Differential gene expression of blood-based ABCA9, CNOT8, SESN1, UCP3, MAP2K1 and DDIT4 in Alzheimer’s disease
Ainon Zahariah Samsudin 1, Kalavathy Ramasamy 2, Siong Meng Lim 2, Ai Vyrn Chin 3, Maw Pin Tan 3, Shahrul Bahyah Kamaruzzaman 3, Baharudin Ibrahim 4 and Abu Bakar Abdul Majeed 1*

1 Brain Degeneration & Therapeutic Group, Faculty of Pharmacy, Universiti Teknologi MARA (UiTM), 42300 Bandar Puncak Alam, Selangor Darul Ehsan, Malaysia.
2 Collaborative Drug Discovery Research (CDDR) Group, Faculty of Pharmacy, Universiti Teknologi MARA (UiTM), 42300 Bandar Puncak Alam, Selangor Darul Ehsan, Malaysia.
3 Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.
4 Faculty of Pharmacy, University of Malaya, 50603 Kuala Lumpur, Malaysia.
* Correspondence: abubakar@uitm.edu.my; Tel.: +603-32584001

Received: 2 June 2023; Accepted: 3 November 2023; Published: 31 December 2023
Edited by: King Hwa Ling (Universiti Putra Malaysia, Malaysia)
Reviewed by: Shahidee Zainal Abidin (Universiti Malaysia Terengganu, Malaysia); Mohamad Halim Mohamad Shariff (Universiti Teknologi MARA, Malaysia)
https://doi.org/10.31117/neuroscirn.v6i4.262

Abstract: This study uncovered differential gene expression in blood to distinguish subjects with probable Alzheimer’s disease (AD) from normal elderly participants (non-demented controls, NDC). The participants were recruited via training (Phase 1) and validation cohorts (Phase 2). The changes of gene expression in blood samples from the training cohort (92 AD vs 92 NDC) were assessed using the microarray technology. The Partial Least Square Discrimination Analysis (PLSDA) was then used to develop a disease classifier algorithm (accuracy = 88.3%). Six differentially expressed genes were validated through RT-qPCR using blood samples from the validation cohort [(25 AD, 25 NDC, 12 mild cognitive impairment (MCI) and 12 vascular dementia (VaD) subjects]. The PLSDA model indicated a good separation between AD and NDC [area under the receiver operating characteristic curve (ROC AUC) = 0.88]. ABCA9, CNOT8, SESN1, UCP3, MAP2K1 and DDIT4 were found to be differentially expressed between the two groups. Validation of the panel of six genes gave an overall accuracy of 82.0% (AUC=0.86). The ABCA9 mRNA level, which was significantly (p < 0.05) lower in the AD group, correctly classified 90.9% of all subjects (AUC=0.94). This group of genes may be responsible for dysregulation of pathways related to inflammation, mitochondrial dysfunction, oxidative injury, DNA damage, apoptosis and lipid metabolism. The disease classifier algorithm discriminated probable AD from MCI and VaD at specificity of 83.3% and 75.0%, respectively. These findings warrant further validation of potential blood-based biomarkers in larger samples of clinical AD.

Keywords: Alzheimer’s disease; ATP-binding cassette; biomarkers; blood; transcriptomics

©2023 by Samsudin et al. for use and distribution in accord with the Creative Commons Attribution (CC BY-NC 4.0) license (https://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.
1.0 INTRODUCTION
Globally, Alzheimer’s disease (AD) is expected to become a major health concern as the proportion of the aged population enlarges (Alzheimer’s Disease International, 2021; GBD 2019 Dementia Forecasting Collaborators, 2022). Amongst the reasons for this is the high cost of currently available methods of diagnosing and the long waiting time for patients to endure. The challenge, therefore, is for health service providers to be able to diagnose AD through the most affordable approach and identify the probability of engaging with the disease as early as possible. The key problem with early AD diagnostics is that the brain tissue is inaccessible for biopsy, whilst cerebrospinal fluid (CSF) collection is not a valid option for population-wide screening; which leaves either blood or saliva as the most likely candidate. Blood biomarkers are promising diagnostic indicators as they offer several advantages in simplicity, convenience, scalability and cost-effectiveness. Although at one time elusive, recent reports have shown that blood-based biomarkers can become a reality in the coming years. Teunissen et al. (2022) and Schindler and Bateman (2021) found that concentrations of amyloid, phosphorylated tau proteins, neurofilament light chain and glial fibrillary acidic protein in the blood were correlated to AD with remarkable consistency across different cohorts. Nevertheless, ultrasensitive detection methods are often required. It was also found that a combination of blood biomarkers may be highly useful in predicting individuals with mild cognitive impairment (MCI) to progress to AD (Cullen et al., 2021).

