ELSEVIER

Neurobiology of Aging xx (2012) xxx

NEUROBIOLOGY OF AGING

www.elsevier.com/locate/neuaging

# Hippocampus neuronal metabolic gene expression outperforms whole tissue data in accurately predicting Alzheimer's disease progression

Shiri Stempler[a,*], Yedael Y. Waldman[b], Lior Wolf[b], Eytan Ruppin[a,b,*]

[a] The Sackler School of Medicine—Tel Aviv University, Tel Aviv, Israel

[b] The Blavatnik School of Computer Science—Tel Aviv University, Tel Aviv, Israel

Received 13 July 2011; received in revised form 1 January 2012; accepted 5 April 2012

**Abstract**

Numerous metabolic alterations are associated with the impairment of brain cells in Alzheimer's disease (AD). Here we use gene expression microarrays of both whole hippocampus tissue and hippocampal neurons of AD patients to investigate the ability of metabolic gene expression to predict AD progression and its cognitive decline. We find that the prediction accuracy of different AD stages is markedly higher when using neuronal expression data (0.9) than when using whole tissue expression (0.76). Furthermore, the metabolic genes' expression is shown to be as effective in predicting AD severity as the entire gene list. Remarkably, a regression model from hippocampal metabolic gene expression leads to a marked correlation of 0.57 with the Mini-Mental State Examination cognitive score. Notably, the expression of top predictive neuronal genes in AD is significantly higher than that of other metabolic genes in the brains of healthy subjects. All together, the analyses point to a subset of metabolic genes that is strongly associated with normal brain functioning and whose disruption plays a major role in AD.



*Keywords:* Alzheimer's disease; Metabolism; Mini-Mental State Examination (MMSE); Gene expression; Classification model; Regression; Hippocampus; Neuron

## 1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disease that is characterized by cognitive decline and is the most common cause of dementia (www.alz.org). Because the incidence and prevalence of AD and other dementias increase with age, the number of patients is expected to grow rapidly as the population ages. Pathological hallmark abnormalities of AD are the extracellular deposits of $\beta$-amyloid peptide (plaques) and intracellular twisted strands of tau protein (tangles) in the brain (Johnson and Bailey, 2002; Masters et al., 1985). Current treatment is mainly symptomatic, and there is no treatment available to stop the deterioration of brain cells in AD. Definitive diagnosis of AD requires postmortem examination of the brain, which must contain sufficient numbers of plaques and tangles to qualify as affected by AD (Mattson, 2004). Plaques and tangles are present mainly in brain regions involved in learning, memory, and emotional behaviors, such as the entorhinal cortex, hippocampus, basal forebrain, and amygdala. Unraveling the mechanisms underlying AD and impaired brain function has been difficult because of the complexity of the cellular networks that drive these changes. At present, the only gene that has been consistently associated with sporadic cases of AD is apolipoprotein E (APOE) (Mihaescu et al., 2010). However, APOE genotyping is not considered clinically useful for screening, testing, or diagnosis of AD.

Until a definite diagnosis is confirmed neuropathologically, the diagnosis of AD is based on clinical examination and neuropsychological testing. The cognitive performance in AD subjects is assessed via the Mini-Mental State Examination (MMSE) (Folstein et al., 1975). In addition to its

* Corresponding author at: The Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel. Tel.: +972 3 6406528; fax: +972 3 6409357

*E-mail address:* shirist2@post.tau.ac.il (S. Stempler) or ruppin@post.tau.ac.il (E. Ruppin).

0197-4580/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neurobiolaging.2012.04.003

value as a screening test for dementia, the MMSE is often used to document cognitive changes over time in individual patients. This is an important clinical measurement because progressive cognitive loss is a characteristic of neurodegenerative dementing illnesses (Clark et al., 1999). Information on the rate of change over time is also valuable for assessing the results of therapeutic interventions, predicting the severity of cognitive decline, and planning for long-term health care.

Several gene expression studies of AD individuals have been reported so far. Blalock et al. (2004) analyzed hippocampal gene expression microarrays of control subjects and AD patients of varying severity. They tested the correlation of each gene's expression with MMSE and neurofibrillary tangle (NFT) scores. Their work revealed upregulation of many transcription factor signaling genes regulating proliferation and differentiation during AD progression, including tumor suppressors, oligodendrocyte growth factors, and protein kinase A modulators. In addition, upregulation of adhesion, apoptosis, lipid metabolism, and initial inflammation processes was reported along with downregulation of protein folding/metabolism transport as well as several energy metabolism and signaling pathways. Ray and Zhang (2009) used the aforementioned microarrays of AD patients to develop a multiple linear regression (MLR) method to find the strength of the association of each subject's NFT score (dependent variable) with the gene expression profile and the MMSE. Using MLR, they selected 500 genes that can distinguish subjects with incipient AD from healthy control subjects in two different brain regions—the hippocampus and the entorhinal cortex. Liang et al. (2008, 2010) profiled the gene expression in non-tangle-bearing neurons in six postmortem brain regions that are differentially affected in the brains of healthy elderly control subjects, nondemented individuals with intermediate AD neuropathology (NDAD), and AD patients. They compared the expression of 80 nuclear genes encoding mitochondrial electron transport chain subunits in the different brain regions. In a second study, they focused on genes that participate in mechanisms that have been previously implicated as being associated with AD, to assess whether these pathogenic pathways may be enacted in NDAD brains (Liang et al., 2010). These mechanisms include pathways leading to the formation of NFTs and amyloid plaques, ubiquitin–proteasomal pathways, and pathways surrounding synaptic degeneration. Indeed, significant overlapping expression changes were identified in the brains of both NDAD and AD patients compared with those of control subjects.

