Progress in Biological Sciences
Vol. 6, Number 1, Winter / Spring 2016/37-46

Identification of Alzheimer disease-relevant genes using a novel hybrid method

Seyede Zahra Paylakhi[1], Sadjaad Ozgoli[1]*, Seyedhassan Paylakhi[2]

[1] *Systems, Life science and Control Engineering Lab., Electrical & Computer Engineering School, Tarbiat Modares University, Tehran, Iran.*

[2] *Department of Ophthalmology, University of California, San Francisco, CA, USA.*

Received: May 3, 2016; Accepted: October 6, 2016

ABSTRACT

Identifying genes underlying complex diseases/traits that generally involve multiple etiological mechanisms and contributing genes is difficult. Although microarray technology has enabled researchers to investigate gene expression changes, but identifying pathobiologically relevant genes remains a challenge. To address this challenge, we apply a new method for selecting the disease-relevant genes from a published microarray dataset. The approach is comprised of combination of fisher criteria, SAM (Significance Analysis for Microarrays), and GA/SVM (Genetic Algorithm/ Support Vector Machine). To get rid of noisy and redundant genes in high dimensional microarray data, the Fisher method is used. SAM technique is utilized and different subsets of highly informative genes are selected by GA/SVM which uses different training sets. The final subset, highly informative genes, is achieved by analyzing the number of times each gene occurs in the different gene subsets. The proposed method was tested on microarray data of Alzheimer's disease (AD) and the biological significance of identified genes was evaluated, and the results were compared with those of previous studies. The results indicate that the proposed method has a good selection and classification performance, which can produce 94.55 of classification accuracy by use of only 44 genes. From biological point of view, at least 24 (55%) of these genes are Alzheimer associated genes. Analysis of these genes by GO and KEGG led to identification of AD-related terms and pathways. These genes can act as predictors of the disease as well as a mean to find new candidate genes.

Keywords: Gene selection; Genetic algorithm; Support Vector Machine; Fisher method.

Introduction

Despite decades of research, the molecular mechanisms underlying many of complex diseases remain largely unknown. The identification of disease-related genes and pathways not only improves our understanding of the disease pathogenesis, but also offers the prospect of finding new therapeutic targets (1). Generally, two experimental approaches are used to identify disease-related genes: linkage analysis and association studies (1). However, the two methods have some drawbacks. For example, linkage analysis can lead to the chromosomal regions containing tens or even hundreds of genes and population-based association studies may lead to

* **Corresponding author:** ozgoli@modares.ac.ir

identification of a number of false positives (2, 3).

In recent years, the advent of microarray technology has enabled researchers to investigate gene expression changes that are associated with diseases (4). Although gene expression profiling is a powerful tool for assessing molecular changes associated with the disease, challenges remain (5, 6). One of the shortcomings of microarray data is that there are abundant genes, but fewer samples. Therefore, microarray data should be preprocessed by feature selection methods, and the genes with a lack of information should be disposed (7). This highly dimensional nature of microarray data along with the obvious need for dimension reduction, have given rise to a wealth of feature selection techniques. These methods are classified as filter and wrapper (8).

Filter methods benefit from the general features of training data, which are set in order to select significant features. They do not use a learning algorithm; in contrast, wrapper methods use a specific learning algorithm. These methods aim to select a subset of features, which reduces the errors associated with learning algorithms (9).

In the present paper, a new framework is proposed for gene selection which includes Fisher, SAM (Significance Analysis of Microarrays) and GA/SVM (Genetic Algorithm/Support Vector Machine) algorithms. In the initial stage, a Fisher criterion was used in order to eliminate the irrelevant and noisy genes. Then, SAM technique is applied for the purpose of raising the number of relevant, useful, and informative genes and of reducing redundant and irrelevant genes. In the final step, by the use of GA/SVM strategy, various training data sets are applied and several subsets of informative genes are determined. At last, the ultimate subset of genes is the result of the analysis of the number of times that each gene is present in the different subsets.

