Three-way interaction model to trace the mechanisms involved in Alzheimer’s disease transgenic mice

Nasibeh Khayer¹, Sayed-Amir Marashi²,³ *, Mehdi Mirzaie⁴, Fatemeh Goshadrou¹

¹ Department of Basic Sciences, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran, ² Department of Biotechnology, College of Science, University of Tehran, Tehran, Iran, ³ School of Biological Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran, ⁴ Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran

* marashi@ut.ac.ir

Abstract

Alzheimer’s disease (AD) is the most common cause for dementia in human. Currently, more than 46 million people in the world suffer from AD and it is estimated that by 2050 this number increases to more than 131 million. AD is considered as a complex disease. Therefore, understanding the mechanism of AD is a universal challenge. Nowadays, a huge number of disease-related high-throughput “omics” datasets are freely available. Such datasets contain valuable information about disease-related pathways and their corresponding gene interactions. In the present work, a three-way interaction model is used as a novel approach to understand AD-related mechanisms. This model can trace the dynamic nature of co-expression relationship between two genes by introducing their link to a third gene. Apparently, such relationships cannot be traced by the classical two-way interaction model. Liquid association method was applied to capture the statistically significant triplets which are involved in three-way interaction. Subsequently, gene set enrichment analysis (GSEA) and gene regulatory network (GRN) inference were applied to analyze the biological relevance of the statistically significant triplets. The results of this study suggest that the innate immunity processes are important in AD. Specifically, our results suggest that H2-Ob as the switching gene and the gene pair {Csf1r, Milr1} form a statistically significant and biologically relevant triplet, which may play an important role in AD. We propose that the homeostasis-related link between mast cells and microglia is presumably controlled with H2-Ob expression levels as a switching gene.

Introduction

Alzheimer’s disease (AD) is the most common cause for dementia in human. AD often affects individuals over the age of 60, with increasing risk by age. Based on the last report of the AD International, more than 46 million people in the world suffer from AD and it is estimated that by 2050 this number increases to more than 131 million [1]. AD is associated with a
complex progression of neurodegeneration that results in memory impairment and loss of other cognitive processes, as well as the presence of non-cognitive symptoms including delusions, agitation and changes in behavior and personality [2]. Although the causes of AD and its progress are not well known yet, it is characterized by the presence of amyloid plaques and neurofibrillary tangles outside and inside the neuronal cells, respectively. Oxidative stress, chronic inflammation, gliosis and excitotoxicity are other pathologic symptoms of this disease. Reduction in neuronal synapses following the cellular death in hippocampus area of brain, can be the reason for loss of spatial memory and anomaly in patients’ behavior [3].

AD pathogenicity procedure is complex. Without a proper understanding of the pathogenicity mechanisms, treatments will only lead to decreasing the disease symptoms and attacks, while the main therapy is not yet available. Therefore, finding pathogenicity mechanisms is a universal concern [4,5].

High-throughput data produced under different biological conditions, includes a large amount of information about gene regulation processes and protein functions. In addition, depending on the applied computational method, different conclusions can be drawn from the same biological data [6–10].

A common way for analyzing functionally-related genes is the use of two-way interaction model, in which expression level of two genes are measured according to their pairwise correlation value [11,12]. Naturally, genes with high correlation tend to be functionally-related. However, because of various reasons, some functionally-related gene pairs may not show a significant correlation. Therefore, such a gene pair may not be discovered with a two-way interaction model. For example, the co-expression relation of two functionally-related genes may change based on the cellular state [13]. Furthermore, some proteins have multiple molecular roles, and therefore, such proteins can be involved in several biological procedures [13]. In particular, some co-expression relations can only be seen in disease condition, but not in healthy tissues [14,15]. Thus, the two-way interaction model may be too simplistic to explain some of the complex molecular relations [16].

In a few previous studies, it had been suggested that classification can help in analyzing the gene expression patterns when the relationship between gene expression values is not monotonic. Aid-Pavlidis and coworkers (2009) traced those genes whose co-expression with BDNF gene is preserved under different conditions, e.g., age, gender and disease state. Dysregulation of this gene is observed in AD, Huntington’s disease and Parkinson’s patients. In another work, Wang and coworkers (2015) introduced a method for estimating the strength of gene interactions in groups. Their method infers the strength of gene group interactions using sparse canonical correlation analysis coupled with repeated random partitioning and subsampling of the gene expression data. More specifically, to select the strength of gene group interaction, the co-expression relationship between genes is traced by considering different subsets of genes, and then, determined the conditional dependencies between genes after removing the influences of a set of other functionally related genes.

