Micro Array Suite 5.0 (MAS 5.0) algorithm (Pepper et al., 2007), and selected via the median method. The data were then screened using the feature filter method of the genefilter package. Each probe is mapped to one gene, and the probe is discarded if it cannot match any genes.

Integrated analysis of gene signatures from multiple microarray datasets

In the present study, the GWRS and GWGS method was used to identify gene signatures. The details of this approach have been described in previous research (Liu et al., 2013).

The degree of differential gene expression in each single microarray database was measured by GWRS on a genome-wide scale (r value) based on fold change. In this method, the number of datasets was denoted by n , and the number of unique genes across n datasets was denoted by m . The GWRS of the i -th gene in the j -th dataset was measured by:

$$S_{ij} = -2 \log\left(\frac{r_{ij}}{m}\right) \quad (\text{Equation 1})$$

where r_{ij} , $i = 1-m$, $j = 1-n$, was the rank number of the i -th gene in the j -th study. In current study, the value of n was 4.

Based on the corresponding GWRS across multiple microarray datasets, the GWGS of a gene was measured by:

$$S_i^r = \sum_{j=1}^n \omega_j S_{ij} \quad (\text{Equation 2})$$

where ω_j represents the relative weight of the j -th dataset. A gene with a large GWGS value was considered to be globally significant across multiple studies. The degree of differential expression was measured by the fold change. We assigned a rank number for each gene according to its degree of differential expression. In the study by Liu et al. (2013), the top 200 genes were considered. In our study, the reliability and robustness of the network should also be taken into account. Therefore, the top 300 genes were finally selected for further analysis, as 300 was the smallest number of genes that would allow optimal performance to be retained for classification accuracy in this study.

Construction of a PPI network

Since genes seldom implement their functions independently, network analysis presents a powerful tool for understanding the functional organization of the proteome. To reveal the dysfunctional clusters and identify genes that may play a vital role in the disease process, the PPI network was constructed by linking causal disease genes with the 300 most responsive genes. For protein interaction data, we utilized a known interactome database from the search tool for the retrieval of interacting genes/proteins (STRING) (Jensen et al., 2009). The PPI network was constructed using Cytoscape (Scardoni et al., 2009), a free software package for visualizing, modeling, and analyzing the integration of biomolecular interaction networks with high-throughput expression data.

Identification of clusters

In order to detect the densely connected regions and their functions in the network, we utilized a plug-in for the Cytoscape software, ClusterONE (Nepusz et al., 2012), to perform the clustering analysis. The top 300 genes with corresponding P values and fold changes were mapped into Cytoscape software. To find the densely connected network module, we discarded the clusters with fewer than 11 nodes and a density lower than 0.2.

Functional enrichment analysis

To further study the functions and enriched pathways of these top 300 genes, functional enrichment and pathway analysis were performed based on the Gene Ontology (GO) database (Version No. 2010.09.03) (<http://www.geneontology.org/>) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (www.genome.jp/kegg/) (Kanehisa and Goto, 2000), respectively. In the functional analysis, GO terms with P values greater than 0.01 were discarded. The enriched terms with more than five genes present and P values < 0.01 were considered to be statistically significant in the pathway enrichment analysis (Duncan et al., 2010). The two analyses were performed using the database for annotation, visualization, and integrated discovery (DAVID, <http://david.abcc.ncifcrf.gov/tools.jsp>) (Huang et al., 2008). The significant categories were identified by the Expression Analysis Systematic Explorer (EASE) score (Hosack et al., 2003).

RESULTS

Integrated analysis of gene signatures

After preprocessing the expression data from the four different datasets, the number of genes in E-GEOD-1297, E-GEOD-5281, E-GEOD-28146, and E-GEOD-36980 were 12493, 20109, 20109, and 10986, respectively, using the GWRS algorithm. A method based on fold change was applied when measuring the GWRS value for differential expression. Figure 1 shows the fold change for the top 44 genes. The GWRS of a gene was then measured based on the GWRS value. A total of 6994 genes were screened using the intersection of the microarray datasets. The top 300 AD genes in the hippocampus were identified for further analysis (Table 2).

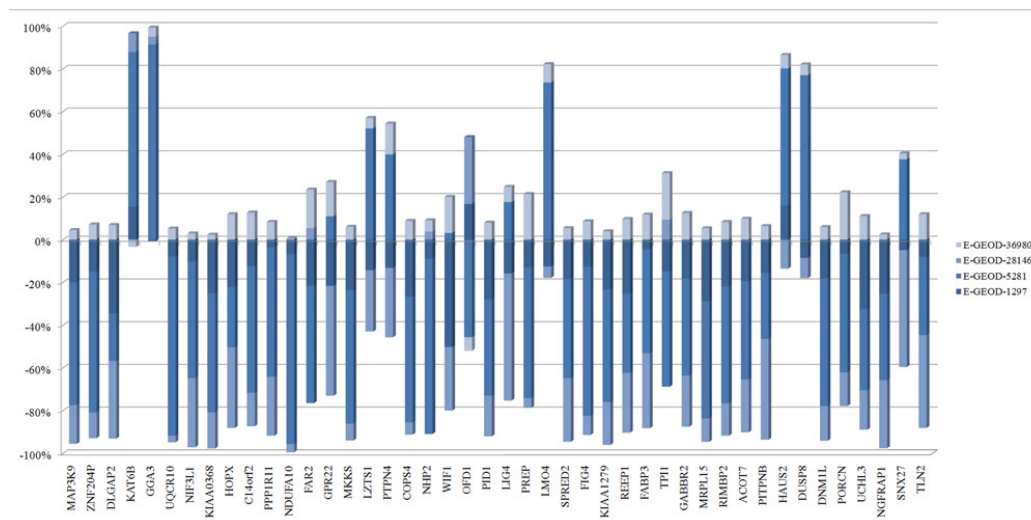


Figure 1. Stacked bar plots of fold changes for the top 44 ranked genes across four transcriptome datasets from the hippocampus of patients with Alzheimer's disease.