Materials and Methods

Microarray data

Microarray expression data (GSE1297) was downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). In this study, the authors examined hippocampal gene expression of nine control and 22 AD subjects on 31 separate microarrays (10). The array data were generated using Affymetrix Human Genome U133A Array. Gene expression data was converted into expression measures through Affymetrix® Expression Console™ Software, using gene-level RMA summarization.

Gene selection and classification

Step 1) Fisher criterion: A Fisher criterion is a form of filtering method. In this model, the gene rankings are shown by the following equation:

$$\frac{(m_1(g) - m_2(g))^2}{(s_1^2(g) + s_2^2(g))}$$

Here $m_1(g)$ and $m_2(g)$ are indicators of means of gene (g) expression in the cancerous and normal samples, respectively. $s_1(g)$ and $s_2(g)$ signify standard deviations of gene (g) across the cancerous and normal samples, respectively. In the present technique, core higher values are allocated to those features with very different means between the two classes, relative to their variances. Those features which have the highest scores are regarded as the most discriminatory features (11).

Step 2) SAM: The Significance Analysis for Microarrays (SAM) approach was proposed for high dimensional microarray data (9). It was an attempt to identify genes with notable changes in their expression, assimilating a set of gene-specific t-tests. To measure gene-specific fluctuations, SAM defines relative difference measure for the i_th gene as follows:

$$d(i) = \frac{xP_i - xN_i}{s_i + s_0}$$

where xP_i and xN_i are the average levels of expression of gene i directly related to the groups P and N, one at a time. The s_i in the denominator shows the gene-specific scatter which is defined by

$$S_i = \frac{|P| + |N|}{|P||N|(|P| + |N| - 2)} \left(\sum_{k \in P} \left[x_k(i) - x_p(i)\right]^2 + \sum_{k \in N} [x_k(i) - x_N(i)]^2\right.$$

In order to make the variance of $d(i)$ independent of gene expression, the parameter s_0 is assigned.

Step 3) GA/SVM:

GA (Genetic Algorithm): GA is usually applied for optimization problems which could take discontinuous or continuous values. The ultimate goal of GA is discovering best and perfect solution within a group of potential solutions. Each set of solutions is named as population. Populations are comprised of vectors, that is to say, chromosome or individual. Each item in the vector is known as gene (12).

SVM (Support Vector Machine): SVM is a kind of supervised machine learning technique. It is a type of method which has a powerful effect against the dispensed noisy data. It has been demonstrated that perform well in many subjects of biological analysis like evaluation of microarray data expression (13).

This training data follows a form $(x_1, y_1), \ldots, (x_n, y_n), x \in R^N$ and $Y \in \{+1, -1\}$. Each data is formed with N dimensional vector and belonging only one of two classes $(+1\ or\ -1)$. Two classes are kept apart from each other by hyper planes in order to provide following forms for all training data. Thus,

$$(w.x_i + w_0) \geq 1 - y_i \qquad if \qquad y_i{=}{+}1$$

$$(w.x_i + w_0) \leq -1 - y_i, \qquad if \qquad y_i = -1$$

Or

$$y_i[w.x_i + w_0)] \geq 1 \qquad i = 1, \ldots, n$$

Here, $y_i \geq 0$ points to slack variables. In fact, it is used to provide a tolerance to some data with small error. If all data satisfy in above relations correctly, P variables will not be used. Optimal hyper plane among all hyperplanes is found by minimizing following formula

$$c\sum_{i=1}^{n} y_i + \frac{1}{2}\|w\|^2$$

In this formula, c indicates a regularization parameter and provides an example of a trade-off between complexity and classification performance. In other words, the margin is maximized by optimal separating hyper plane. Problem is changed into following dual form of quadratic optimization problem.

$$Maximize \qquad w(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,k=1}^{n} \alpha_i\, y_i \alpha_k y_k (x_i, x_k),$$

$$Subject\ to \qquad \sum_{i=1}^{n} \alpha_i\, y_i = 0 \qquad \alpha_i \geq \forall i$$

According to α_i Lagrange multipliers computed in above-mentioned formula, following decision function is built (14).

$$f(x) = sign\left(\sum_{i=1}^{sv} (x, x_i) + b\right)$$

In order to illustrate the basic idea of the operation of our method, in Figure 1, we show a simple scheme of how features are extracted and classified from the initial microarray dataset.