Following on these important analyses of gene expression in AD, we present here a microarray-based study that focuses specifically on the role of metabolic genes in the cognitive decline of this disease. Our focus has been strongly motivated by cumulative evidence demonstrating that numerous metabolic alterations may cause the impairment of brain cells' function and viability in AD. Decreases in cerebral metabolic rate (CMR) characteristically occur in AD and other dementias (Blass, 2001). Reduced CMRs for glucose and $O_2$ are reported in many studies (Blass, 2001). Decreased activities of key enzymes in energy metabolism in brains of AD patients have also been reported in many studies. Examples for such enzymes are the cytochrome c oxidase, pyruvate dehydrogenase complex, and $\alpha$-ketoglutarate dehydrogenase complex (Chandrasekaran et al., 1994; Blass, 2001). Mitochondrial function is specifically altered in AD (Wang et al., 2009). Electron microscope studies have demonstrated the accumulation of abnormal mitochondria in senile plaques in AD (Terry et al., 1964). Damage to both the components and the structure of mitochondria, as well as increased oxidative stress, has been extensively reported in AD (Zhu et al., 2006). The mitochondrial respiratory chain is one of the main sources of reactive oxygen species (ROS) (Gibson et al., 2008), resulting in oxidative damage to varied molecules (Ferrer, 2009). The overall effect assumed is a positive feedback cycle where ROS produce oxidative stress that eventually produces more ROS (Bonda et al., 2010). The nervous system is particularly susceptible to oxidative stress (Barnham et al., 2004), as neurons are extremely energy dependent and therefore particularly sensitive to changes in mitochondrial function (Su et al., 2010). Additionally, several proteins linked with metabolic reactions have been shown to be targets of oxidative damage in AD (Butterfield et al., 2010; Reed et al., 2008; Sultana et al., 2006). Finally, other neurodegenerative diseases such as Parkinson's disease (PD) and amyotrophic lateral sclerosis (ALS) also share these common features of extensive oxidative stress (Su et al., 2010), mitochondrial damage, and apoptosis processes (Reed et al., 2008). The prevalence of neurological diseases associated with mutations in mitochondrial genes underscores the important functional role of mitochondria in neuronal metabolism (Barnham et al., 2004; Butterfield et al., 2010). Finally, our work has been further inspired by the recent work of Khaitovich et al. (2008), who studied the role of metabolic genes in another major brain disease—schizophrenia. They found that genes associated with schizophrenia are heavily involved in energy metabolism of the brain. Remarkably, the expression of many of these genes has also changed rapidly during recent human evolution, leading them to suggest that the evolution of human cognitive abilities was accompanied by important adaptive changes in brain metabolism.

In light of the accumulating evidence briefly reviewed earlier in the article, the aim of the current study is to focus on metabolic dimension to identify the metabolic genes and pathways that strongly correlate with AD progression and cognitive decline in the hippocampus, specifically in hippocampal neurons. Although some prime metabolic determinants of these observables may have previously surfaced during whole-genome mRNA analysis, many important findings would have most likely been lost because of the

masking effects of nonmetabolic genes and the large feature space, hence leading us to a study focused on metabolism per se. For this purpose, we trained random forest models (Breiman, 2001) for the classification of AD severity using metabolic genes solely. Additionally, we generated regression models for the prediction of the MMSE score from metabolic genes' expression. We compared the selected genes obtained from the models of the two data sets (i.e., whole hippocampus tissue vs. hippocampal neurons) to study the unique processes that occur in different cell types. Finally, we analyzed the selection forces acting on these genes in human and primate evolution to highlight mechanisms involved in the evolution of cognitive abilities.

## 2. Methods

### 2.1. Data sets

The microarray data used in this study were obtained from the Gene Expression Omnibus (GEO) site (www.ncbi.nlm.nih.gov). The first data set, containing expression data of hippocampus field CA1 neuronal genes, was taken from studies of six regions from postmortem brains (GSE5281) (Liang et al., 2008, 2010). The data set contains expression profiles of laser-capture microdissected non-tangle-bearing neurons from 29 subjects categorized into three groups termed: "Control," "NDAD," and "diagnosed AD". Because this data set does not contain associated MMSE scores, for the analysis of this database, we use classification models for these three disease-state classes.

For the hippocampus tissue analysis, we used gene expression data of hippocampal (CA1 field) specimens from 31 individuals (GSE1297) (Blalock et al., 2004). In addition, this data set contained the MMSE scores of the subjects, ranging from 2 to 30. Blalock et al. (2004) categorized the samples into four groups of AD severity, termed "Control" (MMSE $>$ 25), "Incipient AD" (MMSE 20–26), "Moderate AD" (MMSE 14–19), and "Severe AD" (MMSE $<$ 14), based on the MMSE scores. Statistical regression models that link the expression data of the metabolic genes and the MMSE scores are used to quantify and study the information content of gene expression with regard to the cognitive test outcome.

In both cases, the data were filtered for metabolic genes (Duarte et al., 2007). A list of these metabolic genes is available in the supplementary data (Table S1).

To study the variation in gene expression among different tissues, a third data set was taken from Su et al. (2004) (GDS596). A subset of 30 different nonredundant healthy adult tissues (Waldman et al., 2010) was used to perform a tissue-specific expression analysis of the genes selected by the classification and regression models, in comparison with the expression of the other metabolic genes.

For gene expression variation analysis between human and primates, we used an additional gene expression data set from the prefrontal cortex of humans, chimpanzees, and rhesus monkeys (Khaitovich et al., 2006). This data set was taken from the ArrayExpress Archive, experiment E-TABM-84 (Parkinson et al., 2011).

### 2.2. Building the regression and classification models

The total samples of each of the two data sets were randomly divided into training and test sets, consisting of 2/3 and 1/3 of the samples, respectively. This process was repeated 20 times to obtain 20 random partitions for each data set in a standard cross-validation procedure. Each of the training and test sets was sampled such that they contained similar ratios of severity classes across the entire data set, where for the hippocampus neuronal genes, we used the three classes mentioned earlier in the text, and for the hippocampus tissue, we used the four MMSE categories (Blalock et al., 2004).

Random forest classification and regression algorithms were used for generating the classification and regression models, respectively (Breiman, 2001). The Matlab implementation of Random Forest was trained on the metabolic gene expression data training sets (code.google.com/p/randomforest-matlab/). Default parameters were used (500 trees were grown). Conforming to standard procedure, training was performed on the training sets and performance evaluation on distinct test sets.