Table 2. Top 300 genes identified by integrated analysis that affect the hippocampus in Alzheimer's disease.

No.	Genes	No.	Genes	No.	Genes
1	<i>MAP3K9</i>	101	<i>RAD51C</i>	201	<i>DLGAP4</i>
2	<i>ZNF204P</i>	102	<i>EHD3</i>	202	<i>RGS4</i>
3	<i>DLGAP2</i>	103	<i>CCT4</i>	203	<i>ATP6V0B</i>
4	<i>KAT6B</i>	104	<i>HIF1AN</i>	204	<i>COP55</i>
5	<i>GGA3</i>	105	<i>SHQ1</i>	205	<i>ELOVL4</i>
6	<i>UQCR10</i>	106	<i>SEPT6</i>	206	<i>ADRBK2</i>
7	<i>NIF3L1</i>	107	<i>RIC8A</i>	207	<i>TOMM70A</i>
8	<i>KIAA0368</i>	108	<i>BEND5</i>	208	<i>SLC20A1</i>
9	<i>HOPX</i>	109	<i>PPP1R7</i>	209	<i>QPCT</i>
10	<i>C14orf2</i>	110	<i>KCNJ3</i>	210	<i>XK</i>
11	<i>PPP1R11</i>	111	<i>ACTR10</i>	211	<i>PRDM2</i>
12	<i>NDUFA10</i>	112	<i>C14orf132</i>	212	<i>BCL11A</i>
13	<i>FAR2</i>	113	<i>NME1</i>	213	<i>CCDC130</i>
14	<i>GPR22</i>	114	<i>WDFY3</i>	214	<i>UBR7</i>
15	<i>MKKS</i>	115	<i>TRIM14</i>	215	<i>ATP5J</i>
16	<i>LZTS1</i>	116	<i>MTMR4</i>	216	<i>UBE2W</i>
17	<i>PTPN4</i>	117	<i>TOMM34</i>	217	<i>PPP1CA</i>
18	<i>COP54</i>	118	<i>ATG9A</i>	218	<i>MYTIL</i>
19	<i>NHP2</i>	119	<i>B4GALNT1</i>	219	<i>PRKCZ</i>
20	<i>WIF1</i>	120	<i>MICU1</i>	220	<i>RHOQ</i>
21	<i>OFD1</i>	121	<i>FIBP</i>	221	<i>CDC5L</i>
22	<i>PID1</i>	122	<i>TSSC1</i>	222	<i>NDUFS3</i>
23	<i>LIG4</i>	123	<i>AMD1</i>	223	<i>POP4</i>
24	<i>PREP</i>	124	<i>MTX2</i>	224	<i>KCNKI</i>
25	<i>LMO4</i>	125	<i>PPIA</i>	225	<i>PFKM</i>
26	<i>SPRED2</i>	126	<i>MICALL2</i>	226	<i>MED24</i>
27	<i>FIG4</i>	127	<i>ZNF839</i>	227	<i>IMP4</i>
28	<i>KIAA1279</i>	128	<i>KCNK10</i>	228	<i>UTP18</i>
29	<i>REEP1</i>	129	<i>TRO</i>	229	<i>WARS</i>
30	<i>FABP3</i>	130	<i>DNAJC16</i>	230	<i>GLRB</i>
31	<i>TP11</i>	131	<i>MRPL16</i>	231	<i>ADSL</i>
32	<i>GABBR2</i>	132	<i>ASNS</i>	232	<i>PFDN2</i>
33	<i>MRPL15</i>	133	<i>PSMB7</i>	233	<i>OGDHL</i>
34	<i>RIMBP2</i>	134	<i>BAZ1B</i>	234	<i>NDUFA9</i>
35	<i>ACOT7</i>	135	<i>TTPAL</i>	235	<i>MXRA8</i>
36	<i>PITPNB</i>	136	<i>DUSP6</i>	236	<i>MTOR</i>
37	<i>HAUS2</i>	137	<i>RNFT2</i>	237	<i>DEAF1</i>
38	<i>DUSP8</i>	138	<i>UGP2</i>	238	<i>ROBO3</i>
39	<i>DNM1L</i>	139	<i>LRP8</i>	239	<i>RSG1</i>
40	<i>PORCN</i>	140	<i>U2AF1</i>	240	<i>PTDSS1</i>
41	<i>UCHL3</i>	141	<i>KIAA1045</i>	241	<i>CACNB2</i>
42	<i>NGFRAP1</i>	142	<i>HCCS</i>	242	<i>NDUFV2</i>
43	<i>SNX27</i>	143	<i>STRAP</i>	243	<i>KPNA2</i>
44	<i>TLN2</i>	144	<i>THOC5</i>	244	<i>TCEAL2</i>
45	<i>POLB</i>	145	<i>FSD1</i>	245	<i>L1CAM</i>
46	<i>SHFM1</i>	146	<i>KCNQ2</i>	246	<i>DMTF1</i>
47	<i>TMEM246</i>	147	<i>ICT1</i>	247	<i>BAG4</i>
48	<i>CAPZA2</i>	148	<i>FAM32A</i>	248	<i>SLC25A44</i>
49	<i>SLC25A4</i>	149	<i>VAPB</i>	249	<i>SYT13</i>
50	<i>INSIG2</i>	150	<i>MAD2L1BP</i>	250	<i>FAM178A</i>
51	<i>CCT3</i>	151	<i>ATP5J2</i>	251	<i>RFX4</i>
52	<i>TM9SF4</i>	152	<i>MZT2B</i>	252	<i>NT1</i>
53	<i>PDIA6</i>	153	<i>ACACA</i>	253	<i>NDUFA13</i>
54	<i>LDB2</i>	154	<i>CCNC</i>	254	<i>BOPI</i>
55	<i>RBM25</i>	155	<i>SYNGR3</i>	255	<i>KCND3</i>
56	<i>EPB41L1</i>	156	<i>COX7B</i>	256	<i>KANSL3</i>
57	<i>LRRC47</i>	157	<i>NDUFB3</i>	257	<i>WWTR1</i>
58	<i>TBC1D30</i>	158	<i>USO1</i>	258	<i>GTF2H3</i>
59	<i>NCALD</i>	159	<i>RIT2</i>	259	<i>ACOT8</i>
60	<i>SBNO1</i>	160	<i>TXNL1</i>	260	<i>GLS2</i>

Continued on next page

Table 2. Continued.