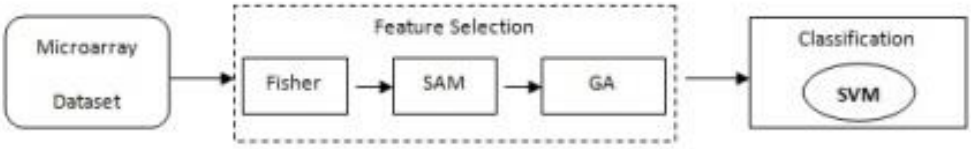

Figure 1. A simple plan of how genes are selected out from the microarray dataset using the suggested method. First, a feature selection method is utilized to removeredundant and noisy genes. Next, the resulted subset is estimated by means of aSVM classifier.

Biological evaluation of selected genes

We tried to address the biological importance of the identified genes through various approaches. First, the identified genes were compared to other AD-related studies to find previously reported genes. To do this, we performed a keyword search on PubMed (http://www.ncbi.nlm.nih.gov/pubmed). Second, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics resource (15) to identify enriched functional pathways and

categories, or gene ontology terms, annotated by KEGG (16) and GO (17). A cutoff p value of 0.05 was used for enriched KEGG pathways or GO functions.

Results

In this paper, the software "Rapid minder 5®" was used for implementation of the SAM and GA/SVM algorithms. After preprocessing of the data, the proposed algorithms were implemented on the data set in order to recognize the discriminatory genes. To evaluate the classification accuracy, 10 folds cross validation (10-CV) was accomplished using SVM Classifier.

Accuracy is the fraction of correctly classified samples over all samples.

$$accuracy = \frac{TN + TP}{TN + FN + FP + TP}$$

Here, TN is a negative sample where the negativity is predicted truly. TP is a positive sample where the positivity is predicted truly. FN is a positive sample which, it is predicted negative as false and FP is a negative sample predicted in a false manner (18).

The results of the average accuracy of the different methods are shown in Table 1. As we can see, when a gene selection method is used before classification, the performance of the classifier improves. Therefore, we reach the highest accuracy with the minimum number of genes.

Table 1. Comparison of different gene selection methods based on accuracy

Methods	Accuracy	Number of identified genes	Ratio of previously reported genes/ identified genes
M1	48.39	12,990	NA*
M2	67.1	6,786	NA*
M3	72.97	122	0.31
M4	94.55	44	0.55

M1: classification without gene selection.
M2: classification after using Fisher criterion for gene selection.
M3: classification after using Fisher criterion with SAM for gene selection
M4: classification after using combination of Fisher criterion and SAM with GA/SVM algorithm for gene selection.
* We did not investigate previously reported genes among identified genes.

As it mentioned before, the software used for applying the method is Rapid minder 5. Carrying out of SAM is achieved by the use of weight by SAM operator, and operators Optimize by Generation (GGA) and Support Vector Machine are used for GA/SVM method equivalently.

Figure 2 shows the connection between the number of selected genes and the equivalent prediction accuracy by use of SVM on dataset. As demonstrated in Figure 2, the best result achieved on Alzheimer's disease dataset is 94.55% accuracy by using only 44 genes. Table 2 shows the names of identified genes. Interestingly, 55% of these genes (24 of 44) have been reported in previous AD-related studies (Table 2).

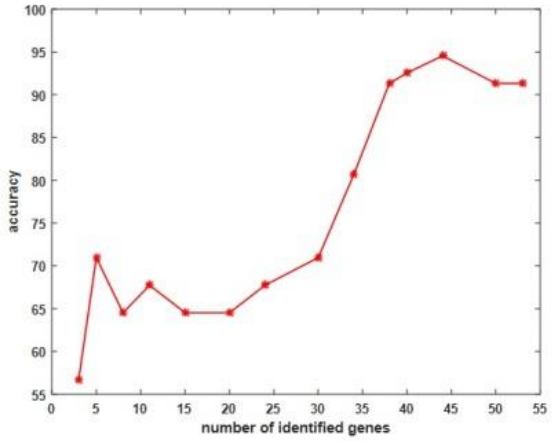

Figure 2. Testing accuracy of classification on dataset. This figure shows the connection between the number of selected genes and the equivalent prediction accuracy by use of SVM on dataset.