### 2.3. Obtaining lists of genes that are informative for AD progression

Performing feature (gene) selection, the genes with the highest importance values obtained from the models in each of the different random partitions were selected. Different cutoffs were chosen, and the same training data were used to retrain a model using the genes that were above the cutoff point. The cutoff that obtained the highest prediction accuracy or correlation for the classification and regression models was chosen to obtain the list of selected genes in each experiment. For the classification models, the accuracy is defined as the number of true predictions divided by the number of samples in the test set.

### 2.4. Determining the statistical significance of the models

To assess the significance of the regression/classification error rates, a permutation test was applied (Radmacher et al., 2002). In the permutation test, outcome MMSE/class labels were randomly assigned to each patient, and the entire model discovery process was repeated. For each of the 20 partitions, 1000 such permuted data sets were produced, and the permutation $p$ value was computed. This test was repeated for each of the models that was generated.

An additional test for assessing the significance of the models' results was carried out by selecting 1000 random sets of genes of the same size as the optimal cutoff that was chosen for each data set (i.e., 100 genes for the neuronal model and 50 genes for the whole tissue model). These genes were selected from the genes available on each of the

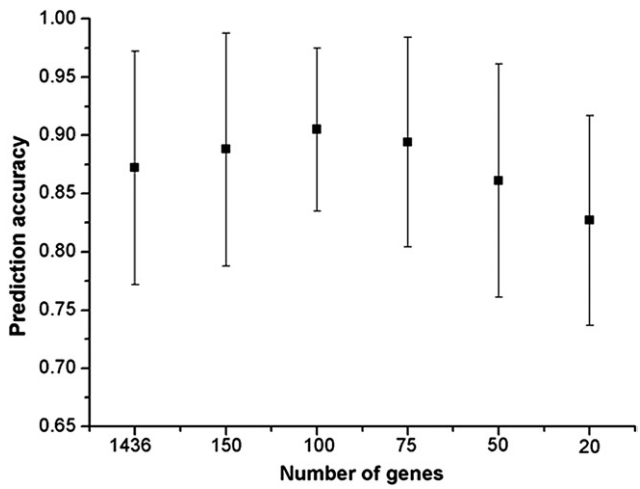


Fig. 1. Prediction accuracy of 20 random forest classification models built using cross-validation (Methods). The figure shows the mean prediction accuracy results (SD is shown as error bars) for the models that were generated based on all 1436 metabolic genes, and for the models based on different cutoffs of genes selected for their highest importance scores (top 150 genes, top 100 genes, etc.)

two microarrays. We then used these gene sets to generate 1000 independent random forest models (with 20 cross-validation models for each) and for calculating an empirical $p$ value.

### 2.5. Determining the tissue specificity of gene expression

To evaluate the functional importance of the genes selected by the models in terms of their expression in the brain, a third data set was used. These data contain gene expression of 30 adult human tissues, among which the entire brain tissue is also included (Su et al., 2004). Expression rate and breadth are defined as the mean expression across all tissues and the number of tissues in which a gene is expressed, respectively. A gene was defined as expressed in a tissue if its expression level was above 200 standard Affymetrix average-difference units (Su et al., 2004). We then compared, using Wilcoxon test, the expression rate and breadth of the list of the selected genes from the feature selection models with that of other metabolic genes.

### 2.6. Evolution patterns in primates

Expression profiles from the prefrontal cortex of 10 humans, 6 chimpanzees, and 6 rhesus monkeys were used for inter- and intraspecies gene expression variation analysis (Khaitovich et al., 2006). For interspecies variation (divergence) analysis, we measured for each gene, its squared difference from the mean expression, for each of two species. For intraspecies variation (diversity) analysis, we measured the standard deviation (SD) of each of the metabolic genes across the different samples for each of the species and divided it by its average. In both cases (inter- and intraspecies analysis), we compared the mean variation score of the model genes with that of 100,000 equal-size random sets of other metabolic genes, to obtain an empirical $p$ value.

## 3. Results and discussion

### 3.1. Neuronal metabolic genes in the hippocampus predict AD severity more accurately than the whole tissue metabolic genes

To compare the influence of the disease-altered metabolism of neurons with that of the whole tissue, we decided to build classification models of neuronal gene expression from the hippocampus and of gene expression of a whole hippocampus tissue, for the prediction of AD progression. For this purpose, we first used the data set of Liang et al. (2008, 2010). It includes gene expression from hippocampal neurons of patients with three levels of disease severity: control, NDAD, and AD. The data set was preprocessed as detailed in the Methods section, focusing on metabolic genes. Classification models were generated using random forest implemented in Matlab, for the prediction of AD severity (Methods). The accuracy of the classification predictions obtained was remarkably high—0.872 ($\pm$ 0.1), $p <$ 0.001 (Fig. 1). Notably, these prediction accuracy levels, obtained by analyzing the metabolic genes, were similar to those obtained when using all the genes on the chip.

For feature selection, we generated models based on different cutoffs on the number of genes selected (Fig. 1). The same training data were used to retrain a model using the selected genes that were above the cutoff point, and the prediction accuracy was evaluated on unseen test data. The results of the 20 cross-validation models containing the top 100 genes were most accurate, compared with other sizes of top selected gene sets (see Fig. 1 and Methods). The 100-gene models (of each of the 20 cross-validation variants) yield an accuracy of 0.9 ($\pm$ 0.07) in predicting AD severity (Fig. 2) (random sets of 100 genes each generated significantly lower results [empirical $p$ value of 0.038, Methods]).

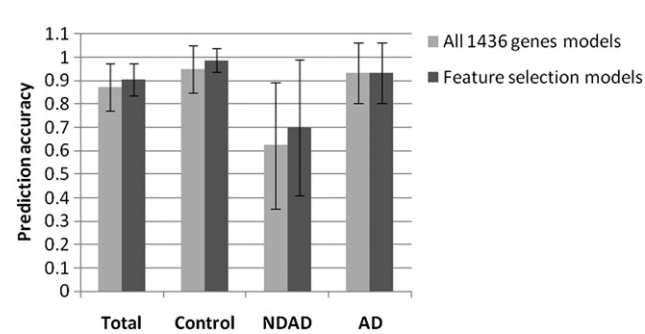


Fig. 2. Mean prediction accuracy of the various AD progression stages obtained across 20 random forest classification models built using cross-validation (Methods). SD is shown as error bar. The results of the models that were generated based on all 1436 metabolic genes and of the feature selection models (based on top 100 genes) are shown.