No.	Genes	No.	Genes	No.	Genes
61	<i>PDHAI</i>	161	<i>KTNI</i>	261	<i>NCL</i>
62	<i>CI2orf43</i>	162	<i>SCG3</i>	262	<i>XRCC6</i>
63	<i>ATP6V1D</i>	163	<i>FAM134C</i>	263	<i>CSRNP2</i>
64	<i>ZAK</i>	164	<i>FBXO21</i>	264	<i>RABGGTB</i>
65	<i>MCTS1</i>	165	<i>AP1S1</i>	265	<i>ANKRD46</i>
66	<i>MARS</i>	166	<i>VPS41</i>	266	<i>PSMA1</i>
67	<i>POLR1B</i>	167	<i>PRAF2</i>	267	<i>ARMCX2</i>
68	<i>NPTX2</i>	168	<i>PTPRO</i>	268	<i>MARK1</i>
69	<i>SMARCC2</i>	169	<i>RGS7</i>	269	<i>CELF3</i>
70	<i>GABRA5</i>	170	<i>ST6GALNAC5</i>	270	<i>AEBP1</i>
71	<i>RALGAPB</i>	171	<i>FBXO34</i>	271	<i>DYNLRB1</i>
72	<i>BABAM1</i>	172	<i>TTBK2</i>	272	<i>ELMO2</i>
73	<i>TOMM22</i>	173	<i>FRMPD4</i>	273	<i>B4GALT6</i>
74	<i>CFLAR</i>	174	<i>RB1CC1</i>	274	<i>ARF5</i>
75	<i>POP7</i>	175	<i>GPRASP1</i>	275	<i>TAC1</i>
76	<i>RAB15</i>	176	<i>SCG5</i>	276	<i>RNMT</i>
77	<i>NOLC1</i>	177	<i>ZHX3</i>	277	<i>ADH5</i>
78	<i>RFK</i>	178	<i>HK1</i>	278	<i>PPM1G</i>
79	<i>ARPC5L</i>	179	<i>SMOX</i>	279	<i>SHANK2</i>
80	<i>OXR1</i>	180	<i>SCN2B</i>	280	<i>EI24</i>
81	<i>BHLHB9</i>	181	<i>THYN1</i>	281	<i>KDM5B</i>
82	<i>RAN</i>	182	<i>AP3B2</i>	282	<i>KCTD2</i>
83	<i>BRD3</i>	183	<i>CHRN2</i>	283	<i>STAU2</i>
84	<i>TFEB</i>	184	<i>SMARCA4</i>	284	<i>SFXN3</i>
85	<i>ITPKB</i>	185	<i>PIP4K2C</i>	285	<i>RALBP1</i>
86	<i>TM2D3</i>	186	<i>LEPROTL1</i>	286	<i>SFSWAP</i>
87	<i>SCAI</i>	187	<i>MRPS7</i>	287	<i>UBE2Q1</i>
88	<i>LGII</i>	188	<i>SLC4A1AP</i>	288	<i>HPRT1</i>
89	<i>RAD23B</i>	189	<i>DYNLL1</i>	289	<i>CCT2</i>
90	<i>NEDD4L</i>	190	<i>ATRN</i>	290	<i>FAM20B</i>
91	<i>SPATS2</i>	191	<i>TSPAN5</i>	291	<i>PENK</i>
92	<i>CHGA</i>	192	<i>PINK1</i>	292	<i>SLC2A4RG</i>
93	<i>UXS1</i>	193	<i>TBLIX</i>	293	<i>SPATS2L</i>
94	<i>NELLI</i>	194	<i>NEUROD6</i>	294	<i>LEMD3</i>
95	<i>LHX6</i>	195	<i>GFPT1</i>	295	<i>ADARB1</i>
96	<i>HERC1</i>	196	<i>ATP2B3</i>	296	<i>CAPRIN2</i>
97	<i>CUL1</i>	197	<i>ARID1A</i>	297	<i>MRPL44</i>
98	<i>MTHFD1</i>	198	<i>C6orf211</i>	298	<i>TAX1BP1</i>
99	<i>NDUFV1</i>	199	<i>MRPL9</i>	299	<i>KDM1A</i>
100	<i>SUCLG1</i>	200	<i>KCNJ6</i>	300	<i>TDRKH</i>

PPI network

According to the PPI data downloaded from STRING, the relationships between these 300 genes relevant to AD in the hippocampus were identified, and a PPI network was constructed (Figure 2). In total, the constructed PPI network included 184 nodes and 661 edges, where nodes refer to gene signatures and the edges between nodes indicate the interactions of genes in the network. From the network, we could observe that there was more than one interaction between some nodes, with reference to experimental interactions, domain fusion, interactions inferred by curator, interolog mapping, phylogenetic profiling, predictive text mining, and unspecified methods such as coexpression. When calculating the degree of each gene, we merged multiple edges between two nodes into one, and counted them as one edge in the following topology analysis. Genes from the PPI network showing a high degree included *NDUFS3* (degree = 23), *COPS5* (degree = 16), *SUCLG1* (degree = 15), *NDUFB3* (degree =

13), *NDUFV2* (degree = 13), *PPP1CA* (degree = 12), *IMP4* (degree = 11), *PSMA1* (degree = 11), *ADSL* (degree = 11), *MRPL15* (degree = 11), *ATP5J* (degree = 11), *NDUFA9* (degree = 11), *NDUFV1* (degree = 10), *SLC2A4RG* (degree = 10), and *CCT3* (degree = 10). These genes are likely to be crucial for maintaining the function and coherence of signaling mechanisms. In addition, the transcription factors (TFs) were chosen according to data for the known regulatory TFs (Vaquerizas et al., 2009). Six TFs were mapped on the PPI network, including *DMTF1*, *DEAF1*, *PRDM2*, *NEUROD6*, *MYT1L*, and *CDC5L*.