Table 2. Forty four identified genes in the present study

Gene Name	Gene Symbol	Gene ID
Acyl-CoA dehydrogenase family member 8	ACAD8	27034
Stimulator of chondrogenesis 1	SCRG1	11341
GLI-Kruppel family member GLI1	GLI1	14632
Tryptophan rich basic protein	WRB	7485
Interleukin 1 receptor associated kinase 4	IRAK4	51135
Kinase D-interacting substrate 220kDa	KIDINS220	57498
Heterogeneous nuclear ribonucleoprotein D	HNRNPD	3184
A-kinase anchoring protein 13	AKAP13	11214
Fatty acid binding protein 7, brain	FABP7	12140
Torsin family 1 member A	TOR1A	1861
Nuclear factor of activated T cells, cytoplasmic, calcineurin dependent 4	NFATC4	73181
Ornithine aminotransferase	OAT	4942
Cullin-associated and neddylation-dissociated 1	CAND1	55832
Fibroblast growth factor 20	FGF20	26281
CREB binding protein	CREBBP	1387
Mitogen-activated protein kinase kinase 3	MAP2K3	5606
Charcot-Leyden crystal galectin	CLC	1178
COP9 signalosome subunit 5	COPS5	10987
Cytoplasmic linker associated protein 2	CLASP2	23122
Epidermal growth factor receptor pathway substrate 15	EPS15	2060
RB1 inducible coiled-coil 1	RB1CC1	9821
GRB2-related adaptor protein	GRAP	10750
Adenomatosis polyposis coli	APC	11789
Heat shock protein family B (small) member 11	HSPB11	51668
Aconitase 2	ACO2	50
Fibroblast growth factor receptor 1	FGFR1	2260
Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma	PIK3CG	5294
Secretogranin III	SCG3	29106
Casein kinase 1 gamma 3	CSNK1G3	1456
Ubiquitin specific peptidase 19	USP19	10869
Fetuin B	FETUB	26998
Mitochondrial ribosomal protein L15	MRPL15	29088
MON1 homolog B, secretory trafficking associated	MON1B	22879
Collagen type V alpha 2	COL5A2	1290
LMBR1 domain containing 1	LMBRD1	55788
CD2 (cytoplasmic tail) binding protein 2	CD2BP2	10421
wingless-type MMTV integration site family, member 5B	WNT5B	22419
Mitogen-activated protein kinase kinase kinase 3	MAP3K3	4215
Pyruvate dehydrogenase (lipoamide) beta	PDHB	5162
Homeobox B1	HOXB1	15407
BTB domain and CNC homolog 2	BACH2	60468
Killer cell immunoglobulin like receptor, three Ig domains X1	KIR3DX1	90011
Ubiquitin-fold modifier 1	UFM1	51569
Hepatitis b virus x-interacting protein-like protein	HBXIP	100528355

Analysis of the 44 genes using DAVID for identifying enriched GO terms led to the identification

of many terms that achieved a significant p-value within the following GO categories: biological process, molecular function and cellular component. This yielded 13 terms for biological process, five for molecular function, and three for cellular component (Table 3). The main biological processes were MAPKKK cascade, intracellular signalling cascade, regulation of microtubule depolymerisation and Wnt receptor signalling pathway. Cellular component-enriched terms were related to the chromosome (kinetochore and centromeric region) and microtubule. Among the five molecular function terms, two of them (40%) are related to microtubule related terms. Among the 10 enriched KEGG pathways (Table 4), six of them are directly involved in the pathogenesis of AD. Neurotrophin, Wnt and MAPK signalling pathways are most interesting pathways that we will discuss more under "Discussion".