In the present study, a three-way interaction model is used for determining the disease-related mechanisms in AD. Such a model can trace the dynamic nature of co-expression relationship of two genes (\(X_1, X_2\)) by the introducing a third gene (\(X_3\)), whose expression level modulates the correlation between \(X_1\) and \(X_2\). Fig 1 illustrates an example of a three-way interaction. One of the statistical methods for this purpose is liquid association analysis. This method describes co-expression changes of two genes based on the expression level of a third gene, which is sometimes referred to as the controller gene [13].
**Materials and methods**

**Gene expression**

The dataset includes gene expression of the cerebral cortex of APP/PS1 and WT mouse littermates (accession number E-MTAB-2121 [17] in ArrayExpress database). The data were generated using the Affymetrix microarray platform and the GeneChip Mouse Gene 1.1 ST Array. The raw data were preprocessed and normalized using robust multi-array analysis method (implemented in the Bioconductor Limma package [18]).

One of the main challenges in microarray data analysis is the large dimensionality. To reduce the data size, duplicate probes were removed using genefilter function in the Bioconductor genefilter package [19]. It should be noted that this function retain the highest interquartile range (IQR) probe for each gene.

**Determining three-way interactions**

**Selection of candidate switching genes.** In this study, a candidate switching gene is defined to have a significantly different gene expression level in diseased and normal samples. Therefore, empirical Bayes t-test was performed using the lmfit function in the Limma package [18] for each gene, given the groups of arrays. Afterward, false discovery rate estimate (FDR) was calculated using the Benjamini-Hochberg (BH) correction method [20]. Differentially-expressed genes with FDR <0.05 were selected as candidate switching genes.

**Liquid association triplets.** Three-way interactions between candidate switching genes and all possible pairwise combinations of genes in the dataset were calculated using fastMLA function in fast LA package [21]. This package uses a modified liquid association algorithm for determining changes in co-expression relations of a gene pair, $X_1$ and $X_2$, based on the expression level of a third gene ($X_3$).

In the fast liquid association algorithm, for each gene triplet, an MLA score is computed. More specifically, MLA ($X_1, X_2 | X_3$) can be estimated as:

$$\hat{MLA} = \frac{\sum_{i=1}^{M} \hat{\rho}_{X_{1i}X_{2i}}}{M}$$
where $M$ is the number of bins over $X_3$, $\hat{p}_i$ is the Pearson’s correlation coefficient of $X_1$ and $X_2$ in samples of the $i$'th bin, and $X_{3i}$ is the mean of expression values of $X_3$ in the $i$'th bin.

Performing two preprocessing steps on the data is necessary before running fastLA.

1. To obtain asymmetric edges for each variable, normal quintile transformation should be performed for each gene, based on the approach proposed by Li [13].

2. Expression level of each gene should be standardized to zero mean and standard deviation of one [22].

The first preprocessing was performed using an in-house implementation, while the second one by using CTT package [23].

False discovery rate (FDR) was estimated by using the Benjamini-Hochberg correction method and liquid association triplets with FDR $< 0.001$ was chosen as statistically significant triplets.

**Gene set enrichment analysis**

Gene set enrichment analysis (GSEA) is a statistical method to assess the significance of the shared association of several genes (proteins) using predefined annotations [24]. For every switching gene, and also for all of the genes involved in all statistically significant triplets, GSEA was separately performed based on biological process and cellular component using the gene ontology (GO) database. Furthermore, the same analyses were performed to find enriched pathways in KEGG database [25]. For the above-mentioned analyses, we used ClueGO tool [26] (with a Kappa threshold of 0.4) within the Cytoscape v.3.3.0 environment [27]. Right-sided hypergeometric test and the Benjamini-Hochberg correction method were used for validation of enrichment analysis.

**Construction of gene regulatory network**

A gene regulatory network (GRN) is a framework to model complex regulatory mechanisms that control gene expression in cells. A GRN is represented as a directed graph, which consists of nodes (genes) and directed edges (regulatory relations) that can be activatory or inhibitory. By using such a network, changes in gene expression can be predicted after a particular stimulation [28]. Here, we used ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) [29] for constructing the GRN. ARACNE is an approach for reverse engineering of cellular networks from gene expression data. This algorithm infers directed regulatory interactions between each transcriptional regulator and its potential targets based on mutual information. Construction of GRN was performed using geWorkbench_2.6.0 software application ARACNE between switching candidate genes as hub markers and all of the genes involved in the statistically significant triplets by considering $p$-value $< 0.0001$.