Previous studies indicated differential gene expression in blood sample of subjects with AD and normal elderly controls (Boooij et al., 2011; Lee & Lee, 2020; Lunnon et al., 2013; Patel et al., 2020). The diagnostic parameters were, however, inconsistent and further studies are required for clinical utility (Donaghy et al., 2022). As part of the effort to uncover useful blood-based biomarkers, the present study was undertaken amongst two independent cohorts. The initial training cohort (Phase 1) involved the selection of potential biomarkers after excluding possible confounding effects. The significant separation of AD from the normal elderly participants (non-demented controls, NDC) group in the training cohort was determined based on the strong analytical power of ≥ two-fold statistical difference. This was then followed by validation of the potential biomarkers in a second cohort (Phase 2). To determine whether the selected genes were able to correctly predict AD versus (vs) non-AD groups by the disease classifier when tested, this study included additional groups of subjects with mild cognitive impairment (MCI) and vascular dementia (VaD). Besides, the present study also investigated the differential changes of gene expression in blood of probable AD relative to healthy subjects, thereby identifying the major pathophysiological pathways involved in AD. Furthermore, the performance of the selected differentially expressed genes was tested to determine their ability to distinguish AD, MCI and VaD.

2.0 MATERIALS AND METHODS
2.1 Recruitment and assessment of participants
This study was approved by the ethics committees of UiTM [reference no: 600-RMI (5/1/6/01)] and the University of Malaya Medical Centre (UMMC) (reference no: PPUM-HU-61/12/1-1). The present experimental design was in accordance with principles and guidelines stipulated by the Declaration of Helsinki, World Medical Association (Carlson et al., 2004). A written informed consent was obtained from each patient or legal representative before blood collection. Recruitment and assessment procedures were conducted as described in detail by Mohd Hasni et al. (2017) and Rehiman et al. (2022). The training cohort (Phase 1) comprised 184 participants [92 NDC subjects vs 92 probable AD patients] whilst the validation cohort (Phase 2) was made up of a total of 74 participants [25 NDCs vs 25 probable AD vs 12 MCI vs 12 VaD] (Method S1). Figure S1 illustrates the workflow of the present study.

2.2 Microarray
Total RNA for microarray was extracted from blood samples using the Ribopure™ – Blood RNA Isolation Kit (Ambion, USA) and stored at −80 °C before use. Only high-integrity RNA with a cut-off > 7.0, 260/280 and 260/230 ratios > 1.8 were used for subsequent analysis. A total of 184 samples were being analysed for gene expression based on a one-colour microarray experiment using the commercial oligonucleotide microarray slide. The Oligonucleotide probe (single-stranded RNA fragment) in the slide was 60 bases (mer) in length. The Agilent SurePrint G3 Human GE 8x60K (Agilent Technologies, CA, USA) with 42,405 oligonucleotide probes (60-mer), representing 29,271 annotated genes, were used for hybridisation according to the manufacturer’s instructions (Method S3). The data has been made available in the National Centre for Biotechnology Information’s (NCBI) Gene Expression Omnibus (GEO) and are accessible through GEO series accession number GSE85426 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85426).
2.3 Real-time quantitative-PCR (RT-qPCR) of blood samples from the training and validation cohorts

Only six genes with FC ≥ 2.0 (namely DDIT4, CNOT8, SESN1, MAP2K1, ABCA9 and UCP3) were selected based on their significant up- or downregulation (p < 0.05) for verification of the microarray results using the RT-qPCR technique. Thirty AD subjects and 30 NDC subjects were randomly selected from the training cohort. Total RNA (1 µg) was reverse transcribed (20 µL) to cDNA using the Tetro cDNA Synthesis Kit (Bioline, USA). The ensemble database was referred for primer design (www.ensembl.org) and synthesised by Integrated DNA Technologies (Table S1). RT-qPCR was performed in triplicates using the Corbett 3000 RotorGene (Corbett Research, NSW, Australia) (Method S4). The conditions of the RT-qPCR used were based on the two-step cycling protocol described in the manual of the SensiFAST SYBR® No-ROX Kit: initial polymerase enzyme activation step (95°C for 2 mins), followed by 40 cycles of denaturation (95°C for 5 secs) and annealing/extension (acquired at the end of step; 60°C for 15 secs). The FC in AD was determined by the Pfaffl method (Pfaffl, 2001). The panel of six genes was finally evaluated in the validation cohort (independent of microarray study, Figure S1), which comprised of 74 subjects (25 AD, 25 NDC, 12 MCI and 12 VaD).

2.4 Statistical analysis

Statistical analyses for demographic and biochemical data were performed using the GraphPad Prism Software version 6.0 (GraphPad Software Inc, CA, USA). The panel of six genes was finally evaluated in the validation cohort (independent of microarray study, Figure S1), which comprised of 74 subjects (25 AD, 25 NDC, 12 MCI and 12 VaD).

2.5 Statistical analysis

Statistical analyses for demographic and biochemical data were performed using the GraphPad Prism Software version 6.0 (GraphPad Software Inc, CA, USA). Data were presented as mean ± SD. All results with a p < 0.05 were considered to be significantly different. Computations of sensitivity and specificity for each possible cut-off point of the individual mRNA were performed using statistical software package (SPSS Version 17.1 for Windows) for analysis.