Table 3. Gene ontology terms enriched by identified genes

Category	Term	Gene Annotation	Number of Genes	P-Value
Biological Process	GO:0000165	MAPKKK cascade	4	0.0118
Biological Process	GO:0007242	intracellular signaling cascade	9	0.0131
Biological Process	GO:0007243	protein kinase cascade	5	0.0148
Biological Process	GO:0010604	positive regulation macromolecule metabolic process	7	0.0213
Biological Process	GO:0045333	cellular respiration	3	0.026
Biological Process	GO:0045333	protein amino acid phosphorylation	6	0.0274
Biological Process	GO:0045893	positive regulation of transcription, DNA-dependent	5	0.0337
Biological Process	GO:0051254	positive regulation of RNA metabolic process	5	0.0346
Biological Process	GO:0007026	negative regulation of microtubule depolymerization	2	0.0406
Biological Process	GO:0031114	regulation of microtubule depolymerization	2	0.0406
Biological Process	GO:0031111	negative regulation of microtubule polymerization or depolymerization	2	0.0431
Biological Process	GO:0016055	Wnt receptor signaling pathway	3	0.0462
Biological Process	GO:0001501	skeletal system development	4	0.0488
Cellular Component	GO:0000776	kinetochore	3	0.0114
Cellular Component	GO:0000775	chromosome, centromeric region	3	0.028
Cellular Component	GO:0005881	cytoplasmic microtubule	2	0.0374
Molecular Function	GO:0016563	transcription activator activity	6	0.0026
Molecular Function	GO:0008017	microtubule binding	3	0.0127
Molecular Function	GO:0004672	protein kinase activity	6	0.0135
Molecular Function	GO:0004674	protein serine/threonine kinase activity	5	0.0184
Molecular Function	GO:0015631	tubulin binding	3	0.0236

Table 4. Biological pathways enriched by KEGG analysis of identified genes

Pathway	Number of Genes	Gene	p-Value
Pathways in cancer (hsa05200)	7	PIK3CG, FGFR1, WNT5B, CREBBP, FGF20, GLI1, APC	0.0009
Neurutrophin signaling pathway (hsa04722)	4	IRAK4, PIK3CG, MAP3K3, KIDINS220	0.010309
MAPK signaling pathway (hsa04010)	5	FGFR1, MAP3K3, MAP2K3, NFATC4, FGF2	0.015412
Basal cell carcinoma (hsa05217)	3	WNT5B, GLI1, APC	0.017461
Wnt signaling pathway (hsa04310)	4	WNT5B, CREBBP, NFATC4, AP	0.017543
Hedgehog signaling pathway (hsa04340)	3	WNT5B, CSNK1G3, GLI1	0.018067
Melanoma (hsa05218)	3	PIK3CG, FGFR1, FGF20	0.028202
Prostate cancer (hsa05215)	3	PIK3CG, FGFR1, CREBBP	0.042711
Regulation of actin cytoskeleton (hsa04810)	4	PIK3CG, FGFR1, FGF20, AP	0.043808
Toll_like receptor signaling pathway (hsa04620)	3	IRAK4, PIK3CG, MAP2K3	0.053643

Discussion

The objective of this study was to introduce a novel method for identifying disease-relevant genes from microarray data sets. Due to the complexity of many common diseases and the asynchrony of pathogenic events, we may miss key changes that are important in disease manifestation, but not uniform within assessed groups, by simple analysis of the expression data (3). With these considerations in mind, we proposed a three-stage selection algorithm of hybridizing the fisher and SAM filter methods and GA/SVM, as a wrapper method for addressing the gene selection problem. According to the results obtained, the performed method greatly decreases the number of features from thousands to 44 and the accuracies are improved to nearly 95%.

The results of the proposed method are compared with those of other similar studies in order to evaluate the proposed method. In one study, three methods of RF (Random Forest), IG (Information Gain), and GA (Genetic Algorithm) were implemented on AD microarray data and then the functions of these three methods were compared (19). In another one, the t-test NMSE (Normalized Mean Square Error) method was also implemented on the AD microarray data (20). Table 5 shows the comparison of the findings of the above-mentioned articles and the results of the current study. As it is clear from Table 5, the proposed method has higher accuracy and shows higher ratio of previously reported genes/identified genes than RF, IG, and GA methods. In addition, although the accuracy of proposed method is a little lower than that of the t-test NMSE method, the proposed method shows much higher ratio of previously reported genes/identified genes than that of the t-test NMSE method. It is reasonable to assume that the higher the ratio, the more biologically relevant the genes to the disease.