The code for above analyses is available from: https://github.com/NKhayer/3wayintraction

**Results**

**Preprocessing and normalization of gene expression data**

After preprocessing and normalization using robust multi-array analysis method, the final dataset consisted of 30 samples including 15 wild type and 15 APP/PS1 transgenic samples (It should be noted that two arrays (i.e., 15.w.3 and 26.w.12) were discarded from dataset due to low quality). In the original dataset, each sample included 35556 probes. As a result of removing duplicated probes, this number dropped 20527 probes.
Additional data including array images, MA plots, box plots and data histograms before and after of normalization are provided in S1 Fig.

Selection of candidate switching genes

In order to determine the candidate switching genes, eBayes t-test was performed on the normalized data. We selected 38 differentially-expressed genes by considering FDR <0.05 as the candidate switching genes. The detailed results of this analysis are available in S1 Table.

Determining significant three-way interactions

Using fastLA package, liquid association analysis was performed for every combination of a candidate switching gene ($X_3$) and every possible pair of genes in the dataset {$X_1$, $X_2$}. The top 300000 triplets with the highest significance levels based on $p$-value were defined as outputs of this analysis. In S2 Fig, $p$-value histogram of these three-way interactions is available. Fig 2 shows the changes in FDR using Benjamini-Hochberg method versus $-\log (p$-value) for the first 300000 triplets. For the rest of our analysis, the set of all three-way interactions were chosen by considering FDR <0.05, which consists of 106 triple combinations. Within these triplets, we found 241 unique genes including, 35 unique $X_3$ genes and 207 unique $X_1$ or $X_2$ genes. These results show that repetition of $X_1$ and $X_2$ in triplets is rare. The list of all statistically significant triplets are presented in S2 Table. To show the switching genes which are involved in the numerous significant triplets, a network from statistically significant triplets was constructed (see S3 Fig). This network shows that $Slc14a1$, $Dock8$, $Slc11a1$ and $Parvg$ (with 10, 7, 7 and 6 connections, respectively) have the largest contribution in the statistically significant triplets rather than other $X_3$. Also the network expansion shows that the switching mechanisms are not limited to specific number of $X_3$, as expected for complex diseases such as AD.

Results of gene set enrichment analysis.

We used GSEA in order to find biologically-relevant triplets in the 106 statistically significant triplets. GSEA was performed with $p$-value <0.05 and FDR <0.1 for $X_3$ genes (35 individual genes) and all of the involved genes in triplets (241 individual genes). Enriched terms based on "biological process", "cellular component" and "KEGG pathway" are reported in Fig 3. Since the terms in lower levels of gene ontology are more general, enriched terms for "biological process" and "cellular component" in levels lower than level 10 and level 9 are not reported. The results of GSEA for two gene groups, i.e., genes in $X_3$ position and all of the involved genes in triplets, show that there are a large number of common terms in two groups. For verifying that the large number of common terms in group 1 (i.e., $X_3$) and group 2 (i.e., $X_1$, $X_2$, $X_3$) are significantly different from random, we extracted 30 random subsets from group 2, each subsets including 35 genes. After gene set enrichment analysis, very rare overlap was observed between group 2 and each of the random subsets. These results confirm that the large number of common terms in group 1 and 2 are not by chance. The results of this analysis are available in S3 Table.

The terms, "Regulation of ERK1 and ERK2 cascade", "Pattern recognition receptor signaling pathway", "T-cell proliferation", and "Neutrophil chemotaxis", based on biological process (Fig 3A) and the term "Lysosome" based on cellular component (Fig 3B) and a number of immunity-related terms, including "Cytokine-cytokine receptor interaction", "Rheumatoid Arthritis", and "Leishmaniasis" based on biological pathway (Fig 3C) are common in two groups. The complete list of enriched terms is available in S4 Table. Furthermore, the categorized biological processes based on the AD-related pathways are available in S5 Table.

Based on the proposed definition of three-way interactions of switching mechanism model, it is expected that in biologically-relevant triplets, $X_1$ and $X_2$ are in the same pathway or biological process. By tracing triplets in the enriched terms, 12 triplets in which $X_1$ and $X_2$ are
involved in the same biological process or pathway were determined (Table 1). As it is reported in Table 1, regulatory and innate immunity processes are frequently repeated in these triplets. Fig 4 indicates two exemplary scatter plots of these triplets in three different ranges of associated $X_3$ expression level which shows a considerable change in the correlation of $X_1$ and $X_2$ as a result of change in $X_3$. Scatter plots of all of the 12 triplets are available in S4 Fig.