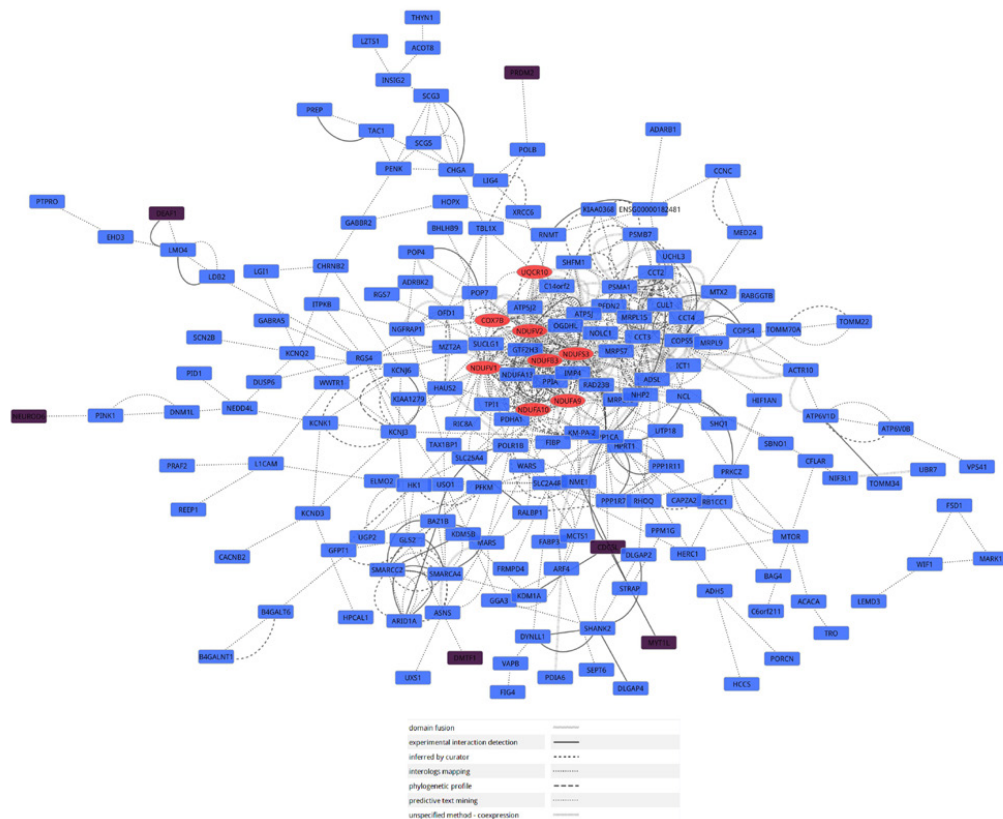


Figure 2. PPI network for the top 300 genes with differential expression in the hippocampus of patients with Alzheimer's disease. Purple refers to transcription factors, red refers to the genes common to cluster 1 and the KEGG pathways for oxidative phosphorylation, Parkinson's disease, Huntington's disease, and Alzheimer's disease.

Clusters

The clusters with densely connected nodes in the PPI network were detected using the ClusterONE plug-in of the Cytoscape software. In the current study, five clusters were identified with parameters set to a minimum size of 11 and a minimum density of 0.2. Figure 3 presents the five significant clusters selected from the top 300 genes. The number of nodes in each cluster was 17, 14, 11, 12, and 13, while the number of corresponding edges was 159, 46, 13, 30, and 16, respectively.

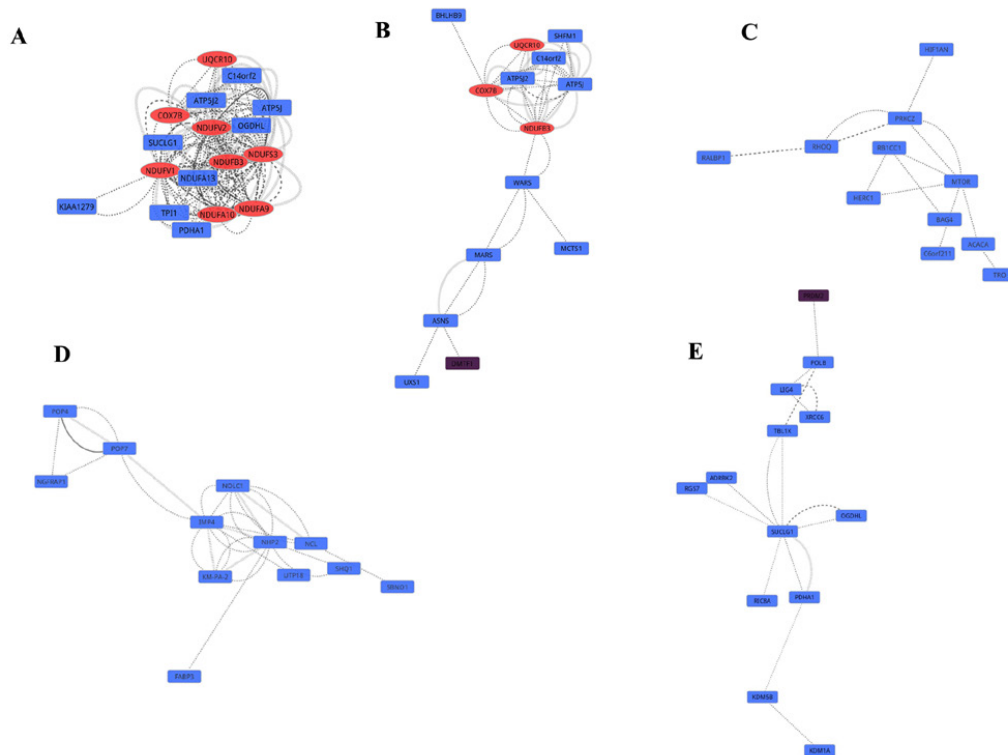


Figure 3. (A) Cluster 1, (B) cluster 2, (C) cluster 3, (D) cluster 4, and (E) cluster 5 of the top 300 genes with differential expression in the hippocampus of patients with Alzheimer's disease. Purple refers to transcription factors, red refers to the genes common to cluster 1 and the KEGG pathways for oxidative phosphorylation, Parkinson's disease, Huntington's disease, and Alzheimer's disease.