3.0 RESULTS

3.1 Training cohort: demography, clinical characteristics and medication history

Table 1 summarises the demography and clinical characteristics of the participants from the training cohort. A total of 184 participants were recruited. Only 180 samples of probable AD and NDCs (n = 90 per group) were included in the final data analysis. Four samples were excluded due to poor microarray sample quality. There was no significant difference in gender between the probable AD and NDC groups. The average age for probable AD patients and NDC subjects was 77.9 and 75.2 years old, respectively (p = 0.02). Although age was a risk factor for AD, the linear regression analysis found no association between the selected genes and age (see Subheading 3.2).

The mean Mini-Mental State Examination (MMSE), Instrumental Activity of Daily Living (IADL) and Basic Activity of Daily Living (BADL) scores were significantly different (p < 0.001) between probable AD patients and NDC subjects. MMSE scores of 21–26 indicate mild dementia, 10–20 indicate moderate dementia and 0–9 indicate severe dementia (Perneckzy et al., 2006). The mean MMSE score for probable AD patients was 17.3 with 40.0% of patients under the mild dementia category, 46.7% of patients under the moderate dementia category and 13.3% of patients under the severe dementia category. As for the NDC group, the mean MMSE score was 29.7. The mean IADL and BADL scores for NDC were 15.9 and 11.9, respectively. Probable AD patients, on the other hand, were associated with lower mean scores of IADL (4.7) and BADL (9.8), respectively. Regarding comorbidities, there was no significant difference between the groups of probable AD patients and NDC. Assessments like clinical dementia rating (CDR), were only performed for probable AD patients. The total homocysteine (tHcy) was significantly (p < 0.01) lower whilst the holotranscoabalamine (holoTC) and folate in plasma samples were significantly (p < 0.01) higher in probable AD subjects when compared to NDC. Probable AD patients and NDC subjects seemed equally involved in physical activities. Both probable AD patients and NDC exhibited no significant difference in physical activities, smoking, alcohol consumption and years of education. Table S2 shows the medication history of the participants from the training cohort. Amongst the medications, the usage of acetylcholinesterase inhibitor (AChEI), N-methyl-D-aspartate (NMDA) receptor antagonists and statins were significantly different (p < 0.05) between the groups of probable AD patients and NDC.

3.2 Training cohort: diagnostic classifier that distinguishes AD patients from NDC

The discovery of significant genes is an essential step in constructing a precise classification model. By using the Benjamini Hochberg false discovery rate (FDR) multiple testing correction with fold change (FC) ≥ 2.0 as cut-off at p < 0.05, a total of 299 genes out of 29,271 annotated genes were selected. Further to the removal of unidentified genes, 172 genes remained, with 132 genes being downregulated and 40 genes being upregulated. To differentiate gene expression between the two groups, Z-score transformation was performed. The 50 highest-ranked probes based on the Z score (Table S3)
were used to build a partial least square discrimination analysis (PLSDA) model for the training cohort (Table 2). The PLSDA model was presented with an overall accuracy of 88.3%, sensitivity of 90.0% and specificity of 86.7%. There was a good separation between the two groups (probable AD and NDC) with an area under the curve (AUC) of 0.88.

Although the set of 50 genes yielded higher accuracy, sensitivity and specificity, using too many genes as a panel for biomarker discovery is certainly not practical. The PLSDA model was then tested on 25, followed by 12 and 10 highest ranked genes based on the Z score (Table 2). The results showed that the set of 12 genes produced better accuracy (83.9%), sensitivity (82.2%) and specificity (85.6%) than the set of 10 genes in distinguishing the groups (Table 2). Of the 12 genes, only six genes (CNOT8, DDIT4, SESN1, MAP2K1, ABCA9 and UCP3) that were unaffected by the stage of AD, gender and ethnicity were selected. For stages of AD, subjects were retrospectively graded according to the MMSE score: MMSE score of ≥ 27 indicates None, 21-26 indicates mild, 10-20 indicates moderate, 0-9 indicates severe dementia. Besides, the six selected genes were also unaffected by age as indicated by the linear regression analysis (CNOT8, p = 0.879; DDIT4, p = 0.344; SESN1, p = 0.578; MAP2K1, p = 0.412; ABCA9, p = 0.268; and UCP3, p = 0.112). The selected six genes from the microarray analysis were then verified using RT-qPCR.

Table 1: Demographics and clinical characteristics of participants of the training cohort