The data set of Blalock et al. (2004) of whole hippocampus tissue genes was used as our second data set and was preprocessed as described in the Methods section. These samples were categorized into four groups of AD severity: control, incipient, moderate, and severe (see Methods). For generating the classification models for predicting AD severity stages, the moderate and severe subjects were grouped into one class of advanced AD to allow for a direct comparison with the classification results obtained with neuronal data (see supplementary data). The classification models' prediction accuracy was lower than that obtained via the neuronal gene models (0.76) (these tissue-based classification models are described in the supplementary data).

Several reported works have focused on and emphasized the advantage of using gene expression from single cells. Hinkle and Eberwine (2003) indicated that single-cell molecular biology technology can help in the identification of fundamental disease mechanisms and in the diagnosis and treatment of neurologic disorders. An additional example is a recent work that compared gene expression from CA1 pyramidal neurons and regional hippocampal dissections in AD—highlighting a dilution effect that is likely to occur in regional microarray (Ginsberg et al., 2012). Furthermore, Kamme et al. (2003) have shown that even a single cell type is not homogeneous at the gene expression level. All the studies support our results, which show higher prediction accuracy when relying on single cell data (neuronal data). Furthermore, we show here that the information that is embedded in the metabolic genes in the neuronal cells is sufficient for predicting AD, highlighting the role of neuronal metabolic processes in the disease.

To further validate our findings and compare the role of metabolic genes in whole tissue data with that of neuronal data, we built additional classification models, generated based on two additional gene expression data sets: (1) whole tissue data of the cortex (mainly temporal cortex) (Webster et al., 2009), and (2) neuronal data from a temporal lobe subregion—the entorhinal cortex (Liang et al., 2008). The analysis of these independent data sets further corroborates the results of the hippocampus models in which the neuronal metabolic genes can predict the AD stage more accurately than whole tissue gene expression (see Fig. S1).

### 3.2. Hippocampus metabolic genes can predict the MMSE score of AD patients

Importantly for our goals, the hippocampus tissue data contain the MMSE scores of the subjects as well (which are regrettably lacking from the neuronal-specific data). To study the correlation between the metabolic genes and the cognitive decline in AD, we generated regression models for the prediction of the MMSE score, using random forest regression (see Methods). The correlation coefficient of the MMSE regression model test set prediction was marked—0.57 (± 0.19) with root mean square error (RMSE) of 6.36 (± 0.6). The significance of the results was computed with permutation tests ($p < 0.001$, Methods). Remarkably, this correlation is higher than the correlations between the MMSE and other sources of biological measurements that were reported in the past in the literature (Ho et al., 2005; Kennedy et al., 2004). Blalock et al. (2004) reported MMSE correlations with NFT levels in the hippocampus ($r = 0.45$) and with amyloid plaque levels ($r = 0.19$). Thus, metabolic gene expression is more strongly associated with MMSE cognitive function score in the AD hippocampus than the classical pathological markers of AD. As in the previous section, these prediction levels (obtained with the metabolic genes) are similar to those obtained from regression models generated from all the genes on the chip.

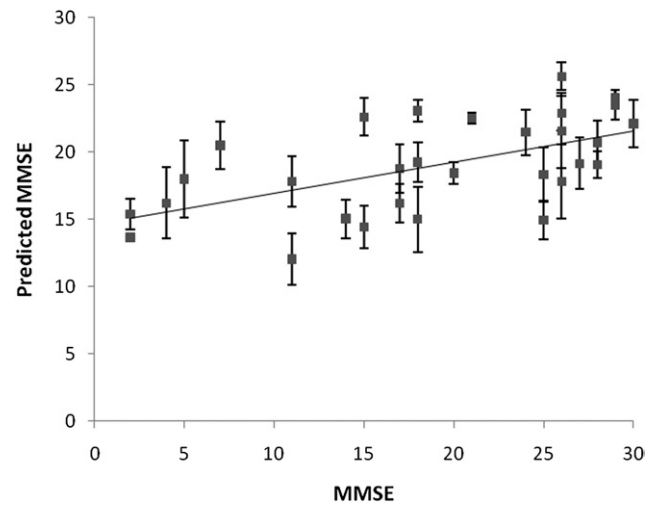


Fig. 3. Predicted MMSE scores vs. actual MMSE scores using feature selection-based regression models (SD of predicted MMSE scores is shown as error bar).

The random forest gene selection importance score was used for the selection of the most important metabolic genes associated with MMSE decline in AD (Fig. S2). Models generated from the top 50 genes were the smallest models yielding high correlations and were selected for further gene selection analysis (obtaining RMSE of 6.28 [± 0.9] and a correlation coefficient of 0.57 [± 0.18] [Fig. 3]). As in the case without gene selection, a permutation test on the MMSE score yielded $p < 0.001$ (Methods).

### 3.3. Analysis of the genes selected in the neuronal analysis

Beginning with a gene-level analysis, we note that three metabolic genes appeared in all of the 20 different neuronal cross-validation models: NDUFA10, which belongs to the electron transport chain; ME3, a mitochondrial enzyme that catalyzes the oxidative decarboxylation of malate to pyruvate and is downregulated in PD (Bossers et al., 2009); and RPN1, an essential subunit of *N*-oligosaccharyl transferase enzyme. It is part of the regulatory subunit of the 26S proteasome and may mediate binding of ubiquitin-like do-

mains to this proteasome (Vernace et al., 2007). Disruption of the ubiquitin/proteasome pathway is relevant to pathophysiological conditions that provoke the accumulation of aberrant proteins, such as in a variety of neurodegenerative disorders, including AD and PD (Vernace et al., 2007). Indeed, as one would expect, these three genes are all underexpressed in AD.