Functional enrichment analysis

Functional enrichment analysis of the resulting gene signature was performed, including GO enrichment and KEGG pathway enrichment. GO analysis was carried out in three categories, including biological processes (BP), molecular functions (MF), and cellular components (CC). The results of GO analysis show that the top 300 genes were significantly enriched in 97 terms. After discarding the terms with P values greater than 0.01, the most significant five GO terms of BP, MF, and CC are presented in Table 3. The most significant terms for BP, MF, and CC were the generation of precursor metabolites and energy ($P = 8.21 \times 10^{-7}$), NADH dehydrogenase activity ($P = 9.04 \times 10^{-5}$), and mitochondrial membrane ($P = 1.40 \times 10^{-6}$), respectively. Pathway enrichment analysis showed that these genes were significantly enriched in four terms. The most significant term was oxidative phosphorylation ($P = 2.05 \times 10^{-5}$), including the genes *NDUFB3*, *ATP5J2*, *UQCRI0*, *NDUFA9*, *NDUFV1*, *NDUFV2*, *COX7B*, *NDUFS3*, *NDUFA10*, *ATP6V1D*, *ATP6V0B*, and *ATP5J*. The other terms were Parkinson's disease ($P = 9.87 \times 10^{-5}$), Huntington's disease ($P = 0.0053$), and Alzheimer's disease ($P = 0.0095$), and all of these pathways are related to neuropathy. The four pathways are presented in Table 4.

Table 3. Significantly enriched GO terms for the hippocampus of patients with Alzheimer's disease.

Terms	P value	Genes
Biological process (BP)		
Generation of precursor metabolites and energy	8.21E-007	NDUFB3, TXNL1, SLC25A4, NDUFA9, SUCLG1, OGDHL, HK1, NDUFA13, PFKM, NDUFA10, ATP6V1D, ATP6V0B, PPP1CA, TP11, UQCR10, NDUFV1, GFPT1, NDUFV2, PDHA1, NDUFS3, ATP5J
Oxidative phosphorylation	6.17E-005	NDUFB3, UQCR10, NDUFA9, NDUFV1, NDUFV2, NDUFS3, NDUFA10, ATP6V1D, ATP6V0B, ATP5J
Mitochondrial ATP synthesis coupled electron transport	4.63E-004	NDUFB3, UQCR10, NDUFA9, NDUFV1, NDUFV2, NDUFS3, NDUFA10
ATP synthesis coupled to electron transport	4.63E-004	NDUFB3, UQCR10, NDUFA9, NDUFV1, NDUFV2, NDUFS3, NDUFA10
Mitochondrial electron transport, NADH to ubiquinone	8.55E-004	NDUFB3, NDUFA9, NDUFV1, NDUFV2, NDUFS3, NDUFA10
Molecular Function (MF)		
NADH dehydrogenase activity	9.04E-005	NDUFB3, NDUFA9, NDUFV1, NDUFV2, NDUFA13, NDUFS3, NDUFA10
NADH dehydrogenase (quinone) activity	9.04E-005	NDUFB3, NDUFA9, NDUFV1, NDUFV2, NDUFA13, NDUFS3, NDUFA10
NADH dehydrogenase (ubiquinone) activity	9.04E-005	NDUFB3, NDUFA9, NDUFV1, NDUFV2, NDUFA13, NDUFS3, NDUFA10
Oxidoreductase activity, acting on NADH or NADPH, with quinone or similar compound as acceptor	1.90E-004	NDUFB3, NDUFA9, NDUFV1, NDUFV2, NDUFA13, NDUFS3, NDUFA10
Oxidoreductase activity, acting on NADH or NADPH	0.003153	NDUFB3, NDUFA9, NDUFV1, NDUFV2, NDUFA13, NDUFS3, NDUFA10
Cellular Component (CC)		
Mitochondrial membrane	1.40E-006	HCCS, NDUFB3, ATP5J2, DNMI1, SLC25A4, NDUFA9, MTX2, SUCLG1, COX7B, HK1, NDUFA13, PINK1, NDUFA10, UQCR10, TOMM70A, NDUFV1, NDUFV2, TOMM22, SLC25A44, MTOR, NDUFS3, TOMM34, ATP5J
Mitochondrion	1.59E-006	NDUFB3, HCCS, MTX2, COX7B, OGDHL, SFXN3, ADH5, PINK1, HK1, C14ORF2, GLS2, MTHFD1, UQCR10, ACOT7, ACOT8, MRPL15, MRPL16, SLC25A44, PDHA1, NDUFS3, OXR1, ATP5J, TOMM34, ATP5J2, DNMI1, FIBP, SLC25A4, NDUFA9, SUCLG1, NDUFA13, MRPL9, REEP1, MRPS7, NDUFA10, CAPRN2, KIAA1279, TOMM70A, NDUFV1, NDUFV2, TOMM22, MTOR, MRPL44
Mitochondrial envelope	3.83E-006	HCCS, NDUFB3, ATP5J2, DNMI1, SLC25A4, NDUFA9, MTX2, SUCLG1, COX7B, HK1, NDUFA13, PINK1, NDUFA10, UQCR10, TOMM70A, NDUFV1, NDUFV2, TOMM22, SLC25A44, MTOR, NDUFS3, TOMM34, ATP5J
Mitochondrial part	4.64E-006	NDUFB3, HCCS, MTX2, OGDHL, COX7B, HK1, PINK1, GLS2, UQCR10, MRPL16, SLC25A44, PDHA1, NDUFS3, ATP5J, TOMM34, ATP5J2, DNMI1, SLC25A4, NDUFA9, SUCLG1, MRPL9, NDUFA13, NDUFA10, TOMM70A, NDUFV1, NDUFV2, TOMM22, MTOR
Mitochondrial membrane part	1.11E-005	NDUFB3, ATP5J2, TOMM70A, NDUFA9, NDUFV1, NDUFV2, COX7B, NDUFA13, TOMM22, NDUFS3, NDUFA10, ATP5J

Table 4. Significantly enriched KEGG pathways for the hippocampus of patients with Alzheimer's disease.

