A Blood Gene Expression Marker of Early Alzheimer’s Disease

Katie Lunnona, Martina Sattleckerb,1, Simon J. Furneya,1,2, Giovanni Coppolaa,
Andrew Simmonsb, Petraela Proitsia, Michelle K. Luptonb, Anbarasu Lourdusamya,b,
Caroline Johnstonb, Hilka Soininenb, Iwona Klozkewskic, Patrizia Mezzocci,
Magda Tsoikaf, Bruno Vella, Daniel Geschwind, Simon Lovestonea,b, Richard Dobsona,b,
Angela Hodgesa,∗ and On behalf of the AddNeuroMed Consortium

a King’s College London, Institute of Psychiatry, King’s Health Partners Centre for Neurodegeneration Research, London, UK
b NIHR Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, King’s College London, London, UK
c Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California at Los Angeles, CA, USA
d Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland
e Medical University of Lodz, Lodz, Poland
f Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy
g 3rd Department of Neurology, “G. Papanicolaou” Hospital, Aristotle University of Thessaloniki, Thessaloniki, Greece
h INSERM U 558, University of Toulouse, Toulouse, France

Accepted: 21 August 2012

Abstract. A marker of Alzheimer’s disease (AD) that can accurately diagnose disease at the earliest stage would significantly support efforts to develop treatments for early intervention. We have sought to determine the sensitivity and specificity of peripheral blood gene expression as a diagnostic marker of AD using data generated on HT-12v3 BeadChips. We first developed an AD diagnostic classifier in a training cohort of 78 AD and 78 control blood samples and then tested its performance in a validation group of 26 AD and 26 control and 118 mild cognitive impairment (MCI) subjects who were likely to have an AD-endpoint. A 48 gene classifier achieved an accuracy of 75% in the AD and control validation group. Comparisons were made with a classifier developed using structural MRI measures, where both measures were available in the same individuals. In AD and control subjects, the gene expression classifier achieved an accuracy of 70% compared to 85% using MRI. Bootstrapping validation produced expression and MRI classifiers with mean accuracies of 76% and 82%, respectively, demonstrating better concordance between these two classifiers than achieved in a single validation population. We conclude there is potential for blood expression to be a marker for AD. The classifier also predicts a large number of people with MCI, who are likely to develop AD, are more AD-like than normal with 76% of subjects classified as AD rather than control. Many of these people do not have

∗Correspondence to: Angela Hodges, King’s College London, De Crespigny Park, London, UK. Tel.: +44 207 848 0772; Fax: +44 207 848 0632; E-mail: angela.k.hodges@kcl.ac.uk.
INTRODUCTION

Alzheimer’s disease (AD) is a common chronic neurodegenerative disorder, accounting for ~60% of dementia cases. Dementia affects 63 million patients worldwide with numbers set to rise to 114 million by 2050 resulting in dramatic social and economic consequences as our care systems struggle to cope [1]. Although much progress has been made in understanding the molecular pathology of AD, the treatments currently available only temporarily alleviate some symptoms and do not modify pathology. The use of biomarkers for identification of individuals with AD prior to the appearance of clinical symptoms, the so-called pre-dementia phase of the disease [2], will be essential to the development of drugs for early intervention [3–5]. Furthermore, if sufficiently powered, some biomarkers could be used as part of a screening program for at-risk elderly people [6].