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Microarray</th>
<th>p-value$^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AD (n = 90)*</td>
<td>NDC (n = 90)*</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>42</td>
<td>48</td>
</tr>
<tr>
<td>Female</td>
<td>48</td>
<td>42</td>
</tr>
<tr>
<td>Age in year*</td>
<td>77.9 ± 5.7</td>
<td>75.2 ± 7.2</td>
</tr>
<tr>
<td>Ethnicity:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>59</td>
<td>56</td>
</tr>
<tr>
<td>Indian</td>
<td>13</td>
<td>23</td>
</tr>
<tr>
<td>Malay</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>Education in year †</td>
<td>9.2 ± 5.4</td>
<td>11.1 ± 3.7</td>
</tr>
<tr>
<td>MMSE</td>
<td>17.3 ± 6</td>
<td>29.7 ± 1.2</td>
</tr>
<tr>
<td>CDR</td>
<td>1.6 ± 0.7</td>
<td>NA</td>
</tr>
<tr>
<td>IADL</td>
<td>9.8 ± 3.3</td>
<td>11.9 ± 0.1</td>
</tr>
<tr>
<td>Total homocysteine</td>
<td>14.8 ± 4.7</td>
<td>12.6 ± 4.1</td>
</tr>
<tr>
<td>Holotranscobalamine</td>
<td>93.4 ± 63.9</td>
<td>120.1 ± 64.0</td>
</tr>
<tr>
<td>Folate</td>
<td>8.0 ± 5.2</td>
<td>10.1 ± 4.9</td>
</tr>
<tr>
<td>Total cholesterol</td>
<td>4.9 ± 1.1</td>
<td>4.8 ± 1.1</td>
</tr>
<tr>
<td>Low density lipoprotein</td>
<td>2.5 ± 0.9</td>
<td>2.6 ± 1.0</td>
</tr>
<tr>
<td>Smoking (yes %)</td>
<td>16 (17.8%)</td>
<td>11 (12.2%)</td>
</tr>
<tr>
<td>Alcohol (yes %)</td>
<td>20 (22.2%)</td>
<td>20 (22.2%)</td>
</tr>
<tr>
<td>Physical activity (yes %)</td>
<td>49 (54.4%)</td>
<td>51 (56.7%)</td>
</tr>
<tr>
<td>Hypertension (yes %)</td>
<td>46 (51.1%)</td>
<td>38 (42.2%)</td>
</tr>
<tr>
<td>Cardiovascular disease (yes %)</td>
<td>14 (15.6%)</td>
<td>13 (14.4%)</td>
</tr>
<tr>
<td>Stroke/ history of stroke (yes %)</td>
<td>5 (5.6%)</td>
<td>1 (1.1%)</td>
</tr>
<tr>
<td>Hyperlipidaemia (yes %)</td>
<td>21 (23.3%)</td>
<td>19 (21.1%)</td>
</tr>
<tr>
<td>Diabetes mellitus (yes %)</td>
<td>29 (32.2%)</td>
<td>24 (26.7%)</td>
</tr>
<tr>
<td>Cancer (yes %)</td>
<td>1 (1.1%)</td>
<td>5 (5.6%)</td>
</tr>
<tr>
<td>Traumatic brain injury (yes %)</td>
<td>3 (3.3%)</td>
<td>8 (8.9%)</td>
</tr>
</tbody>
</table>

*mean ± SD; *p < 0.05; **p < 0.01; ***p < 0.001; †Number of years spent at school, college or university; $The respective p-value of categorical data was derived from chi-squared test whereas the respective p-value of continuous data was derived from independent t-test; Abbreviations: AD, Alzheimer’s disease; BADL, basic activities of daily living; CDR, clinical dementia rate; IADL, instrumental activities of daily living; MMSE, Mini-mental State Examination; NA, data not available; NDC, non-dementia controls; SD, standard deviation.
To verify the microarray results, 60 subjects (30 AD vs 30 NDC) were randomly selected from the training cohort. The expression levels of the six genes (Figure 1), which were significantly different (p < 0.05) between AD and NDC groups from the microarray analysis, yielded a similar trend in the RT-qPCR (Pearson’s r = 0.954, p = 0.0002). The FC and regulation of the genes are as follows: CNOT8 (array = 2.40, RT-qPCR = 5.75); MAP2K1 (array = 2.35, RT-qPCR = 3.80); DDIT4 (array = 2.29, RT-qPCR = 2.94); SESN1 (array = 2.09, RT-qPCR = 3.29); ABCA9 (array = 2.13, RT-qPCR = 4.55) and UCP3 (array = 2.09, RT-qPCR = 4.55). The FC obtained using the microarray, irrespective of whether the genes were upregulated or downregulated.

The 172 genes (p < 0.05) with FC > 2 from the GeneSpring analysis were then determined for the pathways involved by using the IPA software. A total of 15 substantively dysregulated canonical pathways were recognised (Table S4) and they included Granzyme A signalling, chemokine signalling, STAT3 signalling, mitochondrial dysfunction, lipid signalling (protein kinase A) and various nervous system related signalling (e.g., axonal guidance and gap junction signalling). The top-ranked biological functions comprised of infectious disease, respiratory disease, inflammatory response, cell death and survival, cell-to-cell signalling and interaction and organ and tissue morphology. A total of five networks and interactions were identified by the Ingenuity Pathway Analysis (IPA) based on the functional roles of the genes. The genes selected for RT-qPCR validation amongst the probable AD participants were from the top four molecular networks: DDIT4, MAP2K1 (network 1), CNOT8 (network 2), UCP3 (network 3), ABCA9, SESN1 (network 4). Functional annotation was performed using the Database for Annotation Visualization and Integrated Discovery (DAVID) web-accessible program. Out of the 172 genes, 139 genes were identifiable by DAVID, out of which 41 belong to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. DAVID functional annotation chart analysis of genes showed substantial enrichment of acetylation and nucleosome.