Term	Genes	P value
Oxidative phosphorylation	<i>NDUFB3, ATP5J2, UQCR10, NDUFA9, NDUFV1, NDUFV2, COX7B, NDUFS3, NDUFA10, ATP6V1D, ATP6V0B, ATP5J</i>	2.05E-005
Parkinson's disease	<i>NDUFB3, UQCR10, SLC25A4, NDUFA9, NDUFV1, NDUFV2, COX7B, PINK1, NDUFS3, NDUFA10, ATP5J</i>	9.87E-005
Huntington's disease	<i>NDUFB3, UQCR10, SLC25A4, NDUFA9, NDUFV1, NDUFV2, COX7B, NDUFS3, NDUFA10, ATP5J</i>	0.0053
Alzheimer's disease	<i>NDUFB3, UQCR10, NDUFA9, NDUFV1, NDUFV2, COX7B, NDUFS3, NDUFA10, ATP5J</i>	0.0095

Identification of the hub sub-network

For all the resulting clusters, the average rank value based on the P value and degree was calculated, and the cluster with the highest rank value and degree was considered a hub

sub-network. The results are shown in Table 5. Genes in cluster 1 with an average rank value of 5.61 and average degree of 8.94 were significantly connected. Compared with our earlier enrichment analysis, eight of the 17 genes in cluster 1 (*NDUFB3*, *NDUFA9*, *NDUFV1*, *NDUFV2*, *NDUFS3*, *NDUFA10*, *COX7B*, and *UQCRI0*) shared the four significant KEGG pathways mentioned above, further confirming the significance of cluster 1.

Table 5. Average rank value and degree for each cluster.

Cluster	1	2	3	4	5
Number of nodes	17.0	14.0	11.0	12.0	13.0
Rank (average)	5.61	5.60	4.95	5.09	5.19
Degree (average)	8.94	5.00	2.63	4.25	3.38

DISCUSSION

AD affects many people in older age, and is a complex disease with a pathogenesis that is unclear. The goal of this study was to identify important genes, sub-networks, and potential pathways involved in AD to contribute to the understanding of its molecular pathogenesis. We first combined the genes from microarray data using GWRS and GWGS methods, and then analyzed these genes using functional enrichment analysis, pathway enrichment analysis, PPI, and cluster analysis.

In our study, a total of 6994 genes were identified using a robust new model that enabled integrative analysis of multiple microarray datasets produced by different platforms and protocols. PPI network analysis revealed several gene signatures with high connectivity, which might be candidates for the pathogenesis of AD.

In the present study, *NDUFS3* (degree = 23) showed the highest degree in the network. Zhang et al. (2012) verified that *NDUFS3* was significantly upregulated in the oxidative phosphorylation and AD pathways in diabetic nephropathy. Meanwhile, *NDUFS3* had the highest connectivity in the constructed PPI network and was a common gene in the four significant pathways we identified in AD, supporting an association with AD pathogenesis. *NDUFV1* is the core subunit, along with *NDUFS3*, in human mitochondrial complex I, and this is considered the minimal assembly required for catalysis.

NDUFV1 encodes an enzyme subunit of NADH:ubiquinone oxidoreductase complex I, and requires one 4Fe-4S cluster and one flavin mononucleotide (FMN) molecule as cofactors. Defects in *NDUFV1* lead to several diseases, such as Leigh syndrome, a severe neurological disorder characterized by bilaterally symmetrical necrotic lesions in subcortical brain regions. *NDUFA10* and *NDUFA9* encode subunits that each bind one flavin adenine dinucleotide (FAD) cofactor and are considered accessory subunits of complex I that have no involvement in catalysis. *NDUFB3* is also an accessory subunit of complex I that is not believed to be involved in catalysis. A study of Liu et al. (2011) indicated that defects in *NDUFV2* are closely related to AD and other encephalopathies. *COX7B* is one of the nuclear-encoded polypeptide chains of cytochrome c oxidase (complex IV), the terminal oxidase in mitochondrial electron transport. *COX7B* is reportedly increased in AD brains and its over-expression in cells was shown to enhance amyloid-beta peptide (1-40) toxicity, which has been implicated in neuronal cell death in AD (Nagai et al., 2004). *UQCRI0* is a component of complex III and interacts with cytochrome c1. *ATP5J2* and *ATP5J* in complex V produce ATP from ADP in the presence of a proton gradient across the membrane, which is generated by electron transport complexes

of the respiratory chain. These genes exhibited a high degree in our network, showing their potential importance for the molecular pathogenesis of AD.