**Determining the position of triplets in gene regulatory network.** As another attempt for analyzing the functional relevance of three-way interactions, we reconstructed a GRN based on ARACNE. By considering $p$-value < 0.0001 as the threshold, the resulting network

![Graph](https://doi.org/10.1371/journal.pone.0184697.g002)

Fig 2. FDR vs. -log ($p$-value). The changes in FDR (BH-corrected $p$-value) versus -log ($p$-value) for the first 300000 results of fastLA [21]. As shown FDR = 0.001 corresponds to -log ($p$-value) = 6.817.

https://doi.org/10.1371/journal.pone.0184697.g002

![Gene set enrichment analysis](https://doi.org/10.1371/journal.pone.0184697.g003)

Fig 3. Gene set enrichment analysis. Enriched terms based on (A) “biological process”; (B) “cellular component”; and (C) “KEGG pathway” for two gene groups, genes in $X_3$ position and all of the genes involved in the triplets. The common terms in these two groups are shown in red. The high frequency of common terms suggest that the results of liquid association method are consistent with the biological expectation from three-way interactions, that is, the presence of switching and switched genes in the same biological pathway.

https://doi.org/10.1371/journal.pone.0184697.g003
included 42 nodes and 82 edges. This network is available S1 File, Section A. The regulatory relationship of significant triplets obtained from liquid association method was traced in this network and the results are shown as a subnetwork in Fig 5. As it is observed, the regulatory relationship between $X_3$ from 72nd triplet ($H2-Ob$) and two other genes in this triplet ($Milr1$ and $Csf1r$) can be seen with $Slc14a1$ as an intermediate. In addition, a direct regulatory interaction between $X_3$ ($Slc14a1$) and $X_2$ ($Slamf6$) from 97th triplet is observed.

The details of construction GRN and detection of significant triplets in this network is available in S1 File, Section B.

### Discussion

Although pairwise gene interactions have been widely investigated in AD [30–32], in the present study, for the first time, we investigated the switching mechanism in AD using a three-way

<table>
<thead>
<tr>
<th>Biological Process</th>
<th>GO Term</th>
<th>GO Levels</th>
<th>p-value</th>
<th>FDR (BH)</th>
<th>Triplet Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biologica l Process</td>
<td>Regulation of immune system process</td>
<td>[2, 3]</td>
<td>1.13E-05</td>
<td>5.20E-03</td>
<td>34,72,90</td>
</tr>
<tr>
<td></td>
<td>Defense response</td>
<td>[3]</td>
<td>2.15E-04</td>
<td>6.20E-03</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Response to other organism</td>
<td>[2, 4]</td>
<td>3.32E-04</td>
<td>9.01E-03</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Regulation of defense response</td>
<td>[4, 5]</td>
<td>9.20E-04</td>
<td>1.21E-02</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Cytokine production</td>
<td>[3]</td>
<td>7.46E-04</td>
<td>1.23E-02</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Negative regulation of apoptotic signaling pathway</td>
<td>[4, 5, 6, 7, 8]</td>
<td>1.24E-03</td>
<td>1.50E-02</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Regulation of response to stimulus</td>
<td>[2, 3]</td>
<td>1.28E-03</td>
<td>1.52E-02</td>
<td>22,29,34</td>
</tr>
<tr>
<td></td>
<td>Positive regulation of cellular metabolic process</td>
<td>[3, 4, 5]</td>
<td>1.46E-03</td>
<td>1.64E-02</td>
<td>34,45</td>
</tr>
<tr>
<td></td>
<td>Cellular response to chemical stimulus</td>
<td>[3]</td>
<td>1.91E-03</td>
<td>1.80E-02</td>
<td>34,38,22</td>
</tr>
<tr>
<td></td>
<td>Regulation of response to external stimulus</td>
<td>[3, 4]</td>
<td>1.76E-03</td>
<td>1.81E-02</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Cellular response to hydrogen peroxide</td>
<td>[5, 6]</td>
<td>2.89E-03</td>
<td>2.30E-02</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Programmed cell death</td>
<td>[4]</td>
<td>3.27E-03</td>
<td>2.48E-02</td>
<td>34,22</td>
</tr>
<tr>
<td></td>
<td>Negative regulation of apoptotic process</td>
<td>[5, 6, 7]</td>
<td>5.06E-03</td>
<td>3.16E-02</td>
<td>22,34</td>
</tr>
<tr>
<td></td>
<td>Regulation of apoptotic signaling pathway</td>
<td>[4, 5, 6, 7]</td>
<td>5.58E-03</td>
<td>3.30E-02</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Positive regulation of biological process</td>
<td>[1, 2, 3]</td>
<td>7.43E-03</td>
<td>3.90E-02</td>
<td>34,45,60,90</td>
</tr>
<tr>
<td></td>
<td>Positive regulation of metabolic process</td>
<td>[2, 3, 4]</td>
<td>8.37E-03</td>
<td>4.16E-02</td>
<td>34,45</td>
</tr>
<tr>
<td></td>
<td>Positive regulation of macromolecule metabolic process</td>
<td>[3, 4, 5]</td>
<td>8.60E-03</td>
<td>4.23E-02</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Cellular response to oxygen-containing compound</td>
<td>[4]</td>
<td>9.06E-03</td>
<td>4.27E-02</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Cell surface receptor signaling pathway</td>
<td>[4, 5]</td>
<td>1.00E-02</td>
<td>4.64E-02</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Positive regulation of cellular process</td>
<td>[2, 3, 4]</td>
<td>1.15E-02</td>
<td>5.18E-02</td>
<td>34,45,90</td>
</tr>
<tr>
<td></td>
<td>Regulation of apoptotic process</td>
<td>[5, 6]</td>
<td>1.35E-02</td>
<td>5.68E-02</td>
<td>22,34</td>
</tr>
<tr>
<td></td>
<td>Oxidation-reduction process</td>
<td>[3]</td>
<td>1.53E-02</td>
<td>5.94E-02</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Negative regulation of cell proliferation</td>
<td>[3, 4, 5]</td>
<td>1.50E-02</td>
<td>5.98E-02</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Response to wounding</td>
<td>[3]</td>
<td>1.59E-02</td>
<td>6.11E-02</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Apoptotic signaling pathway</td>
<td>[4, 5, 6]</td>
<td>1.61E-02</td>
<td>6.11E-02</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Regulation of protein metabolic process</td>
<td>[4, 5]</td>
<td>1.82E-02</td>
<td>6.66E-02</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Regulation of response to stress</td>
<td>[3, 4]</td>
<td>2.04E-02</td>
<td>7.09E-02</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Regulation of signal transduction</td>
<td>[3, 4, 5]</td>
<td>2.52E-02</td>
<td>8.01E-02</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Response to oxygen-containing compound</td>
<td>[3]</td>
<td>2.67E-02</td>
<td>8.34E-02</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Single-organism transport</td>
<td>[3, 4]</td>
<td>3.02E-02</td>
<td>8.83E-02</td>
<td>72</td>
</tr>
</tbody>
</table>