### Table 2: Prediction based on the PLSDA using microarray data*

<table>
<thead>
<tr>
<th>Number of genes</th>
<th>Accuracy (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>88.3</td>
<td>90.0</td>
<td>86.7</td>
<td>ULK3, PSMD3, POLR2B, RTCB, CCND2, SPG7, SEC16A, ACTG1, TTC38, SNORA73A, XLOC_014512, SNORD38-1, DEFA3, CNOT8, DDIT4, SESN1, MAP2K1, SPOC1D1, CSAR1, CAMP, HAPLN2, FRBSL1, UCP3, IQSEC3, ABCA9</td>
</tr>
<tr>
<td>25</td>
<td>85.0</td>
<td>83.3</td>
<td>86.7</td>
<td>CNOT8, DDIT4, SESN1, MAP2K1, SPOC1D1, CSAR1, CAMP, HAPLN2, FRBSL1, UCP3, IQSEC3, ABCA9</td>
</tr>
<tr>
<td>12</td>
<td>83.9</td>
<td>82.2</td>
<td>85.6</td>
<td>CNOT8, DDIT4, SESN1, MAP2K1, SPOC1D1, CSAR1, CAMP, HAPLN2, FRBSL1, UCP3, IQSEC3, ABCA9</td>
</tr>
<tr>
<td>10</td>
<td>78.3</td>
<td>78.9</td>
<td>77.8</td>
<td>CNOT8, DDIT4, SESN1, MAP2K1, SPOC1D1, CAMP, HAPLN2, UCP3, IQSEC3, ABCA9</td>
</tr>
</tbody>
</table>

*Values are based on GeneSpring PLSDA from 90 AD and 90 NDC subjects

### 3.3 Training cohort: identification of pathways and biologically relevant network based on microarray

A total of 172 genes were tested for expression between the AD and NDC groups using both microarray and RT-qPCR. The 172 genes with FC > 2 were subjected to functional annotation using the DAVID web-accessible program. The top four molecular networks identified were:

1. **DDIT4, MAP2K1** (network 1)
2. **CNOT8** (network 2)
3. **UCP3** (network 3)
4. **ABCA9, SESN1** (network 4)

Functional annotation chart analysis of the genes showed substantial enrichment of acetylation and nucleosome.

### 3.4 Validation cohort: demography and clinical characteristics

The validation cohort (Table 3) was made up of 30 (40.5%) male and 44 (59.5%) female participants. The MMSE, IADL and BADL scores for the AD patients were significantly lower when compared to the NDC individuals (p < 0.001), followed by VaD as well as MCI. No significant difference was observed for all other tested parameters.
3.5 Validation cohort: Prediction of AD based on the expressions of the six selected genes using RT-qPCR

Six selected genes were analysed using the RT-qPCR technique (Table S5). Figure 2a illustrates the FC of gene expression in probable AD, MCI and VaD relative to NDC. The FC (AD vs NDC) of the upregulated genes (CNOT8, DDIT4, SESN1 and MAP2K1) were 9.21 (p = < 0.0001), 3.24 (p = 0.0244), 6.32 (p < 0.0001) and 9.07 (p = 0.0001), respectively. Except for the SESN1, the FC (MCI vs NDC) of ABCA9, UCP3, CNOT8, DDIT4 and MAP2K1 expression levels were also significantly different (p < 0.05). The SESN1 gene expression was not significantly different between VaD patients (p = 0.468) and NDC.

It was found that the gene expression of ABCA9 was able to correctly classify probable AD patients at a sensitivity and specificity of about 91% and an AUC of 0.94 (Table 4). The respective sensitivity and specificity of the remaining genes were between 73-86% and 74-83%,
respectively, with AUC ranging between 0.81-0.93. When the classifier was based on a combination of all six genes (Table 4), 41 out of 50 subjects were correctly classified, yielding an accuracy of 82.0%. More specifically, 21 of 25 (sensitivity of 84.0%) probable AD patients and 20 of 25 (specificity of 80.0%) NDC participants were correctly classified. The Positive Likelihood Ratio (PLR) was 8.20. The combined six genes produced an AUC value of 0.86.

Figure 2: The expression of six genes in the validation cohort and the efficacy of the identified disease classifier in other neurodegenerative groups. (A) Bars indicate mean fold of 74 participants: 25 probable AD, 12 MCI, 12 VaD and 25 NDC subjects. The NDC subjects are set at 1. Bars with common superscripts differ significantly at *p < 0.05 AD vs NDC, #p < 0.05 MCI vs NDC, +p < 0.05 VaD vs NDC. Before generating the graph, the raw data were log transformed as the log transformed data were suitable for plotting graphs. The data was found to be non-normally distributed as determined by the Wilks-Shapiro test for normality, thus the Mann-Whitney U test was used. (B) Gene expression scores between probable AD patients (in red) and MCI (in black). In the test cohort, 10 of the 12 MCI subjects were assigned to the correct class with the specificity of 83.3% for MCI as compared to probable AD. (C) Gene expression scores between probable AD patients (in red) and VaD (in black). Nine of 12 VaD patients were correctly classified NDC, with the specificity of 75.0% when compared to probable AD.
Table 4. Predictive capacity of six AD-associated genes from probable AD patients (n = 25) and NDC (n = 25)