For further analysis, we selected a group of top 50 metabolic genes that were most frequent in the different 100-gene models (Table S2). These 50 genes appeared in at least 10 models (of the total 20 cross-validation models) and hence were considered the most important for the prediction of AD severity. Of the 50 genes, 34 genes were underexpressed as the disease progressed and 16 genes were overexpressed. According to Alzgene (www.alzgene.org), three metabolic genes were found to be significantly associated with AD in genome-wide association studies (GWAS) (Bertram et al., 2007). Of these genes, one gene is also found in the 50 genes selected by the model ($p$ value = 0.13). This gene is *MTHFR*, a member of the folate metabolism pathway. Recent studies have revealed association between this gene and the susceptibility to develop late-onset Alzheimer's disease (LOAD) (Kim et al., 2008; Li et al., 2009; Wang et al., 2005). Several other genes (among the 50 metabolic genes) were found to be related to AD in various experimental studies, and they are listed in the supplementary Table S2. The top AD predictive genes of the whole hippocampus model as well as the unique pathways which appear in each of the models are detailed in Tables S3 and S4.

To better understand the significance of the 50 genes selected from the neuronal classification model, with respect to brain tissue-specific activity, we compared the expression pattern of these genes with that of other metabolic genes: First, we wanted to see whether the activity of these genes is higher in the brain as compared with other tissues. Because energy demands for brain activity are relatively high, it is indeed expected that we find that overall metabolic genes show, on average, a 1.34-fold increase in their expression levels in the brain as compared with other tissues (Fig. 4). Nevertheless, we found that the increase in expression in the normal brain of the 50 AD-predictive genes is significantly higher, with a 2-fold increase ($p$ value = $4.71 \times 10^{-5}$; Wilcoxon test; Fig. 4). Moreover, although these genes are expressed at higher levels across all tissues as compared with other metabolic genes ($p$ value = $6.83 \times 10^{-4}$ and $p$ value = $4.97 \times 10^{-5}$ for expression rate and breadth, respectively; see Methods and Table S5), their increase in expression in the brain is higher as compared with other metabolic genes, further strengthening their functional importance in the brain. Two additional tissues that showed high expression levels of these 50 metabolic genes were the liver and the heart, which also have high metabolic demands (2.74- and 2.15-fold increase with $p$ values of $1.25 \times 10^{-5}$ and $3.46 \times 10^{-7}$, respectively).

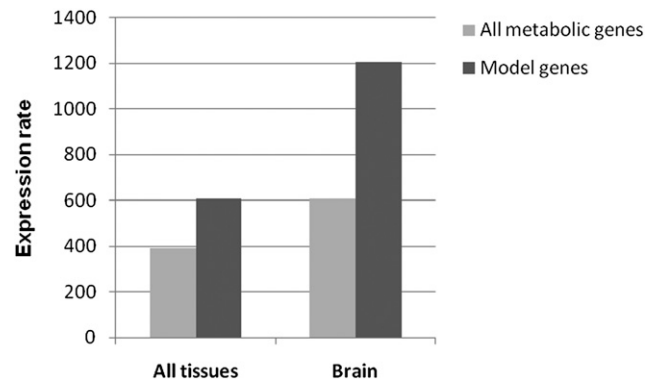


Fig. 4. Expression profile of all metabolic genes and the genes selected by the neuron classification model. For each group (metabolic genes and model genes), we calculated the mean expression in the brain and the mean expression rate across all 30 adult human tissues (Methods). Although highly expressed across all tissues, model genes show stronger increase in brain expression as compared with all metabolic genes, testifying to their importance in neuronal functioning.

Next, we set out to analyze the evolution patterns of the neuronal-selected genes in primates. It was already suggested that the evolution of human cognitive abilities was accompanied by changes in metabolism and that neurological disorders are a costly by-product of such evolution (Crow, 1995; Fu et al., 2011). Indeed, Khaitovich et al. (2008) showed that metabolic alterations found in schizophrenia were under positive selection in the human lineage. Hence, we set out to analyze the evolution of the expression of these genes as compared with that of other metabolic genes in primates. For that purpose, we compared the expression of metabolic genes in the human brain with those of two different primates: chimpanzees and rhesus monkeys (see Methods). Comparison between humans and chimpanzees did not reveal significant differences in expression of the selected metabolic genes ($p$ value = 0.187, Fig. S3A). Nevertheless, human–rhesus comparison revealed significant differences in expression of the selected genes as compared with other metabolic genes ($p$ value = $2.94 \times 10^{-3}$, Fig. 5A). Similar results were also obtained for chimpanzees and rhesus monkeys ($p$-value = $7.29 \times 10^{-3}$, Fig. S3B).

These high interspecies differences suggest that the expression of these genes has undergone positive selection since the last common ancestor of humans, chimpanzees, and rhesus monkeys. This may be a result of either functional specialization of these genes or evolution of the entire brain between the species. Previous studies have shown that AD-like neuropathologies increase during aging in most primates, and that chimpanzees show more deposits of $\beta$-amyloid as compared with rhesus monkeys (Finch and Sapolsky, 1999). This may hint to the difference in the results obtained in our study from those found previously in