In the current study, four significant pathways (oxidative phosphorylation, Parkinson's disease, Huntington's disease, and Alzheimer's disease pathways) were detected by KEGG enrichment analysis of those top 300 genes. Among them, the most significant was the oxidative phosphorylation pathway, with the lowest P value of 2.05×10^{-5} . Oxidative phosphorylation is a pervasive and highly efficient metabolic pathway that provides energy for basal metabolism from ATP reformation in the mitochondria of cells (Rolfe and Brown, 1997). In this pathway, redox reactions occur via the transfer of electrons from electron donors to acceptors. The electron transport chain includes five main protein complexes (NADH dehydrogenase, succinate dehydrogenase, the cytochrome bcl complex, cytochrome c oxidase, and ATP synthase) and is considered the vital energy producer. Coincidentally, several genes that we identified, such as *NDUFB3*, *NDUFA9*, *NDUFV1*, *NDUFV2*, *NDUFS3*, and *NDUFA10* in cluster 1, are related genes that are associated with NADH dehydrogenase. During oxidative phosphorylation, some reactive oxygen species are produced simultaneously, such as superoxide and hydrogen peroxide, which lead to free radical propagation, cell damage, disease, and senescence. Previous studies have highlighted the relationship between oxidative phosphorylation and AD. According to a study by Sun et al. (2012), an energy deficiency in the brain might be the commonest etiological agent for AD. Shoffner (1997) also showed that functional decreases in the activity of enzymes involved in oxidative phosphorylation appeared to occur in AD and may be related to other neurodegenerative processes, which supports the concept that oxidative phosphorylation plays an important role in the pathophysiology of AD. The oxidative phosphorylation pathway is shown in Figure 4. The importance of the oxidative phosphorylation pathway in the present study was further confirmed by the KEGG pathway for AD (Figure 5). In the Parkinson's disease and Huntington's disease pathways, oxidative phosphorylation has also been identified as one of the most significant processes involved in these diseases. This suggests that the oxidative phosphorylation pathway is highly relevant to the pathogenesis of neurological disease.

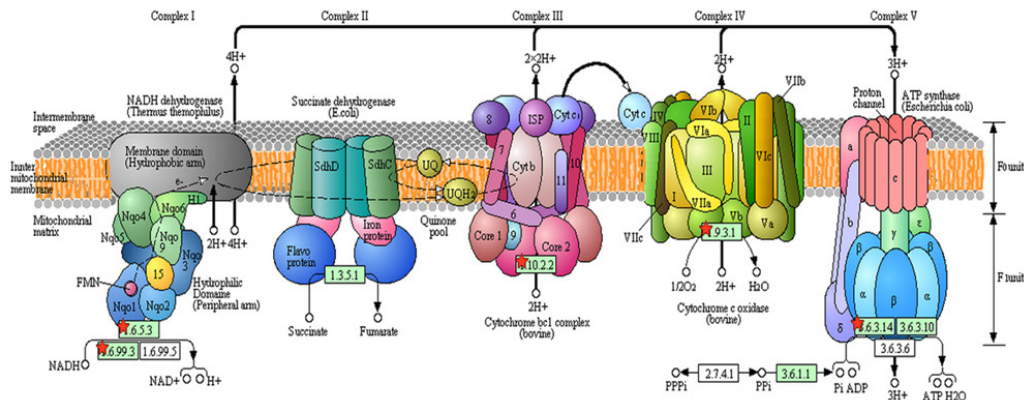


Figure 4. KEGG oxidative phosphorylation pathway. Genes involved in the oxidative phosphorylation pathway are shown. Red stars denote the detected genes.

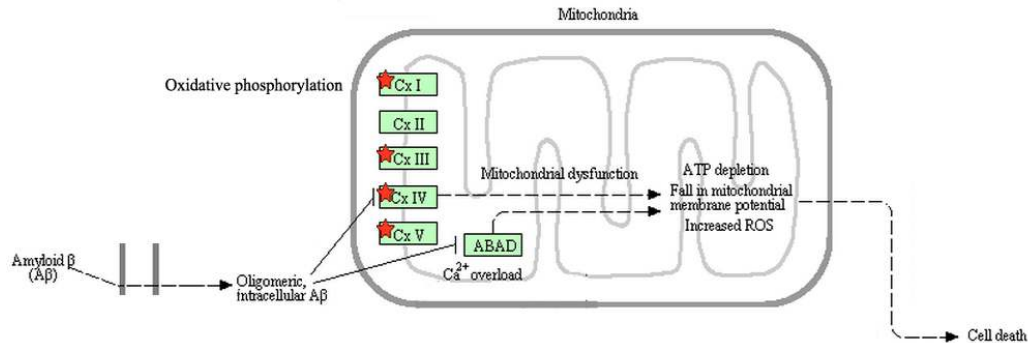


Figure 5. A part of the KEGG Alzheimer's disease pathway. Our detected genes were found in four respiratory chain complexes (Cx I, Cx III, Cx IV, and Cx V).

Using cluster analysis, we found five clusters with densely connected genes. Genes in cluster 1 are the most closely connected, with the highest values for rank and degree, and are considered the hub sub-network for AD. Moreover, eight genes in cluster 1 also participate in the significant pathways that we identified, including *NDUFB3*, *NDUFA9*, *NDUFV1*, *NDUFV2*, *NDUFS3*, *NDUFA10*, *COX7B*, and *UQCRI0*. Most of these genes showed a high degree in the network and some have a proven role in AD. This further confirms the significance of cluster 1 in AD.

In this paper, several hub genes were identified, and many of these genes have seldom been reported in previous research on AD. The bioprocesses and significant signaling pathways associated with AD were presented systematically. Comprehensive network analysis was conducted on the dysregulation of gene expression in AD, and a hub sub-network was shown for the development of AD. The identification of these pathways and several critical genes might give new insight into potential therapies for AD. However, the current study was based on previous reports. The results need to be confirmed by further research and more clinical evidence.

ACKNOWLEDGMENTS

Research received no specific grants from any funding agency in public, commercial, or not-for-profit sectors.

REFERENCES

- Babic T (1999). The cholinergic hypothesis of Alzheimer's disease: a review of progress. *J. Neurol. Neurosurg. Psychiatry* 67: 558.
- Barrett T, Troup DB, Wilhite SE, Ledoux P, et al. (2011). NCBI GEO: archive for functional genomics data sets - 10 years on. *Nucleic Acids Res.* 39: D1005-1010.
- Blalock EM, Geddes JW, Chen KC, Porter NM, et al. (2004). Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc. Natl. Acad. Sci. USA* 101: 2173-2178.
- Blalock EM, Buechel HM, Popovic J, Geddes JW, et al. (2011). Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. *J. Chem. Neuroanat.* 42: 118-126.
- Boada M, Tárraga L, Hernández I, Valero S, et al. (2014). Design of a comprehensive Alzheimer's disease clinic and research center in Spain to meet critical patient and family needs. *Alzheimers Dement.* 10: 409-415.