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>AUC</th>
<th>PLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCA9</td>
<td>90.90</td>
<td>91.30</td>
<td>91.10</td>
<td>0.94</td>
<td>9.57</td>
</tr>
<tr>
<td>UCP3</td>
<td>77.30</td>
<td>73.90</td>
<td>75.60</td>
<td>0.81</td>
<td>6.06</td>
</tr>
<tr>
<td>CNOT8</td>
<td>72.70</td>
<td>78.30</td>
<td>75.60</td>
<td>0.84</td>
<td>3.40</td>
</tr>
<tr>
<td>DDIT4</td>
<td>77.30</td>
<td>77.30</td>
<td>77.30</td>
<td>0.80</td>
<td>3.25</td>
</tr>
<tr>
<td>SESN1</td>
<td>86.40</td>
<td>82.60</td>
<td>84.40</td>
<td>0.93</td>
<td>6.06</td>
</tr>
<tr>
<td>MAP2K1</td>
<td>80.00</td>
<td>78.30</td>
<td>79.10</td>
<td>0.84</td>
<td>4.19</td>
</tr>
<tr>
<td>Combination of all six genes</td>
<td>84.00</td>
<td>80.00</td>
<td>82.00</td>
<td>0.86</td>
<td>8.20</td>
</tr>
</tbody>
</table>

Abbreviations: AUC, Area under curve; PLR, Positive likelihood ratio; ABCA9, ATP Binding Cassette Subfamily A Member 9; UCP3, Uncoupling Protein 3; CNOT8, CCR4-NOT Transcription Complex Subunit 8; DDIT4, DNA-damage-inducible transcript 4; SESN1, Sestrin 1; MAP2K1, Mitogen-Activated Protein Kinase 1

Note: The analysis was based on individual gene using SPSS software

In terms of correlation between gene expression and MMSE scores (Figure 3), ABCA9, MAP2K1 and SESN1 were strongly correlated with MMSE scores which indicated the severity of the disease (ABCA9, r = 0.72, p = 0.0002; MAP2K1, r = -0.74, p = 0.0002; SESN1, r = -0.73, p = 0.0001). Lower expression of the ABCA9 gene was associated with a lower MMSE score whereas higher expression of MAP2K1 and SESN1 genes was associated with lower MMSE scores. Other genes showed a moderate correlation between their expression level and MMSE scores (UCP3, r = 0.47, p = 0.03; CNOT8, r = -0.65, p = 0.001 and DDIT4, r = -0.62, p = 0.002).

Figure 3. The correlation between MMSE score against six selected genes for RT-qPCR of the validation cohort. All genes showed a significant correlation with MMSE score in AD patients and NDC (ABCA9, r = 0.72; p = 0.0002; UCP3, r = 0.47; p = 0.03; CNOT8, r = -0.65; p = 0.001; DDIT4, r = -0.62; p = 0.002; SESN1, r = -0.73; p = 0.0001; MAP2K1, r = -0.74; p = 0.0002).
3.6 Validation cohort: performance of diagnostic classifier for MCI and VaD
The MCI group was included in the test cohort to evaluate whether the diagnostic blood gene expression classifier could be a biomarker for the early stage of cognitive dysfunction. Of the 12 MCI subjects, 10 individuals were classified as MCI (specificity of 83.3%) whereas two were as probable AD (Figure 2B). Although the number of subjects was relatively low, the present findings indicated the potential predictive power of the identified gene expression signature. As a high percentage of MCI was expected to possess a likely AD-endpoint, the current results showed that they were sufficiently different from probable AD to be classified as MCI.

The AD classifier appeared to make only a small distinction, in favour of classifying VaD subjects who shared a lot of pathological neurodegeneration processes and were close to be diagnosed as probable AD. Of the 12 VaD patients with acceptable RT-qPCR quality, 9 (75.0%) were correctly predicted as non-AD by the disease classifier (Figure 2C). These results might represent a marker of diseases sharing common aetiology.

4.0 DISCUSSION
Given that a list of <10 biomarkers would be more robust and more practical for clinical testing purposes (Xia et al., 2013), the present study selected six genes (DDIT4, CNOT8, MAP2K1, SESN1, ABCA9 and UCP3), which were not associated with stage of disease, gender or ethnicity, for validation using RT-qPCR in both training and validation cohorts. Validation of these six genes gave rise to excellent sensitivity (84.0%), specificity (80.0%), accuracy (82.0%) and AUC of 0.86. ABCA9 gene, in particular, discriminated probable AD patients from NDCs with high sensitivity (90.9%), specificity (91.3%), accuracy (92%), AUC of 0.94 and high correlation with MMSE test (r = 0.72, p = 0.002).