By the time an individual is diagnosed with AD, there is already considerable neuronal cell loss, plaque deposition, and neurofibrillary tangles within the brain [7, 8], which may have emerged up to 10 years or more before clinical diagnosis [9–11]. Biomarkers linked to the pathophysiological process in AD can greatly increase the confidence of concluding a person will have an AD-endpoint even in the pre-dementia phase [12]. Cerebrospinal fluid (CSF) biomarkers such as increased tau and decreased levels of amyloid-β (Aβ)1-42 in the CSF correlate with postmortem AD pathology [13, 14]. These measurements together not only differentiate AD from normal elderly controls with high accuracy, but can also predict which subjects with mild cognitive impairment (MCI) are likely to progress to AD within 5 years [15]. However, a lumbar puncture to collect CSF is an invasive procedure, which may not be suitable for use in large-scale trials or for screening populations. Similarly, positron emission tomography (PET) imaging of amyloid burden in the brain correlates with clinical diagnosis of AD, Aβ neuropathology at autopsy [13, 14, 16], and CSF Aβ1-42 levels [17–20]. PET imaging is expensive and impractical to be used in large groups of frail elderly people and is restricted to specialist centers. Our aim was to first identify and validate an AD diagnostic gene expression marker in blood able to distinguish people clinically diagnosed with AD from normal elderly controls with high sensitivity and specificity. We then sought to evaluate its specificity and sensitivity compared to another marker which...
incorporates structural magnetic resonance imaging (MRI) measures of regional brain atrophy and cortical thickness assessed in the same individuals, using a hypothesis-free approach. To date, the outcome of comparing a gene expression marker of AD alongside a classifier using measures of brain atrophy has not yet been reported. Classifiers combining measures of cortical thickness and/or regional brain atrophy changes in AD have previously been shown to very accurately distinguish AD from normal elderly controls [34]. Structural brain changes can also be used to identify people with MCI who will subsequently receive a diagnosis of AD or experience a more rapid drop in Mini-Mental Status Examination (MMSE) score over time [35, 36]. Structural brain changes generally appear after changes in other modalities such as CSF Aβ and tau and coincide with neuronal cell loss and the onset of clinical symptoms [11, 37].

Individuals with MCI are likely to have an AD-like rather than normal using an appropriate biomarker. The proportion of people with MCI having prodromal AD varies between studies as it depends on the method of ascertainment, the criteria for identifying people with MCI, the period of follow-up, and the confirmatory methods for subsequent AD diagnosis. Typical estimates suggest around ∼40–75% of people who have cognitive symptoms consistent with MCI will eventually progress to pathologically confirmed AD [38–40], with only ∼5% “reverting” to normal [41, 42] and ∼30–55% developing other dementias [43–45]. We were therefore interested to investigate whether a set of blood gene expression changes found in AD patients also exist in a MCI population and whether a classifier combining measures of brain atrophy has not yet been reported. Classifiers combining measures of AD [38–40], with only ∼5% “reverting” to normal [41, 42] and ∼30–55% developing other dementias [43–45]. We were therefore interested to investigate whether a set of blood gene expression changes found in AD patients also exist in a MCI population and whether or not there was any distinction between MCI subjects close to receiving a clinical diagnosis of AD and those who remained MCI within a subsequent two year follow-up period. While our AD diagnostic classifier was adequately powered, we did not have sufficient power to build a separate predictive classifier within our MCI population to determine which subjects would develop AD within two years.

METHODS AND MATERIALS

Clinical subjects and samples

Subjects used were from AddNeuroMed, a large cross-European prospective biomarker study (2005–2007) [46, 47]. Informed consent was obtained for all subjects according to the Declaration of Helsinki (1991) and protocols and procedures were approved by the relevant local ethical committees at each site. Subjects were excluded from the study if they were younger than 65 years, had significant neurological or psychiatric illness other than AD, significant unstable systematic illness or organ failure, or a geriatric depression rating scale score ≥45 [48].

Normal elderly control subjects were recruited from non-related family members of AD patients, caregivers’ relatives, social centers for the elderly, or GP surgeries and had no evidence of cognitive impairment. A small number of individuals recruited in this way met the MCI criteria detailed below and were therefore included in the MCI cohort for analyses. AD and MCI subjects were recruited primarily from local memory clinics, and as such the MCI cohort was expected to be composed largely of subjects with a likely AD-endpoint. AD subjects were diagnosed using the National Institute of Neurological and Communicative Disease and Stroke and Alzheimer’s disease (NINCDS-ADRDA) [49] and Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [50] criteria for possible or probable AD. All MCI subjects reported problems with memory, corroborated by an informant, but had normal activities of daily living as specified in the Petersen’s criteria for amnestic MCI [51, 52]. MCI subjects scored 0.5 on the total Clinical Dementia Rating Scale or had a memory score of 0.5 or 1 [53]. Within 2 years of the baseline visit, some MCI subjects progressed to a clinical diagnosis of AD (MCIc), while others remained MCI (MCInc). Diagnoses were made by trained researchers following a previously validated protocol [54].