### Table 1. Biologically relevant triplets

By tracing statistically significant triplets in the enriched terms, 12 triplets in which $X_1$ and $X_2$ are involved in the same biological process or pathway were determined.

https://doi.org/10.1371/journal.pone.0184697.t001

https://doi.org/10.1371/journal.pone.0184697.1001
interactions model. In two-way interaction model, the co-expression level of two genes, $X_1$ and $X_2$, is measured based on their correlation coefficient. As a result, highly correlated genes are identified as functionally related and their encoded proteins are assumed to be involved in the same biological pathway, to participate in the same structural complex or to be regulated by the same mechanism. Nevertheless, some of the functionally related genes may not be co-expressed. A reason for this fact is that gene expression is often sensitive to changes in cellular states such as the presence or absence of the hormones and metabolites or ionic homeostasis [33]. Since biological pathways are interwoven, and each protein may have several cellular roles, co-expression of all of the functionally-related genes might change depending on cellular states, which is often not well-known [13]. Therefore, degree and pattern of correlation of two

Fig 4. Examples of the statistically triplets. In each case, a considerable change in the correlation of $X_1$ and $X_2$ occurs as a result of change in $X_3$.

https://doi.org/10.1371/journal.pone.0184697.g004

Fig 5. Regulatory relationships within triplets. The regulatory relationships of significant triplets obtained from liquid association were traced in the GRN. The more the intensity of the red color, the more the up-regulation of the gene in Alzheimer’s disease. As shown there are regulatory relationships between Mi lr1, Csf1r and H2-Ob in triplet 72. Additionally, regulatory relationships are observed between Slc14a1 and Slamf6 in the triplet 97.

https://doi.org/10.1371/journal.pone.0184697.g005
gene expression profiles can be affected by changes in internal cellular states. In other words, if the condition changes, two functionally-related genes with direct correlation may become uncorrelated or even inversely correlated. Hence, two-way interaction model might be too simplistic to explain complex molecular relations. In contrast, three-way interaction model, which can capture more complicated behavior compared to the two-way interaction model, might be able to demystify more complex molecular relations [16].

The main difference of Aid-Pavlidis et al.’s study with the three-way interaction studies is that the former aims to trace the co-expression relations within a gene expression dataset, while three-way interaction studies trace differential co-expression relations. The common assertion in Wang et al.’s study and the three-way interaction studies is that the co-expression relationship between genes may be influenced by some other genes.

Three common methods exist for finding three-way interactions with switching mechanism: liquid association [13], differential correlation coefficient [34], and interaction test [35]. Here, we used fast Liquid Association (fastLA) algorithm [21]. The reason behind this selection is that liquid association is applicable to X\textsuperscript{3} with quantitative data such as gene expression data, while interaction test and differential correlation coefficient methods are only used for X\textsuperscript{3} with qualitative data such as genotype data or different classes of gene expression data. In addition, in liquid association method, the high computational load, which is the main challenge in investigating the genome-wide three-way interactions, is decreased. Therefore, this algorithm can be practically applied to high-throughput data like mouse microarray data.