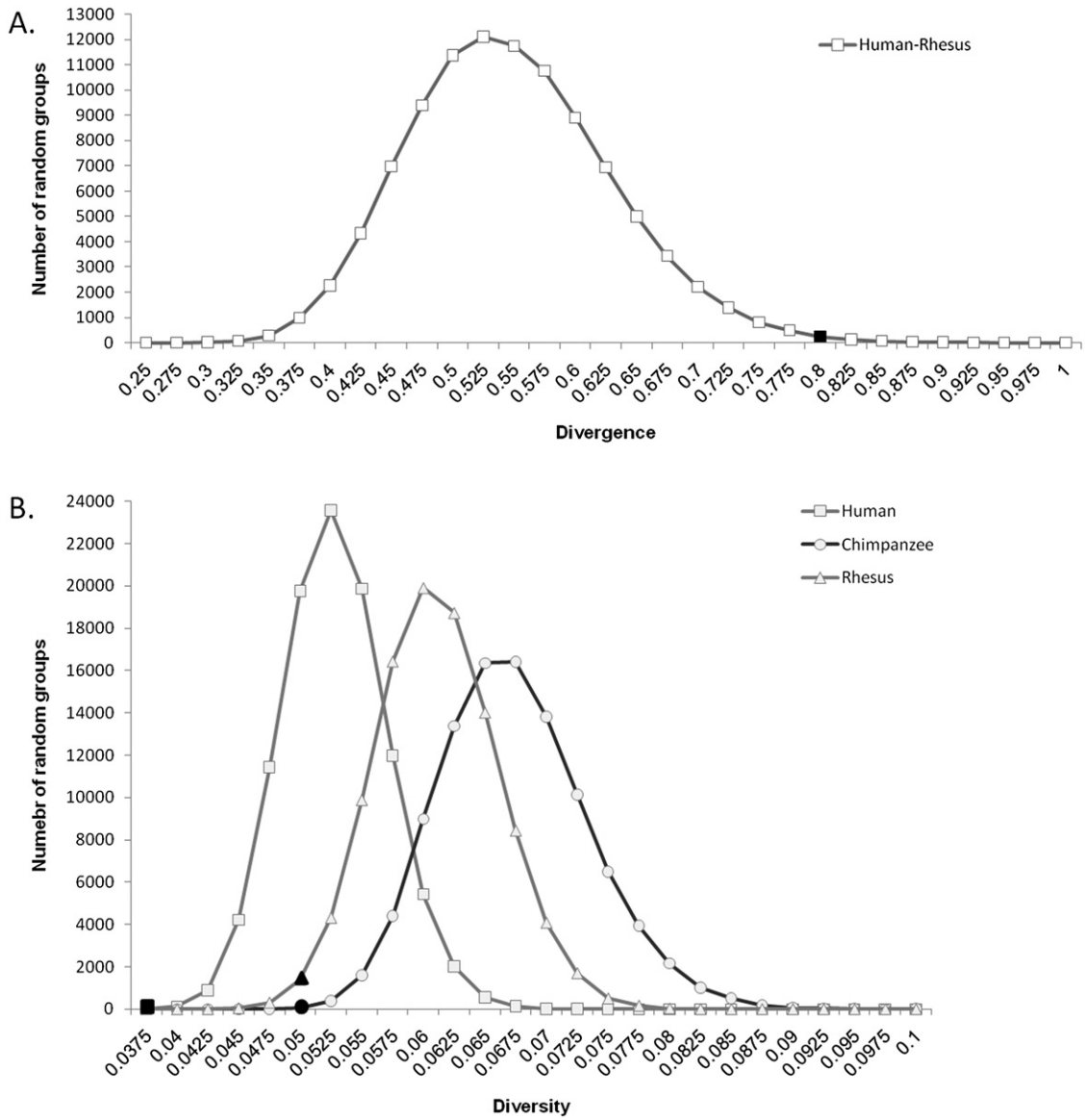


Fig. 5. Distribution of (A) divergence and (B) diversity of gene expression for randomly selected 100,000 random equal-size groups of metabolic genes (Methods). We compared these values with the mean divergence of the model-selected genes between humans and rhesus monkeys (black square in [A]) and the diversity in each of the three primates (black square, circle, and triangle in [B] for humans, chimpanzees, and rhesus monkeys, respectively). Model genes show significantly higher divergence levels in their expression between humans and rhesus monkeys, whereas they show lower diversity within each of the species.

a study on schizophrenia (Khaitovich et al., 2008), where significant differences were found between humans and chimpanzees in the latter but not in our study. These intriguing preliminary findings warrant a further and deeper exploration, which is beyond the scope of the current study.

This brain gene expression data also allowed us to look into the intraspecies variance of these genes (Fig. 5B). Interestingly, we find that AD-related genes show lower intraspecies variability (diversity) in all three primates as compared with other metabolic genes (empirical $p$ values of $1.0 \times 10^{-4}$, $6.9 \times 10^{-4}$, and $1.72 \times 10^{-2}$ for humans, chimpanzees, and rhesus monkeys, respectively). The low diversity suggests that these genes are under stronger regulatory regime in normal condition in all three primates, presumably owing to their functional importance. Nevertheless, because they were selected by the classification model, they show more variability between normal and AD cases (see Table S6). Hence, these findings give rise to the possibility that this regulatory regime is compromised as AD progresses.

Taken together, the analysis of the expression pattern of these neuronal genes across human tissues and between and within primates marks their importance in neuronal activity and hence suggests why alterations in the expression of

these genes are associated with AD. A parallel analysis of top-ranked genes in the whole tissue analysis reveals that, as for the neuronal model genes, the expression of these genes was significantly higher in the brain than the other metabolic genes (1.77-fold increase and 1.34-fold increase for the 50 genes and the other metabolic genes, respectively, $p$ value $= 2.35 \times 10^{-2}$; Wilcoxon test). However, the divergence and diversity expression variation analyses did not yield significant results for these genes.

## 4. Conclusions

The role of metabolism in the progression of AD is studied by inspecting the gene expression of metabolic genes. These genes are shown to be as effective in predicting the severity of AD as the entire gene list. Furthermore, the higher prediction accuracy obtained with neuronal expression per se (vs. whole tissue expression) points to the importance of metabolic processes in this specific cell type. In addition, metabolic whole tissue gene expression can predict the MMSE score better than conventional AD pathological markers that have been reported. Moreover, a selected group of metabolic genes that is chosen based on its AD-predictive ability is shown to be both associated with normal brain functioning and to have similar expression patterns in humans and chimpanzees and a different one in rhesus monkeys.

## Disclosure statement

All authors state that there are no conflicts of interest of any kind. There are no contracts relating to this research. There are no any agreements of the authors or their institutions that could be seen as involving a financial interest in this work.

The data contained in the manuscript being submitted have not been previously published, have not been submitted elsewhere, and will not be submitted elsewhere while under consideration at *Neurobiology of Aging*.

All authors have reviewed the contents of the manuscript being submitted, approve of its contents, and validate the accuracy of the data.