All subjects underwent a structured interview and a battery of neuropsychological assessments including the Mini Mental State Examination (MMSE) [55], Global Deterioration Scale (GDS) [56], and Clinical Dementia Rating Scale (CDR) [53] by trained researchers. Control and MCI subjects were further assessed using the CERAD battery [57]. Detailed information on subject recruitment and assessments can be found in other published studies [46, 47]. Where possible, whole blood samples were collected from each subject for DNA (APOE genotyping) and RNA analyses (gene expression). Consecutive subjects who met MRI inclusion criteria (no claustrophobia, no trauma or surgery which may have left ferromagnetic material in the body, ferromagnetic implants or pacemakers, and the ability to lie still for at least one hour) were invited to undertake an MRI scan until a total of ∼20 subjects per diagnostic group per center were scanned.
Whole blood RNA extraction

Venous blood (2.5 ml) was collected into a PAXgene™ Blood RNA tube (BD) for RNA analysis from subjects who had fasted 2 hours prior to collection. The sample was frozen at −20 °C overnight and then transferred to −80 °C for long-term storage. Prior to RNA extraction, samples were thawed at room temperature overnight. RNA was extracted using the PAXgene™ Blood RNA Kit (Qiagen), according to the manufacturer’s protocol. The yield and quality of extracted RNA were assessed using the NanoDrop™ 1000 spectrophotometer (NanoDrop Technologies) and the Agilent 2100 Bioanalyzer (Agilent Technologies) respectively. Only samples with an RNA Integrity Number ≥ 7.0 were analyzed in this study.

Microarray analysis

Illumina Human HT-12 v3 Expression BeadChips (Illumina) were used to analyze whole genome transcript expression according to the protocol supplied by the manufacturer in 356 subjects (116 control, 127 MCI, and 113 AD). Samples from subjects who had also undergone an MRI scan or who had subsequently changed their diagnosis from MCI to AD within 2 years of the sample being analyzed were prioritized for inclusion. The remaining control, MCI, and AD subjects were randomly chosen from the cohort available at the time of analysis to provide roughly equal numbers across the three groups with the final number chosen for practical reasons. The beadchips contained 48,803 probes designed using data from RefSeq (Build 36.2, Rel 22) and the UniGene (Build 199) databases. Briefly, the TotalPrep RNA Amplification Kit (Ambion) was used to synthesize cDNA from 200 ng total RNA followed by amplification and biotinylation of cRNA and hybridization. Following hybridization, gene expression values were variance-stabilization transformed and quantile normalized using the R Bioconductor package lumi [58]. A total of 30 chips were excluded from further analysis for a number of reasons including very low BeadChip detection rate, disparity in XIST gene expression gender calling (Illumina probe ID: ILMN_1764573) with recorded gender and/or gender markers derived from SNP chip data from the same individuals or diagnostic re-classification at subsequent visits prior to final data analysis (other than MCI-AD). This left a total of 326 subjects (104 AD, 118 MCI (77 MCIc, 41 MCIa), and 104 normal elderly control subjects) which were used for classifier development and validation.

Probes that were not expressed in any samples were removed from further analysis, leaving 38,311 probes for classification purposes. Data were adjusted for the independent variables of age, gender, collection site, and RNA quality (RIN).

Neuroimaging

Highly automated bilateral regional cortical thickness measures from 34 areas and regional brain volume measures from 41 areas totaling 109 measures were obtained from subjects who underwent MRI (Supplementary Table 1). Detailed information about data acquisition, pre-processing, and quality control assessment have been described for this cohort in detail elsewhere [34, 36, 47, 59, 60]. Data from 91 AD, 92 MCI, and 90 control subjects were included in the analyses.

Diagnostic classifier development

Using the clinical diagnosis of AD as the diagnostic standard, we developed a classifier containing those measures of blood gene expression able to achieve the greatest sensitivity and specificity in a training cohort. We based our methods on those recommended by the MicroArray Quality Control II (MAQC-II) study [61]. Samples from AD and normal elderly control subjects were randomly divided into a 75% training cohort (78 AD, 78 Control; Table 1) and a 25% independent validation cohort (26 AD, 26 Control; Table 1). Initially a t-test was performed in the training cohort to reduce the number of gene expression features, leaving a total of 203 probes with p-value <0.01 (False discovery rate (FDR)-corrected). These 203 probes were then used to optimize the parameters of a random forest model as the Random Jungle [62] implementation requires. In the optimization procedure, various combinations of the random forest parameters including mTry (number of randomly selected variables) and nTrees (number of trees) were used. Additionally, we applied the Meng score as a measure of variable importance [63], backward elimination, and a conditional importance setting of 0.8, which was kept constant throughout the optimization procedure. The best training result was achieved by a setting of nTrees = 750 and mTry = 15, with 50 probes included in the model. These settings and the 50 highest ranked probes were then used to further optimize the random forest parameters, 50 highest ranked probes, and all samples in the training cohort were then used.
Table 1
Subject characteristics of the out of bag (training) and validation (test) cohorts used in the development of an AD-Control classifier