Several studies on gene expression data have uncovered valuable patterns from biopsy or autopsy-based samples but these findings are difficult to be extrapolated to clinical settings. Some of the early studies on blood gene expression in AD had successfully identified a list of biomarkers with sensitivity and specificity > 80.0% (Bai et al., 2014; Booij et al., 2011; Fehlbaum-Beurdeley et al., 2010; Maes et al., 2007). The expression values of AD-related genes obtained from recent studies using blood samples of AddNeuroMed1 and 2 (ANM1 and ANM2) datasets also exhibited AUC >0.8 (Lee & Lee, 2020) and could classify AD from healthy control. Voyle et al. (2016), who used gene expression data from the ANM and Dementia Case Registry (DCR) cohorts, obtained an AUC of 0.74. There was, however, another gene expression study (Patel et al., 2020) that had found biomarkers with low sensitivity and low specificity (<80.0%). The list of significantly dysregulated genes in probable AD patients obtained in this study was, however, different from those of previous AD blood-based gene expression studies (Griswold et al., 2020; Lunnnon et al., 2013; Nho et al., 2020; Niculescu et al., 2020; Ou et al., 2021; Panitch et al., 2022; Park et al., 2020; Park et al., 2021; Patel et al., 2019). The present gene enrichment analysis found dysregulated pathways related to oxidative stress, mitochondrial dysfunction, apoptosis, inflammation, DNA damage and perturbed lipid metabolism in probable AD. In spite of the different genes obtained when compared to the previous studies, inflammation and mitochondrial dysfunction seemed to be the common pathways involved in AD pathogenesis (Griswold et al., 2020; Lee & Lee, 2020; Lunnnon et al., 2012; Voyle et al., 2016).

The dysregulated genes were analysed to determine the diagnostic potential of whether these genes were part of AD pathology or un specific to the pathology by calculating the changes in gene expression, sensitivity and specificity of selected genes. The comparison was performed between probable AD patients and NDC subjects and also other neurological patients such as MCI and VaD. The importance of using other neurodegenerative diseases (MCI and VaD as in the present study) was to test whether the selected genes were specific towards only probable AD. MCI, which is a transition between normal aging and early dementia (Lovell & Markesbery, 2007), is regarded as a risk of dementia, especially AD. In spite of the small sample size, the present results of the six genes showed good separation in MCI. Profiles of ABCA9, SESN1 and CNOT8 genes, in particular, supported the hypothesis that MCI is a transition between normal aging and early dementia. DDIT4 gene was upregulated in MCI and AD, with a greater extend of upregulation in MCI. Nevertheless, the upregulation of UCP3 gene and downregulation of MAP2K1 gene exhibited profiles opposite to that in probable AD patients.

The present study had also included VaD in the test cohort. VaD is a syndrome and pathologic subtype that includes ischaemic and haemorrhagic strokes, cerebral hypoxic-isaemic events and senile leukoencephalopathic lesions (Román et al., 1993). VaD was chosen in the present study because it is the second
commonest type of dementia. By using the six selected
genes, this study found the specificity of VaD to be 75%.
VaD may possess only minor neuropathological changes
of AD (Meyer et al., 2002). As such, the specificity of VaD
in the current study was believed to adequately indicate
that the constructed classification algorithm was specific
in picking up the changes in gene expression that might
have occurred in the blood of AD patients. There is,
however, a lack of previous study that had used VaD as
part of their test cohort.

**Figure 4** illustrates the involvement of the six genes in
pathways related to the pathogenesis of AD. Chronic
activation of the NMDA receptor may upregulate the
mitogen-activated protein kinase (Amadoro et al., 2006;
Wan et al., 2012) via extracellular-signal-regulated
kinase (ERK) phosphorylation (Sun et al., 2016) (Figure
4A) that could be accompanied by increased MAP2K1
expression that may cause inflammatory response
(Wang et al., 2014). Dysregulated mitochondria which
were manifested through the downregulation of UCP3
gene (Figure 4B), would result in oxidative stress
(Thanan et al., 2015) and DNA damage.

With regards to oxidative stress, upregulation of SESN1
gene has been identified as an implication of perturbation of the mitochondria process in AD (Figure
4C). Oxidative stress could be derived from excessive
mitochondria ROS production. Their reactive end
products could damage DNA through the upregulation of
the DDIT4 gene (Figure 4D). Lipid dysregulation
(Figure 4E), which was primarily found to be related to
AD (Wong et al., 2017), could be associated with
downregulation of ABCA9 gene. In this study, ABCA9
gene was found to exhibit the highest sensitivity and
specificity. This warrants further investigation to
elucidate the function of this gene in AD pathogenesis.
ABCA9 gene may play a role in monocyte differentiation
and lipid homeostasis (Piehler et al., 2002). Transcriptional expression of this gene could be induced
during monocyte differentiation into macrophages and
suppressed by cholesterol import. As an ATP-binding
cassette (ABC) transporter gene, ABCA9 plays essential
roles in mediating cholesterol efflux by regulating
cellular cholesterol homeostasis (Li et al., 2013).