The E-MTAB-2121 dataset includes gene expression data of APP/PS1 and WT mouse littermates. In S1 Text, Section A, we discussed about reasons for choosing the mouse rather than human dataset.

The low repetition at X\textsubscript{1} and X\textsubscript{2} genes in the obtained triplets from fastLA approves the critical role of X\textsubscript{3} in the change of \{X\textsubscript{1}, X\textsubscript{2}\} expression correlation, because in the absence of X\textsubscript{3} it is expected that in every run of fastLA the same \{X\textsubscript{1}, X\textsubscript{2}\} are reported as the significant switched gene pairs.

From the set of 35 "X\textsubscript{3}" genes, only 6 genes are reported in the Lopez-Gonzalez et al.’s study as the differentially expressed genes. In S1 Text, Section B, we discussed about this small similarity.

A high frequency of common terms is observed between the GSEA results of the two gene groups, i.e., the set of X\textsubscript{3} genes and the set of all genes in the triplets (Fig 3A). This consistency suggests that the switching genes and the switched genes are associated to biologically related pathways. In addition, previous studies showed the importance of these common enriched terms in AD (see below) which confirm the validity of the results obtained in this study.

### Biological processes enrichment analysis

The common enriched "biological process" terms between both of the gene groups (i.e., genes in X\textsubscript{3} position and all of the genes involved in triplets) are "Regulation of ERK1 and ERK2 cascade", "Pattern recognition receptor signaling pathway", "T-cell proliferation", and "Neutrophil chemotaxis" (Fig 3A). All of the four common enriched terms are also biologically relevant to AD. See below.

Up-regulation of ERK1/2 is associated to hyper-phosphorylation of tau protein, and consequently, production of neurofibrillary tangles [36,37]. This kinase is suggested as drug target for AD [38]. Pattern recognition receptors are the intermediate proteins to trigger the innate immune system in AD [39–41]. T cell proliferation is induced by interleukin 15, which is produced following the activation of macrophages in AD [42]. Furthermore, monocyte-derived
dendritic cells in AD patients are able to interfere in induction of T cell proliferation [43]. Neutrophil chemotaxis is induced by β-amyloid peptides in AD [44, 45].

**KEGG pathways enrichment analysis**

The three KEGG pathways that are commonly enriched in both groups are "Cytokine-cytokine receptor interaction", "Rheumatoid Arthritis", and "Leishmaniasis" (Fig 3C). The importance of cytokine/cytokine receptor interaction in AD is reported previously [44,45]. Both rheumatoid arthritis and AD are known to be associated with inflammation [46–48]. Moreover, studies such as Refs. [49–51] have reported the relationship between the two diseases. The common aspect of AD and Leishmaniasis is the activation of innate immune system. Activation of innate immune system in Leishmaniasis and AD is mediated through glycoalyx on the surface of Leishmania parasite [52] and Aβ Aggregates [53], respectively. Interestingly, glycogen synthase kinase-3 inhibitors, which are introduced as anti-parasite drugs for treating leishmaniasis in the first place, are also suggested as candidates for AD treatment [54,55].

**Cellular components enrichment analysis**

"Lysosome" is the common enriched cellular component term in two group of genes (Fig 3B). Mutation of proteins involved in lysosomal/endosomal transports are identified in many neurological diseases and dysfunction of lysosomal system is a definite characteristic of AD [56–58]. Based on these facts, one can explain why lysosome has a significant contribution in the onset and progression of AD.

More particularly, the 12 triplets in which X₁ and X₂ genes are involved in the same pathway or biological process (Table 1) show that regulatory processes and innate immunity are important in the switching mechanisms involved in AD. According to the concept of switching mechanism, the presence of triplets in regulatory pathways is expectable. The 72nd triplet, which consists of genes H2-Ob and {Csf1r, Milr1}, is one of the most statistically significant triplets in our analysis. Furthermore, this triplet is biologically relevant, based on both gene set enrichment analysis (Table 1) and gene regulatory network analysis (Fig 5). As it is shown in Table 1, genes in this triplet are involved in regulatory of immune system processes. Innate immune system has a prominent role in the onset and progression of AD. Neuropathology of AD indicates a strong innate immunity response which is characterized by activated microglia [59], de novo expression or increase of various macrophage antigens [60], and inflammatory cytokine production [61]. Proliferation and activation of microglial cells is a noticeable feature of several neurodegenerative disorders such as AD[62]. This mechanism is regulated by activation of Csf1r (colony-stimulating factor1 receptor) [62]. Up-regulation of Csf1r is reported both in amyloid model of mouse [63] and post-mortem human samples [62]. Olmos-Alonso et al., suggest Csf1r gene as a drug target for prevention of AD progression and show that Csf1r inhibition in transgenic mice APP/PS1 using GW2580, a tyrosine kinase inhibitor, results in blockage of microglia proliferation, change of microglia inflammatory profile to non-inflammatory phenotype and also improvement in memory and behavioral tasks, and prevention of synapse destruction [64]. Spangenberg et al., showed that Csf1r inhibition in 5xfAD AD mouse model leads to clearance of active chronic microglia. This clearance not only is not harmful in the animal model, but also results in improvement of synapse disorders compared to the disease state and additionally lead to decrease in neuronal inflammation and prevents neuron destructions [65].