TABLE 5. Comparison of proposed method with other methods in Alzheimer gene identification

Methods	Accuracy	Number of identified genes	Ratio of previously reported genes/ identified genes
GA + SVM	91	20	0.15
IG + SVM	87	20	0.15
RF + SVM	87	20	0.15
t-test NMSE + SVM	98.3	20	0.25
Proposed Method	94.55	44	0.55

Analysis of the 44 genes by GO led to the identification of many Alzheimer related terms.

Among the seven cellular component-enriched terms, two of them are related to the kinetochore and centromeric region of the chromosome that play important roles in pathogenesis of AD. It has been shown that the presenilins, in which the mutations account for most early-onset familial AD, are localized in the nuclear membrane, its associated interphase kinetochores, and the centrosomes (21). Level of Sgt1, a protein required for the mitotic activity of kinetochore, significantly decreases in the temporal (up to 25-fold), angular (up to 11-fold) and posterior cingulate cortex (up to five-fold) of AD patients compared to control group (22).

Analysis of the 44 identified genes by KEGG led to the identification of some important biological pathways that achieved a significant p-value ($p \leq 0.05$). Among these pathways neurutrophin signalling pathway ($p = 0.007$), MAPK signaling pathway (0.01) and Wnt signalling pathway ($p = 0.03$) are noteworthy. There is a wealth of evidence to show that the neurotrophins are involved in cognitive processes in rats, monkeys, and humans (23-25). There are four genes in this pathway: IRAK4, PIK3CG, MAP3K3, KIDINS220. The fact that this pathway is important in cognitive processes and that there is evidence of involvement of IRAK4 and KIDINS220 in AD progression (26, 27), it is reasonable to speculate that two other genes of this pathway are also important in the disease pathogenesis.

There are five genes in MAPK signalling pathway: FGFR1, MAP3K3, MAP2K3, NFATC4, and FGF20. There are direct or indirect connections between all of these genes and AD (28-32). Since the MAP3K3 gene is common in both pathways we discuss more about this gene. MAP3K3 gene product is a 626-amino acid polypeptide that is 96.5% identical to mouse Mekk3 (33). Although there is no direct experimental evidence showing role of this gene in the disease but there is some evidence about the possible role of this gene in AD pathogenesis. Liang and colleagues in an attempt to demonstrate the robust and specific perturbation of the hub network in AD by analysing several datasets, reported MAP3K3 gene among genes strongly correlated with AD progression (29). Boxall and colleague re-annotated the interacting partners of the neuronal scaffold protein DISC1, one of the candidate genes for schizophrenia, using a knowledge-based approach (34). They revealed two highly connected networks and MAP3K3 was among the hub proteins in one of these networks. Although they initially aimed to find biological commonality between Huntington's disease and of schizophrenia but, based on the list of identified genes, they suggested there is a degree of biological commonality between other neurological disorders (34).

In addition to the genes which have been reported in other AD related studies, some other genes identified in this study are of particular interest. Here, we do not intend to discuss about all the genes but we will discuss about most important ones, USP19 and UFM1. USP19 (Ubiquitin Specific Peptidase 19) is a protein coding gene which codes a deubiquitinating enzyme that plays roles in the degradation of various proteins (35). UFM1 (Ubiquitin-Fold Modifier 1) codes a protein that is conjugated to target proteins by E1-like activating enzyme UBA5 and E2-like conjugating enzyme UFC1 in a manner similar to ubiquitylation (36). The formation of amyloid fibers, the hallmark structures of AD, could be the product of ubiquitin-mediated protein degradation defects and recent studies have also indicated that components of the USP could be linked to the early phase of AD (37, 38). Given these well-known facts, it is very tempting to suggest these genes as potential candidate genes for AD.

In conclusion, the promising implementation of the fisher, SAM and GA/SVM algorithm method on AD, a complex and late-onset disease, suggests that this strategy could be useful for identifying new candidate genes and therapeutic targets in other diseases. Multiple genes are introduced here as candidates for a better understanding of AD pathogenesis. The observation that experimental data has already implicated many of the identified genes in brain functions supports the premise that the remaining genes are also prime candidates for further study.