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Out of bag cohort (Training)</th>
<th>Validation cohort (Test)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AD</td>
<td>Control</td>
</tr>
<tr>
<td>Samples used in analysis</td>
<td>78</td>
<td>76</td>
</tr>
<tr>
<td>Gender (F/M)</td>
<td>45/33</td>
<td>52/26</td>
</tr>
<tr>
<td>No. with MRI measures</td>
<td>67</td>
<td>68</td>
</tr>
<tr>
<td>Age in years (±SD)</td>
<td>72.1(6.2)</td>
<td>75.5(6.9)</td>
</tr>
<tr>
<td>Disease duration in years (±SD)</td>
<td>N/A</td>
<td>3.7(2.3)</td>
</tr>
<tr>
<td>MMSE (±SD)</td>
<td>29.0(1.2)</td>
<td>21.1(4.4)</td>
</tr>
<tr>
<td>ADAS-Cog (±SD)</td>
<td>N/A</td>
<td>23.6(9.4)</td>
</tr>
<tr>
<td>CDR sum of boxes (±SD)</td>
<td>0.1(0.2)</td>
<td>6.4(3.0)</td>
</tr>
</tbody>
</table>

To build a final random forest model, which was then tested with the independent validation cohort.

In order to allow us to directly compare the predictive power of gene expression with sMRI, we re-developed the expression classifier to include only those samples for which both imaging and gene expression data were available. AD and normal elderly control subjects were divided into a 75% training cohort (68 AD, 67 Control; Table 1) and a 25% independent validation cohort (23 AD, 26 Control; Table 1), maintaining sample groups as before and removing those where only expression data were available in order to achieve a similar sample distribution across training and test cohorts. The reduced training cohort was then used to develop three additional diagnostic models; the first included only sMRI imaging measures, the second included sMRI and gene expression, and the third only gene expression in these samples. For all three models, MCI samples were included with the validation cohort. For imaging classifier development, the 109 structural brain measures derived from MRI were used to optimize the parameters of a random forest model as before with the 20 highest ranked measures taken further into the final model. For the combined sMRI and gene expression model, 30 imaging and 45 gene expression measures were in the model and for the gene expression model using samples with available sMRI measures, there were 50 measures. The sample workflow, analysis, and model development scheme for classification is illustrated in Fig. 1.

All four diagnostic models were further validated by bootstrapping. First the data set was randomly split into a bootstrap training-set (75% of samples) and a bootstrap test-set (25% of samples). Each bootstrap training-set was used to build a random forest model, which was subsequently tested in the bootstrap test-set. Thus overall, 200 bootstrap random forest models were generated and tested. The bootstrapping procedure generated a list summarizing and ranking probes based on their importance and predictive power across the 200 bootstrap models.

Using the R package MVpower (http://cran.r-project.org/web/packages/MVpower/MVpower.pdf) in the above approach, we estimate that for effect sizes of 0.7 and 0.2 we have power of 100 and 80% respectively to develop a Random Forest classifier with 203 features and 50 probes based on sample sizes of 78 in each group. Our effect size of 0.7 was based on differences between cases and controls in the selected 203 features, 0.2 was chosen as a conservative lower limit.

**RESULTS**

**Performance of diagnostic classifier in distinguishing AD from normal elderly control subjects**

We sought to identify and evaluate a blood gene expression diagnostic classifier of AD that could distinguish people with AD from normal elderly control subjects. We found 48 genes, represented by 50 illumina probes selected through backward elimination of 203 probes identified by t-tests (Supplementary Table 2; available online: http://www.j-alz.com/issues/33vol33-3.html#supplementarydata03), and achieved the highest “out of bag” accuracy in our training cohort. The 50 probes along with their Meng score of relative importance within the classifier are listed in Table 2. When this AD-Control classifier was applied to the independent validation group of samples, it was able to correctly classify 39/52 subjects (75.0% accuracy), of which 21/26 AD subjects were correctly classified (80.8% sensitivity) and 18/26 normal elderly control subjects were correctly excluded (69.2% specificity) (Table 3A), achieving a positive predictive value (PPV) of 72.4% and negative
1. Sample workflow:

AddNeuroMed Participants:

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>MCI</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>206</td>
<td>153</td>
<td>158</td>
</tr>
</tbody>
</table>

AddNeuroMed samples selected for gene expression:

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>MCI</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>116</td>
<td>127</td>
<td>113</td>
</tr>
</tbody>
</table>

Gene expression samples remaining after quality control:

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>MCI</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>sMRI measures</td>
<td>116</td>
<td>127</td>
<td>113</td>
</tr>
</tbody>
</table>

2. Use of training cohort for classifier development:

Classifiers:
- All expression data (all samples)
- sMRI data
- Expression and sMRI data
- Expression data (only samples with available sMRI measures)

Feature reduction:
A. 203 probes remain after t-test (p<0.01), adjusted
B. No feature reduction for sMRI measures
C. Probes as in A, sMRI measures as in B
D. Probes as in A

Random Forest model optimization (same for all classifiers):
- ntree parameter
- Number of trees
- Backwards elimination
- Mean score for variable importance estimation

Result:
Final model selection based on out of bag error
(for all four classifiers)

3. Final validation of classifiers:

Test of final model with validation cohort:
- All expression data (all samples)
- sMRI data
- Expression and sMRI data
- Expression data (only samples with available sMRI measures)

Result:
Accuracy for all classifier
Cohen's kappa coefficient for inter-rater agreement

4. Bootstrapping:

Repeat the following steps 200 times:
- Sampling bootstrap training cohort
- Build model
- Test models with bootstrap test set

Result:
200 classifier accuracies reflecting variation in population

---

Fig. 1: Sample flow, model development, and model validation scheme for classification. Samples with a minimum RNA integrity number of 7 were selected from AddNeuroMed participants, with priority given to those who had undergone MRI scans and MCI subjects known to have converted from MCI to AD. The remaining control, MCI, and AD subjects were randomly chosen from available samples to provide approximately equal numbers across the three groups. When generating the AD-Control classifier, AD and Control samples were randomly categorized as training (out of bag) (75%) or validation (25%) from samples passing quality control. Samples from all MCI subjects were examined as part of the validation cohort to assess how they were categorized relative to AD and normal elderly control subjects. Each classifier was developed in the training cohort and then tested in the validation cohorts to generate a final probe (gene) or imaging measure list along with a measure of accuracy, sensitivity, and specificity. Finally, data underwent bootstrapping, whereby in each bootstrap iteration (total of 200) a new development and validation set was drawn from the complete data set used for each model. In each round of bootstrapping, the training set was used to build a classification model, which was then assessed with the validation set. This method gives a range of classifier accuracies, reflecting variation in the population, and allowing an additional assessment to the traditional independent test set approach.
### Table 2