The protein encoded by Milr1 gene (mast cell immunoglobulin-like receptor 1, equivalent to allergy inhibitory receptor 1), is an immunoglobulin-like receptor which is often expressed at the surface of mast cells. This protein has an immunoreceptor tyrosine-based inhibitory
motif in cytoplasmic region and acts as a negative regulator of FcεRI-dependent signaling pathway in mast cells [66,67]. Mast cells are abundant cells in central nervous system and contain pro-inflammatory mediators such as histamine, cytokines, ATP, proteases and leukotrienes, which are stored in their internal granules. Anaphylaxis is an immediate hypersensitivity reaction that is started with antigen-carrying immunoglobulin E binding to FcεRI receptor and results in the degranulation of mast cells and secretion of pro-inflammatory mediators to extracellular space [68,69]. Some report suggest that amyloid peptides induce mast cell degranulation response [70,71]. The resulting release of different pro-inflammatory molecules from mast cells may be associated to the onset or even progress of AD [71].

H2-Ob is the switching gene in the 72nd triplet. HLA-DOB is the orthologous of this gene in human, which is related to MHC class II. This protein is commonly expressed in professional antigen presenting cells such as macrophages, B cells and specific dendritic cells. The importance of HLA-DOB gene has been reported in a wide range of disease, including schizophrenia [72], type I diabetes, [73], rheumatoid arthritis [74], Kawasaki disease [75], multiple myeloma [76] and chronic lymphocytic leukemia [77]. Furthermore, two genes associated with immune system, namely HLA-DRB5 and HLA-DRB1, which are both related to MHC class II are suggested to be linked to AD [78]. Therefore, HLA-DOB might be involved in AD.

As shown in Fig 1, when normalized expression level of H2-Ob gene is between 0.37 and 1.84 (as in the most AD samples (see S1 Table)), there is a direct correlation between Milr1 and Csf1r expression levels. Note that Milr1 and Csf1r are associated with biological processes “negative regulation of FcεRI-dependent signaling pathway” and “proliferation and activation of microglial cells”, respectively. Therefore, the positive correlation between Milr1 and Csf1r expression levels implies that the two above-mentioned biological processes are activated in concert with each other. In contrast, when normalized expression level of H2-Ob gene is between -1.84 and -0.37 (as in the most wild type samples (see S1 Table)), the two mentioned processes probably act in opposite directions, because of the inverse correlation between Milr1 and Csf1r expression levels. In other words, changes in expression level of H2-Ob gene can act as the switching factor for altering the cellular behavior. The homeostasis-related link between mast cells and microglia is presumably controlled with H2-Ob as switching genes in AD. See below.

Mast cells have an important role in starting the inflammation of central nervous system and are the first responder cells to the brain injury [79]. Mast cells act as the inducer for microglial activation. It is previously reported that mast cell activation results in induction of microglia for releasing neurotrophin [80]. Moreover, it is reported that induction of microglia activation is done via histamine and tryptase, which are inflammation mediators released by mast cells. Moreover, they showed that activated mast cells directly cause the activation of microglia [81–83].

Mast cells and microglial cells have endogenous homeostatic molecules that can be over-produced as a result of tissue destruction or induction of inflammatory responses. Palmitoyl-lethanolamide, which is produced and hydrolyzed by microglia [84], has a key role in maintaining cellular homeostasis through inhibition of mast cell function, especially when there is an exogenously induced stress such as inflammation [85,86].