Although cholesterol is a major component of the
mammalian cell membrane, the accumulation of excessive cholesterol is toxic to cells. This would in turn,
impair cell signalling which would cause impairment in
synaptic integrity and neurotransmission. Recent
findings in AD research indicated disturbance of Aβ
exportation at the brain’s barriers, which was
physiologically facilitated by the ABC transporter
superfamily, might play a fundamental role in AD
initiation and progression. Previous studies showed
several ABC transporters, such as ABCA1, ABCB1,
ABCG1, ABCG5, and ABCG8 to play essential roles in
mediating cholesterol efflux by the regulation of cellular
cholesterol homeostasis (Chen et al., 2011; ElAli &
Rivest, 2013; Li et al., 2013). Only little is known about
the function of the subgroup of ABCA6-like transporters
which form a compact gene cluster located on chr
17q24.2-3. This cluster comprises the transporters
ABCA5, ABCA6, ABCA8, ABCA9, and ABCA10. Although
ABCA9 is expressed at detectable levels in the brain and
is likely involved in lipid transport processes, the
potential implication in neurodegeneration remains
purely speculative at this point (Pereira et al., 2012,
2018). On the other note, all putative mechanisms that
lead to neuronal death in AD (by apoptosis) could be
correlated with the upregulation of CNOT8 gene
expression (Figure 4F). Any functional defects in the
regulation of the deadenylation activity by CNOT8 gene
could induce p53 level, which could lead to apoptosis.
The analysis revealed that molecular perturbation in AD
patients tend to be shared widely, vary significantly and
substantially overlaps within several confounding
factors.

The present study acknowledges several limitations.
There was a lack of clinical information from
neuroimaging data, CSF analysis and more established
dementia-rating scale. Besides, this study also
encountered challenges in identifying and characterising
unknown genes. The microarray technique relies upon
existing knowledge about the genome sequence and is
limited by the availability of only several databases.
Furthermore, microarray has limited dynamic detection
range owing to background and saturation signals. Given
these limitations, the results reported in this study are
exploratory and should be interpreted conservatively.
On another notes, the present study acknowledges the
usefulness of longitudinal gene expression studies in
supporting AD diagnosis and monitoring from the
prodromal to the symptomatic stage. As such, it would
be beneficial for future validation work to include more
patients with well-characterised MCI and other
dementing disorders (PD, Lewy Body Dementia, VaD) as
well as asymptomatic patients with preclinical disease to
validate AD-specific biomarkers.
Figure 4: Genes in AD-related biological pathways. AD-related pathways could be accelerated via several pathways that lead to degeneration. (A) Activation of immune response by stressors such as Aβ and tau activates physiological activators of extracellular-signal-regulated kinase (ERK) via elevation of MAP2K1 gene, leading to aberrant protein phosphorylation and enhanced inflammation. (B) Dysregulated mitochondria resulting from inhibition of UCP3 gene could lead to increased mitochondrial reactive oxygen species (ROS) production. (C) ROS causes oxidative stress, leading to brain oxidative impairment with the involvement of SESN1 gene. (D) Oxidative DNA damage have been largely found in brain region of AD associated with the upregulation of DDIT4 gene. (E) Lipid dysregulation related to the downregulation of ABCA9 gene would lead to neuronal cell death. (F) Multiple cellular functions of p53 appear to be associated with increased CNOT8 gene expression that could lead to induction and regulation of cell cycle arrest and apoptosis. Deadenylation activity by CNOT8 could induce p53 level in response to hypoxia, DNA damage and then cell death through apoptosis.

5.0 CONCLUSIONS
The present study had revealed six genes (i.e., ABCA9, UCP3, MAP2K1, SESN1, CNOT8 and DDIT4) that might be implicated in AD pathogenesis. This gene panel seems to be associated with inflammation, mitochondrial dysfunction, oxidative injury, DNA damage and apoptosis. Another important pathway highlighted in this study is the lipid metabolism pathway through the downregulation of ABCA9 gene, that would lead to neuronal cell death.

Supplementary Materials: The following are available online at https://neuroscirn.org/ojs/index.php/nrnotes/article/view/262, Figure S1: Workflow of the present study, Table S1: Primers used in the study for validation of microarray data, Table S2: Medication history, Table S3: Fifty genes that best differentiated probable AD patients from NDC subjects in the training cohort, Table S4: Canonical pathways that were significant in AD based on the Ingenuity Pathway Analysis, Table S5: Standard curves of the RT-qPCR analysis, Method S1: Recruitment and assessment of participants, Method S2: Blood biochemical profile and RNA extraction, Method S3: Microarray, Method S4: Real time quantitative-PCR (RT-qPCR) of blood samples from the training and validation cohorts

Acknowledgements: This Transforming Cognitive Frailty to Later-Life Self-sufficiency (AGELESS) study is funded by the Ministry of Higher Education Malaysia under the Long-Term Research Grant Scheme [LRGS/1/2019/UM/01/1/3].

Author Contributions: KR was involved in conceptualisation and project administration; ABAM was responsible for funding acquisition; KR and ABAM were involved in supervision; AZS, AVC, MPT, SBK were involved in investigation; AZS, KR, BI performed formal analysis; AZS wrote the original draft; AZS, KR, SML, AVC, MPT, SBK, BI, ABAM reviewed and edited the original draft.

Conflicts of Interest: The authors declare no conflict of interest.
References