- Brookmeyer R, Johnson E, Ziegler-Graham K and Arrighi HM (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement.* 3: 186-191.
- Desai AK and Grossberg GT (2005). Diagnosis and treatment of Alzheimer's disease. *Neurology* 64: S34-S39.
- Duncan D, Prodduturi N and Zhang B (2010). WebGestalt2: an updated and expanded version of the Web-based Gene Set Analysis Toolkit. *BMC Bioinformatics* 11: 10.
- Freund T and Buzsáki G (1996). Interneurons of the hippocampus. *Hippocampus* 6: 347-470.
- Hokama M, Oka S, Leon J, Ninomiya T, et al. (2014). Altered expression of diabetes-related genes in Alzheimer's disease brains: the Hisayama study. *Cereb. Cortex* 24: 2476-2488.
- Hosack DA, Dennis G Jr., Sherman BT, Lane HC, et al. (2003). Identifying biological themes within lists of genes with EASE. *Genome Biol.* 4: R70.
- Huang DW, Sherman BT and Lempicki RA (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57.
- Jensen LJ, Kuhn M, Stark M, Chaffron S, et al. (2009). STRING 8 - a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 37: D412-416.
- Kanehisa M and Goto S (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28: 27-30.
- Liang D, Han G, Feng X, Sun J, et al. (2012). Concerted perturbation observed in a hub network in Alzheimer's disease. *PLoS One* 7: e40498.
- Liang WS, Dunckley T, Beach TG, Grover A, et al. (2007). Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain. *Physiol. Genomics* 28: 311-322.
- Liang WS, Dunckley T, Beach TG, Grover A, et al. (2008a). Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. *Physiol. Genomics* 33: 240-256.
- Liang WS, Reiman EM, Valla J, Dunckley T, et al. (2008b). Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. *Proc. Natl. Acad. Sci. USA* 105: 4441-4446.
- Liao YC, Lee WJ, Hwang JP, Wang YF, et al. (2014). ABCA7 gene and the risk of Alzheimer's disease in Han Chinese in Taiwan. *Neurobiol. Aging* 35: 2423.e7-13.
- Liu HY, Liao PC, Chuang KT and Kao MC (2011). Mitochondrial targeting of human NADH dehydrogenase (ubiquinone) flavoprotein 2 (NDUFV2) and its association with early-onset hypertrophic cardiomyopathy and encephalopathy. *J. Biomed. Sci.* 18: 29.
- Liu W, Peng Y and Tobin DJ (2013). A new 12-gene diagnostic biomarker signature of melanoma revealed by integrated microarray analysis. *Peer J.* 1: e49.
- Lopez OL, Becker JT, Sweet RA, Klunk W, et al. (2003). Psychiatric symptoms vary with the severity of dementia in probable Alzheimer's disease. *J. Neuropsychiatry Clin. Neurosci.* 15: 346-353.
- Ma L, Robinson LN and Towle HC (2006). ChREBP•Mlx is the principal mediator of glucose-induced gene expression in the liver. *J. Biol. Chem.* 281: 28721-28730.
- Nagai Y, Ogasawara A and Heese K (2004). Possible mechanisms of A beta(1-40)- or A beta(1-42)-induced cell death and their rescue factors. *Nihon Yakurigaku Zasshi.* 124: 135-143.
- Nepusz T, Yu H and Paccanaro A (2012). Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Methods* 9: 471-472.
- Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, et al. (2007). ArrayExpress - a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.* 35: D747-750.
- Pepper SD, Saunders EK, Edwards LE, Wilson CL, et al. (2007). The utility of MAS5 expression summary and detection call algorithms. *BMC Bioinformatics* 8: 273.
- Reisberg B, Borenstein J, Salob SP and Ferris SH (1987). Behavioral symptoms in Alzheimer's disease: phenomenology and treatment. *J. Clin. Psychiatry* 48: S9-15.
- Rifai N and Ridker PM (2001). Proposed cardiovascular risk assessment algorithm using high-sensitivity C-reactive protein and lipid screening. *Clin. Chem.* 47: 28-30.
- Rolfe DF and Brown GC (1997). Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol. Rev.* 77: 731-758.
- Scardoni G, Petterlini M and Laudanna C (2009). Analyzing biological network parameters with CentiScaPe. *Bioinformatics* 25: 2857-2859.
- Schoenberg BS, Kokmen E and Okazaki H (1987). Alzheimer's disease and other dementing illnesses in a defined United States population: incidence rates and clinical features. *Ann. Neurol.* 22: 724-729.
- Shoffner JM. 1997. Oxidative phosphorylation defects and Alzheimer's disease. *Neurogenetics* 1: 13-19.
- Sun J, Feng X, Liang D, Duan Y, et al. (2012). Down-regulation of energy metabolism in Alzheimer's disease is a protective response of neurons to the microenvironment. *J. Alzheimers Dis.* 28: 389-402.
- Thambisetty M, An Y and Tanaka T (2013). Alzheimer's disease risk genes and the age-at-onset phenotype. *Neurobiol. Aging* 34: 2696.e1-5.

- Thies W, Bleiler L and Alzheimer's Association (2013). 2013 Alzheimer's disease facts and figures. *Alzheimers Dement.* 9: 208-245.
- Vaquerizas JM, Kummerfeld SK, Teichmann SA and Luscombe NM (2009). A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* 10: 252-263.
- West MJ, Kawas CH, Martin LJ and Troncoso JC (2000). The CA1 region of the human hippocampus is a hot spot in Alzheimer's disease. *Ann. N. Y. Acad. Sci.* 908: 255-259.
- Yin HL, Wang YL, Li JF, Han B, et al. (2014). Effects of curcumin on hippocampal expression of NgR and axonal regeneration in A β -induced cognitive disorder rats. *Genet. Mol. Res.* 13: 2039-2047.
- Zhang Q, Xiao X, Li M, Li W, et al. (2012). Gene expression profiling in glomeruli of diabetic nephropathy rat. *Exp. Biol. Med.* 237: 903-911.