REFERENCES

1. Botstein, D. and Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.*, **33**, 228-237.
2. McCarthy, M.I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. and Hirschhorn, J. N. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.*, **9**, 356-369.3.
3. Wu, X., Jiang, R., Zhang, M. Q. and Li, S. (2008) Network-based global inference of human disease genes. *Mol. Syst. Biol.*, **4**, 189.
4. Tanaka, F., Ikenaka, K., Yamamoto, M. and Sobue, G. (2012) Neuropathology and omics in motor neuron diseases. *Neuropathology*, **32**, 458-462.
5. Jakobs, T. C. (2014). Differential gene expression in glaucoma. *Cold Spring Harbor Perspect. Med.*, **4**, a020636-a020636.
6. Paylakhi, S. H., Yazdani, S., April, C., Fan, J. B., Moazzeni, H., Ronaghi, M. and Elahi, E. (2012) Non-housekeeping genes expressed in human trabecular meshwork cell cultures.
7. Saeys, Y., Inza, I. and Larrañaga, P. (2007) A review of feature selection techniques in bioinformatics. *Bioinformatics*, **23**, 2507-2517.
8. Canedo, V. B. and Marono, N. S. (2014) Novel feature selection methods for high dimensional data (Doctoral dissertation, Universidade da Coruña).
9. Abou-Taleb, A. S., Mohamed, A. A., Mohamed, O. A. and Abedelhalim, A. H. (2013) Hybridizing filters and wrapper approaches for improving the classification accuracy of microarray dataset. *IJSCE*, **3**, 155-159.
10. Blalock, E. M., Geddes, J. W., Chen, K. C., Porter, N. M., Markesbery, W. R. and Landfield, P. W. (2004) Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 2173-2178.
11. Mohammadi, A., Saraee, M. H. and Salehi, M. (2011) Identification of disease-causing genes using microarray data mining and gene ontology. *BMC Med. Genomics*, **4**, 1.
12. Babaoglu, İ., Findik, O. and Ülker, E. (2010) A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine. *Expert Syst. Appl.*, **37**, 3177-3183.
13. Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M. and Haussler, D. (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, **16**, 906-914.
14. Law, M. (2006) A simple introduction to support vector machines. *Lecture for CSE*, Course no. 802. Michigan State University, Michigan, USA.
15. Sherman, B. T., Huang, D. W., Tan, Q., Guo, Y., Bour, S., Liu, D. and Lempicki, R. A. (2007) DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. *BMC Bioinformatics*, **8**, 426.
16. Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27-30.
17. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M. and Harris, M. A. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25-29.
18. Zhu, W., Zeng, N. and Wang, N. (2010) Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. NESUG proceedings: health care and life sciences, Baltimore, Maryland, 1-9.
19. Li, J., Xu, M., Zhou, H., Ma, J. and Potter, H. (1997) Alzheimer presenilins in the nuclear membrane, interphase kinetochores, and centrosomes suggest a role in chromosome segregation. *Cell*, **90**, 917-927.