The 50 Illumina probes (48 genes) that best classify AD from control samples in the training cohort.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Illumina probe</th>
<th>Meng scorea</th>
<th>q-valueb</th>
<th>Bootstrap countc</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPS27A</td>
<td>IMRG.2088326</td>
<td>1.67E-02</td>
<td>2.11E-05</td>
<td>200</td>
</tr>
<tr>
<td>CHMP4A</td>
<td>IMRG.715807</td>
<td>1.71E-02</td>
<td>4.10E-03</td>
<td>200</td>
</tr>
<tr>
<td>SFRS17A</td>
<td>IMRG.207737</td>
<td>1.68E-02</td>
<td>9.76E-03</td>
<td>198</td>
</tr>
<tr>
<td>POMP</td>
<td>IMRG.609328</td>
<td>1.67E-02</td>
<td>3.62E-05</td>
<td>147</td>
</tr>
<tr>
<td>C5ORF41</td>
<td>IMRG.205821</td>
<td>1.66E-02</td>
<td>9.41E-05</td>
<td>142</td>
</tr>
<tr>
<td>FTHL7</td>
<td>IMRG.203616</td>
<td>1.61E-02</td>
<td>3.28E-03</td>
<td>67</td>
</tr>
<tr>
<td>LOC401206</td>
<td>IMRG.792538</td>
<td>1.59E-02</td>
<td>3.62E-05</td>
<td>115</td>
</tr>
<tr>
<td>AK1</td>
<td>IMRG.201653</td>
<td>1.58E-02</td>
<td>2.34E-05</td>
<td>161</td>
</tr>
<tr>
<td>UQCRB</td>
<td>IMRG.759483</td>
<td>1.55E-02</td>
<td>3.64E-03</td>
<td>200</td>
</tr>
<tr>
<td>LOC633505</td>
<td>IMRG.778620</td>
<td>1.55E-02</td>
<td>4.93E-05</td>
<td>134</td>
</tr>
<tr>
<td>PDG</td>
<td>IMRG.2075051</td>
<td>1.51E-02</td>
<td>2.34E-03</td>
<td>58</td>
</tr>
<tr>
<td>NRPB2</td>
<td>IMRG.733348</td>
<td>1.50E-02</td>
<td>1.85E-03</td>
<td>197</td>
</tr>
<tr>
<td>UTF14A</td>
<td>IMRG.2095820</td>
<td>1.47E-02</td>
<td>2.34E-03</td>
<td>57</td>
</tr>
<tr>
<td>Dicer1</td>
<td>IMRG.772692</td>
<td>1.46E-02</td>
<td>4.32E-04</td>
<td>180</td>
</tr>
<tr>
<td>RPS36A</td>
<td>IMRG.2149998</td>
<td>1.44E-02</td>
<td>3.62E-05</td>
<td>200</td>
</tr>
<tr>
<td>ATP6E2</td>
<td>IMRG.2255847</td>
<td>1.42E-02</td>
<td>5.76E-03</td>
<td>145</td>
</tr>
<tr>
<td>SRPG</td>
<td>IMRG.771801</td>
<td>1.39E-02</td>
<td>2.34E-03</td>
<td>20</td>
</tr>
<tr>
<td>LMS3</td>
<td>IMRG.2229242</td>
<td>1.36E-02</td>
<td>8.15E-03</td>
<td>200</td>
</tr>
<tr>
<td>AFB3</td>
<td>IMRG.2320513</td>
<td>1.32E-02</td>
<td>2.49E-03</td>
<td>177</td>
</tr>
<tr>
<td>MKPL5a</td>
<td>IMRG.2097421</td>
<td>1.32E-02</td>
<td>2.19E-07</td>
<td>200</td>
</tr>
<tr>
<td>REL1.2</td>
<td>IMRG.603340</td>
<td>1.30E-02</td>
<td>5.08E-03</td>
<td>187</td>
</tr>
<tr>
<td>CETN2</td>
<td>IMRG.4695445</td>
<td>1.29E-02</td>
<td>2.11E-05</td>
<td>195</td>
</tr>
<tr>
<td>PWP1</td>
<td>IMRG.7183049</td>
<td>1.11E-02</td>
<td>5.07E-03</td>
<td>81</td>
</tr>
<tr>
<td>KIAA0146</td>
<td>IMRG.20481774</td>
<td>1.11E-02</td>
<td>7.55E-04</td>
<td>100</td>
</tr>
<tr>
<td>LOC388720</td>
<td>IMRG.704990</td>
<td>1.11E-02</td>
<td>2.12E-04</td>
<td>101</td>
</tr>
<tr>
<td>CIP29</td>
<td>IMRG.608907</td>
<td>1.11E-02</td>
<td>5.41E-03</td>
<td>191</td>
</tr>
<tr>
<td>LOC388621</td>
<td>IMRG.877262</td>
<td>1.11E-02</td>
<td>4.75E-03</td>
<td>139</td>
</tr>
<tr>
<td>SFRS17A</td>
<td>IMRG.311716</td>
<td>1.09E-02</td>
<td>4.87E-04</td>
<td>182</td>
</tr>
<tr>
<td>PCBP1</td>
<td>IMRG.607331</td>
<td>1.09E-02</td>
<td>4.23E-04</td>
<td>100</td>
</tr>
<tr>
<td>ATP5D1</td>