Taking into account the mast cell-microglia communications and the importance of homeostasis for survival, one can conclude that the same direction of two biological processes “negative regulation of FcεRI-dependent signaling pathway” and “proliferation and activation of microglial cells”, which is observed in the 72nd triplet in conditions related to AD, is indicative of a type of homeostatic mechanism in the cell that occurs due to changes in expression level of H2-Ob gene. Following the increase in activation of mast cells as the first responder cells to brain damages in AD, up-regulation of H2-Ob gene, either directly or through mediators, may cause change in the correlation of Csf1r and Milr1 expression levels. As a result, the
biological process of “proliferation and activation of microglial cells”, which happens in response to disease-related factors such as β-amyloid plaques, and the biological process of “negative regulation of FcεRI-dependent signaling pathway”, which is related to maintaining cellular homeostasis, may act in the same direction. In contrast, in normal condition, as mast cells are inactive and these cells are in steady condition. Therefore, an inverse correlation is expected between Csf1r and Milr1 genes when the expression level of H2-Ob is low. Consequently it is expected that the biological processes related to these two genes act in the opposite directions probably.

The possible link between Csf1r and neuronal loss

Csf1r has been previously suggested to be involved in microglia activation, and therefore, it was suggested to be a drug target for AD [87]. Interestingly, there are two recent studies in which inhibition of Csf1r, which in turn results in eliminating microglia, is investigated. In the first study, Spangenberg and coworkers (2016) found that Csf1r inhibition prevents neuronal loss in AD mice model [65]. On the other hand, Hilla and coworker (2017) have observed that Csf1r inhibition does not have a preventing effect on neuronal loss [88]. These seemingly contradictory observations should certainly be studied more carefully. A possible explanation for these observations might be that Csf1r inhibition may not be the only determinant of the neuronal loss. In other words, if we assume that there are other players involved in this process one would expect that the simultaneous activity of both proteins, and not Csf1r alone, may determine neuronal degeneration. Based on our results, we already know that the cell involves a switch to fine-tune the simultaneous activity of Csf1r and Milr1. This switch, or other similar switches might be involved in the process of neuronal degeneration, and explain why neuronal loss cannot simply be inferred from Csf1r activity alone.

Conclusion and future work

A huge number of disease-related high-throughput “omics” datasets are freely available today. Such datasets contain valuable information about disease-related pathways and their corresponding gene interactions. In the present study, for the first time, we used the three-way interaction model for tracing the switching genes involved AD. The advantage of this approach compared to the pairwise co-expression analysis is that the three-way interaction model can cope with the dynamic nature of co-expression relations. Hence, the three-way interaction model can potentially shed light on some of the (other) causes of cellular alterations. More specifically, in the present study, we suggest that H2-Ob as the switching gene and the gene pair \{Csf1r, Milr1\} form a statistically significant and biologically relevant triplet in AD. The relationship among the pathways associated with these three genes reveals that two biological processes, namely “negative regulation of FceRI-dependent signaling pathway” and “proliferation and activation of microglial cells” are in the same direction in conditions related to AD. This is indicative of a type of homeostatic mechanism in the cell that occurs when H2-Ob is up-regulated.

In the next step, we plan to experimentally validate the relationship between H2-Ob gene and the \{Csf1r, Milr1\} gene pair.

Supporting information

S1 Table. Differentially-expressed genes.
(PDF)

S2 Table. Statistically significant triplets.
(PDF)
S3 Table. Comparison of GSEA results in group1, 2 and random groups.
(PDF)

S4 Table. Gene set enrichment analysis results based on biological process, cellular component, and KEGG pathway.
(PDF)

S5 Table. The categorized biological processes based on the AD-related pathways.
(PDF)

S1 Fig. This file includes array images, MA plots, box plots and data histograms before and after of normalization.
(PDF)

S2 Fig. The p-value histogram of the top 300000 three-way interactions.
(PDF)

S3 Fig. Statistically significant triplets Network.
(PDF)

S4 Fig. Scatter plots of 12 triplets in which X1 and X2 are involved in the same biological process.
(PDF)

S1 Text. This file includes “the reasons for choosing the mouse rather than human dataset” (Section A) and “comparison of present study results with the Lopez-Gonzalez et al.’s study” (Section B).
(PDF)

S1 File. This file includes “figure of Gene regulatory network” (Section A) and “the steps of construction gene regulatory network” (Section B).
(PDF)

Author Contributions

Conceptualization: Nasibeh Khayer, Sayed-Amir Marashi, Mehdi Mirzaie, Fatemeh Goshadrou.

Data curation: Nasibeh Khayer.

Formal analysis: Nasibeh Khayer.

Funding acquisition: Nasibeh Khayer.

Investigation: Sayed-Amir Marashi.

Methodology: Sayed-Amir Marashi, Mehdi Mirzaie.

Project administration: Sayed-Amir Marashi.

Software: Nasibeh Khayer, Mehdi Mirzaie.


Validation: Mehdi Mirzaie.

Writing – review & editing: Nasibeh Khayer, Sayed-Amir Marashi.
References