20. Spiechowicz, M., Bernstein, H. G., Dobrowolny, H., Leśniak, W., Mawrin, C., Bogerts, B. and Filipek, A. (2006) Density of Sgt1-immunopositive neurons is decreased in the cerebral cortex of Alzheimer's disease brain. *Neurochem. Int.*, **49**, 487-493.
21. Zhang, L., Wang, L., Tian, P. and Tian, S. (2015) Pathway-based feature selection algorithms identify genes discriminating patients with multiple sclerosis apart from controls. *Cornell University Library*, ArXivPrepr. ArXiv150801509. https://arxiv.org/ftp/arxiv/papers/ 1508/1508.01509.pdf
22. Wang, R., Wang, H., Carrera I., Xu, S. and Lakshmana, M. K. (2015) COPS5 protein overexpression increases amyloid plaque burden, decreases spinophilin-immunoreactive Puncta, and exacerbates learning and memory deficits in the mouse brain. *J. Biol. Chem.*, **290**, 9299–9309.
23. Mandel, R. J., and Thal, L. J. (1988) Physostigmine improves water maze performance following nucleus basalismagnocellularis lesions in rats. *Psychopharmacol.*, **96**, 421-425.
24. Ridley, R. M., Barratt, N. G. and Baker, H. F. (1984) Cholinergic learning deficits in the marmoset produced by scopolamine and ICV hemicholinium. *Psychopharmacol*., **83**, 340-345.
25. Crow, T. J. and Grove-White, I. G. (1973) An analysis of the learning deficit following hyoscine administration to man. *Br. J. Pharmacol.*, **49**, 322-327.
26. Cameron, B., Tse, W., Lamb, R., Li, X., Lamb, B. T. and Landreth, G. E. (2012) Loss of interleukin receptor-associated kinase 4 signaling suppresses amyloid pathology and alters microglial phenotype in a mouse model of Alzheimer's disease. *J. Neurosci.*, **32**, 15112-15123.
27. López-Menéndez, C., Gamir-Morralla, A., Jurado-Arjona, J., Higuero, A. M., Campanero, M. R., Ferrer, I. and Iglesias, T. (2013) Kidins220 accumulates with tau in human Alzheimer's disease and related models: modulation of its calpain-processing by GSK3β/PP1 imbalance. *Hum. Mol. Genet.*, **22**, 466-482.
28. MU, L. and YANG, N. (2005) Gene expression profile of amyloid beta proteininjected mouse model for Alzheimer disease. *Acta Pharmacol. Sin.*, **26**, 666-672.
29. Liang, D., Han, G., Feng, X., Sun, J., Duan, Y. and Lei, H. (2012) Concerted perturbation observed in a hub network in Alzheimer's disease. PLoS One, **7**, e40498.
30. Zhou, F., Xu, Y. and Hou, X. Y. (2014) MLK3-MKK3/6-P38MAPK cascades following N-methyl-D-aspartate receptor activation contributes to amyloid-β peptide-induced apoptosis in SH-SY5Y cells. *J. Neurosci. Res.*, **92**, 808-817.
31. Hudry, E., Wu, H. Y., Arbel-Ornath, M., Hashimoto, T., Matsouaka, R., Fan, Z. and Hyman, B. T. (2012) Inhibition of the NFAT pathway alleviates amyloid beta neurotoxicity in a mouse model of Alzheimer's disease. *J. Neurosci.*, **32**, 3176-3192.
32. Alemany, S., Ribasés, M., Vilor-Tejedor, N., Bustamante, M., Sánchez-Mora, C., Bosch, R. and Sunyer, J. (2015) New suggestive genetic loci and biological pathways for attention function in adult attention-deficit/hyperactivity disorder. *Am. J. Med. Genet., Part B*, **168**, 459-470.
33. Nakamura, K., Kimple, A. J., Siderovski, D. P. and Johnson, G. L. (2010) PB1 domain interaction of p62/sequestosome 1 and MEKK3 regulates NF-κB activation. *J. Biol. Chem.*, **285**, 2077-2089.
34. Boxall, R., Porteous, D. J. and Thomson, P. A. (2011) DISC1 and Huntington's disease–overlapping pathways of vulnerability to neurological disorder. PLoS One, **6**, e16263.
35. Wiles, B., Miao, M., Coyne, E., Larose, L., Cybulsky, A. V. and Wing, S. S. (2015) USP19 deubiquitinating enzyme inhibits muscle cell differentiation by suppressing unfolded-protein response signaling. *Mol. Biol. Cell*, **26**, 913-923.
36. Daniel, J. and Liebau, E. (2014) The ufm1 cascade. Cells, **3**, 627-638.
37. Riederer, B. M., Leuba, G., Vernay, A. and Riederer, I. M. (2011) The role of the ubiquitin proteasome system in Alzheimer's disease. *Exp. Biol. Med.*, **236**, 268-276.
38. Upadhya, S. C. and Hegde, A. N. (2007). Role of the ubiquitin proteasome system in Alzheimer's disease. *BMC Biochem.*, **8**, S12.