## Acknowledgements

The authors thank the members of the Ruppin laboratory and especially Ori Folger for helpful discussions. E.R. acknowledges the generous support provided by grants from the Israeli Science Foundation (ISF) to this research. S.S. gratefully acknowledges the support of the Josef Sagol Fellowship for brain research at Tel Aviv University. Y.Y.W. was supported in part by Eshkol (the Israeli Ministry of Science and Technology) and Dan David fellowships.

## Appendix. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.neurobiolaging.2012.04.003.

A list of the metabolic genes that were used in this study, lists of the top genes selected by the various models, hippocampus classification model, and other supplementary data indicated in the text.

## References

Barnham, K.J., Masters, C.L., Bush, A.I., 2004. Neurodegenerative diseases and oxidative stress. Nat. Rev. Drug Discov. 3, 205–214.

Bertram, L., McQueen, M.B., Mullin, K., Blacker, D., Tanzi, R.E., 2007. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat. Genet. 39, 17–23.

Blalock, E.M., Geddes, J.W., Chen, K.C., Porter, N.M., Markesbery, W.R., 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.

Blass, J.P., 2001. Brain metabolism and brain disease: is metabolic deficiency the proximate cause of Alzheimer dementia? J. Neurosci. Res. 66, 851–856.

Bonda, D.J., Wang, X., Perry, G., Nunomura, A., Tabaton, M., Zhu, X., Smith, M.A., 2010. Oxidative stress in Alzheimer disease: a possibility for prevention. Neuropharmacology 59, 290–294.

Bossers, K., Meerhoff, G., Balesar, R., van Dongen, J.W., Kruse, C.G., Swaab, D.F., Verhaagen, J., 2009. Analysis of gene expression in Parkinson's disease: possible involvement of neurotrophic support and axon guidance in dopaminergic cell death. Brain Pathol. 19, 91–107.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Butterfield, D.A., Hardas, S.S., Lange, M.L., 2010. Oxidatively modified glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and Alzheimer's disease: many pathways to neurodegeneration. J. Alzheimers Dis. 20, 369–393.

Chandrasekaran, K., Giordano, T., Brady, D.R., Stoll, J., Martin, L.J., Rapoport, S.I., 1994. Impairment in mitochondrial cytochrome oxidase gene expression in Alzheimer disease. Brain Res. Mol. Brain Res. 24, 336–340.

Clark, C.M., Sheppard, L., Fillenbaum, G.G., Galasko, D., Morris, J.C., Koss, E., Mohs, R., Heyman, A., 1999. Variability in annual Mini-Mental State Examination score in patients with probable Alzheimer disease: a clinical perspective of data from the Consortium to Establish a Registry for Alzheimer's Disease. Arch. Neurol. 56, 857–862.

Crow, T.J., 1995. A theory of the evolutionary origins of psychosis. Eur. Neuropsychopharmacol. 5 suppl, 59–63.

Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R., Palsson, B.O., 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc. Natl. Acad. Sci. U. S. A. 104, 1777–1782.

Ferrer, I., 2009. Altered mitochondria, energy metabolism, voltage-dependent anion channel, and lipid rafts converge to exhaust neurons in Alzheimer's disease. J. Bioenerg. Biomembr. 41, 425–431.

Finch, C.E., Sapolsky, R.M., 1999. The evolution of Alzheimer disease, the reproductive schedule, and apoE isoforms. Neurobiol. Aging 20, 407–428.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12, 189–198.

Fu, X., Giavalisco, P., Liu, X., Catchpole, G., Fu, N., Ning, Z.B., Guo, S., Yan, Z., Somel, M., Pääbo, S., Zeng, R., Willmitzer, L., Khaitovich, P., 2011. Rapid metabolic evolution in human prefrontal cortex. Proc. Natl. Acad. Sci. U. S. A. 108, 6181–6186.

Gibson, G.E., Ratan, R.R., Beal, M.F., 2008. Mitochondria and oxidative stress in neurodegenerative disorders. Preface. Ann. N. Y. Acad. Sci. 1147, xi–xii.

Ginsberg, S.D., Alldred, M.J., Che, S., 2012. Gene expression levels assessed by CA1 pyramidal neuron and regional hippocampal dissections in Alzheimer's disease. Neurobiol. Dis. 45, 99–107.

Hinkle, D.A., Eberwine, J.H., 2003. Single-cell molecular biology: implications for diagnosis and treatment of neurologic disease. Biol. Psychiatry 54, 413–417.

Ho, G.J., Hashimoto, M., Adame, A., Izu, M., Alford, M.F., Thal, L.J., Hansen, L.A., Masliah, E., 2005. Altered p59Fyn kinase expression accompanies disease progression in Alzheimer's disease: implications for its functional role. Neurobiol. Aging 26, 625–635.

Johnson, G.V., Bailey, C.D., 2002. Tau, where are we now? J. Alzheimers Dis. 4, 375–398.

Kamme, F., Salunga, R., Yu, J., Tran, D.T., Zhu, J., Luo, L., Bittner, A., Guo, H.Q., Miller, N., Wan, J., Erlander, M., 2003. Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. J. Neurosci. 23, 3607–3615.

Kennedy, B.P., Bottiglieri, T., Arning, E., Ziegler, M.G., Hansen, L.A., Masliah, E., 2004. Elevated S-adenosylhomocysteine in Alzheimer brain: influence on methyltransferases and cognitive function. J. Neural Transm. 111, 547–567.

Khaitovich, P., Lockstone, H.E., Wayland, M.T., Tsang, T.M., Jayatilaka, S.D., Guo, A.J., Zhou, J., Somel, M., Harris, L.W., Holmes, E., Pääbo, S., Bahn, S., 2008. Metabolic changes in schizophrenia and human brain evolution. Genome Biol. 9, R124.

Khaitovich, P., Tang, K., Franz, H., Kelso, J., Hellmann, I., Enard, W., Lachmann, M., Pääbo, S., 2006. Positive selection on gene expression in the human brain. Curr. Biol. 16, R356–R358.

Kim, J.M., Stewart, R., Kim, S.W., Yang, S.J., Shin, I.S., Shin, H.Y., Yoon, J.S., 2008. Methylenetetrahydrofolate reductase gene and risk of Alzheimer's disease in Koreans. Int. J. Geriatr. Psychiatry 23, 454–459.

Li, K., Liu, S., Yao, S., Wang, B., Dai, D., Yao, L., 2009. Interaction between interleukin-8 and methylenetetrahydrofolate reductase genes modulates Alzheimer's disease risk. Dement. Geriatr. Cogn. Disord. 27, 286–291.

Liang, W.S., Dunckley, T., Beach, T.G., Grover, A., Mastroeni, D., Ramsey, K., Caselli, R.J., Kukull, W.A., McKeel, D., Morris, J.C., Hulette, C.M., Schmechel, D., Reiman, E.M., Rogers, J., Stephan, D.A., 2010. Neuronal gene expression in non-demented individuals with intermediate Alzheimer's Disease neuropathology. Neurobiol. Aging 31, 549–566.

Liang, W.S., Reiman, E.M., Valla, J., Dunckley, T., Beach, T.G., Grover, A., Niedzielko, T.L., Schneider, L.E., Mastroeni, D., Caselli, R., Kukull, W., Morris, J.C., Hulette, C.M., Schmechel, D., Rogers, J., Stephan, D.A., 2008. Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. Proc. Natl. Acad. Sci. U. S. A. 105, 4441–4446.

Masters, C.L., Simms, G., Weinman, N.A., Multhaup, G., McDonald, B.L., Beyreuther, K., 1985. Amyloid plaque core protein in Alzheimer disease and Down syndrome. Proc. Natl. Acad. Sci. U. S. A. 82, 4245–4249.

Mattson, M.P., 2004. Pathways towards and away from Alzheimer's disease. Nature 430, 631–639.

Mihaescu, R., Detmar, S.B., Cornel, M.C., van der Flier, W.M., Heutink, P., Hol, E.M., Rikkert, M.G., van Duijn, C.M., Janssens, A.C., 2010. Translational research in genomics of Alzheimer's disease: a review of current practice and future perspectives. J. Alzheimers Dis. 20, 967–980.

Parkinson, H., Sarkans, U., Kolesnikov, N., Abeygunawardena, N., Burdett, T., Dylag, M., Emam, I., Farne, A., Hastings, E., Holloway, E., Kurbatova, N., Lukk, M., Malone, J., Mani, R., Pilicheva, E., Rustici, G., Sharma, A., Williams, E., Adamusiak, T., Brandizi, M., Sklyar, N., Brazma, A., 2011. ArrayExpress update—an archive of microarray and high-throughput sequencing-based functional genomics experiments. Nucleic Acids Res. 39 (Database issue), D1002–D1004.

Radmacher, M.D., McShane, L.M., Simon, R., 2002. A paradigm for class prediction using gene expression profiles. J. Comput. Biol. 9, 505–511.

Ray, M., Zhang, W., 2009. Integrating gene expression and phenotypic information to analyze Alzheimer's disease. J. Alzheimers Dis. 16, 73–84.

Reed, T., Perluigi, M., Sultana, R., Pierce, W.M., Klein, J.B., Turner, D.M., Coccia, R., Markesbery, W.R., Butterfield, D.A., 2008. Redox proteomic identification of 4-hydroxy-2-nonenal-modified brain proteins in amnestic mild cognitive impairment: insight into the role of lipid peroxidation in the progression and pathogenesis of Alzheimer's disease. Neurobiol. Dis. 30, 107–120.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M.P., Walker, J.R., Hogenesch, J.B., 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. U. S. A. 101, 6062–6067.

Su, B., Wang, X., Zheng, L., Perry, G., Smith, M.A., Zhu, X., 2010. Abnormal mitochondrial dynamics and neurodegenerative diseases. Biochim. Biophys. Acta 1802, 135–142.

Sultana, R., Boyd-Kimball, D., Poon, H.F., Cai, J., Pierce, W.M., Klein, J.B., Merchant, M., Markesbery, W.R., Butterfield, D.A., 2006. Redox proteomics identification of oxidized proteins in Alzheimer's disease hippocampus and cerebellum: an approach to understand pathological and biochemical alterations in AD. Neurobiol. Aging 27, 1564–1576.

Terry, R.D., Gonatas, N.K., Weiss, M., 1964. Ultrastructural studies in Alzheimer's presenile dementia. Am. J. Pathol. 44, 269–297.

Vernace, V.A., Schmidt-Glenewinkel, T., Figueiredo-Pereira, M.E., 2007. Aging and regulated protein degradation: who has the UPPer hand? Aging Cell 6, 599–606.

Waldman, Y.Y., Tuller, T., Shlomi, T., Sharan, R., Ruppin, E., 2010. Translation efficiency in humans: tissue specificity, global optimization and differences between developmental stages. Nucleic Acids Res. 38, 2964–2974.

Wang, B., Jin, F., Kan, R., Ji, S., Zhang, C., Lu, Z., Zheng, C., Yang, Z., Wang, L., 2005. Association of MTHFR gene polymorphism C677T with susceptibility to late-onset Alzheimer's disease. J. Mol. Neurosci. 27, 23–27.

Wang, X., Su, B., Zheng, L., Perry, G., Smith, M.A., Zhu, X., 2009. The role of abnormal mitochondrial dynamics in the pathogenesis of Alzheimer's disease. J. Neurochem. 109 suppl 1, 153–159.

Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., McCorquodale, D.S., 3rd, Cuello, C., Leung, D., Bryden, L., Nath, P., Zismann, V.L., Joshipura, K., Huentelman, M.J., Hu-Lince, D., Coon, K.D., Craig, D.W., Pearson, J.V., Heward, C.B., Reiman, E.M., Stephan, D., Hardy, J., Myers, A.J., 2009. Genetic control of human brain transcript expression in Alzheimer disease. Am. J. Hum. Genet. 84, 445–458.

Zhu, X., Perry, G., Moreira, P.I., Aliev, G., Cash, A.D., Hirai, K., Smith, M.A., 2006. Mitochondrial abnormalities and oxidative imbalance in Alzheimer disease. J. Alzheimers Dis. 9, 147–153.