<td>IMRG.220862</td>
<td>1.08E-02</td>
<td>3.09E-05</td>
<td>65</td>
</tr>
<tr>
<td>H2AFY</td>
<td>IMRG.2053495</td>
<td>1.07E-02</td>
<td>6.30E-03</td>
<td>163</td>
</tr>
<tr>
<td>COX7</td>
<td>IMRG.214771</td>
<td>1.07E-02</td>
<td>2.18E-05</td>
<td>5</td>
</tr>
<tr>
<td>ID3</td>
<td>IMRG.793662</td>
<td>1.07E-02</td>
<td>2.34E-03</td>
<td>162</td>
</tr>
<tr>
<td>SHFM1</td>
<td>IMRG.221828</td>
<td>1.07E-02</td>
<td>2.80E-04</td>
<td>173</td>
</tr>
<tr>
<td>LOC525164</td>
<td>IMRG.712417</td>
<td>1.01E-02</td>
<td>6.06E-03</td>
<td>53</td>
</tr>
<tr>
<td>STX16</td>
<td>IMRG.204912</td>
<td>9.95E-03</td>
<td>1.91E-03</td>
<td>160</td>
</tr>
<tr>
<td>GDPD1</td>
<td>IMRG.204635</td>
<td>9.93E-03</td>
<td>4.88E-03</td>
<td>179</td>
</tr>
<tr>
<td>CSORF166</td>
<td>IMRG.214848</td>
<td>9.85E-03</td>
<td>3.13E-03</td>
<td>27</td>
</tr>
<tr>
<td>ING3</td>
<td>IMRG.227374</td>
<td>9.63E-03</td>
<td>5.42E-03</td>
<td>158</td>
</tr>
<tr>
<td>CACNA2D4</td>
<td>IMRG.406631</td>
<td>9.61E-03</td>
<td>7.73E-03</td>
<td>47</td>
</tr>
<tr>
<td>PLEK</td>
<td>IMRG.795762</td>
<td>9.22E-03</td>
<td>2.84E-03</td>
<td>3</td>
</tr>
<tr>
<td>NUDFA1</td>
<td>IMRG.224326</td>
<td>9.04E-03</td>
<td>2.45E-06</td>
<td>198</td>
</tr>
<tr>
<td>CDK1B</td>
<td>IMRG.722811</td>
<td>9.12E-03</td>
<td>7.93E-03</td>
<td>199</td>
</tr>
<tr>
<td>RGS19</td>
<td>IMRG.767058</td>
<td>7.60E-03</td>
<td>5.40E-03</td>
<td>165</td>
</tr>
<tr>
<td>RPS27A</td>
<td>IMRG.755883</td>
<td>7.61E-03</td>
<td>2.77E-04</td>
<td>180</td>
</tr>
<tr>
<td>LOC731640</td>
<td>IMRG.666174</td>
<td>7.41E-03</td>
<td>8.33E-03</td>
<td>66</td>
</tr>
<tr>
<td>AHS1A</td>
<td>IMRG.203637</td>
<td>6.98E-03</td>
<td>1.74E-03</td>
<td>56</td>
</tr>
<tr>
<td>BXC1</td>
<td>IMRG.646187</td>
<td>6.43E-03</td>
<td>4.29E-04</td>
<td>0</td>
</tr>
<tr>
<td>KARS</td>
<td>IMRG.777587</td>
<td>5.54E-03</td>
<td>2.56E-03</td>
<td>5</td>
</tr>
<tr>
<td>UBE2G1</td>
<td>IMRG.481448</td>
<td>-4.62E-04</td>
<td>1.74E-03</td>
<td>105</td>
</tr>
</tbody>
</table>

*aAn estimate of variable (probe) importance in the AD versus Control classifier Random Forest model. b*p*-value from a *t*-test comparing AD and control samples. cThe number of times each probe from the out of bag classifier appeared in the 200 bootstrap iterations.

*Predictive value (NPV) of 78.3%. As APOEε4 dosage is a known risk factor for developing AD, we repeated the model creation step including the APOEε4 genotype in the list of attributes. During backward elimination, we found this attribute was removed in the early rounds of the iterative process and therefore excluded it from further analysis. As the accuracy of a classifier varies depending on the population sampled, we bootstrapped the sampled population 200 times to assess the accuracy of the 203 probes identified.
Table 3

Classifiers’ performance in out of bag and validation data sets following Random Forests modeling, implemented in Random Jungle

(A) Performance of AD-Control classifier in AD and Control samples

<table>
<thead>
<tr>
<th>Classifier</th>
<th>AD–Control training cohort (expression)</th>
<th>AD–Control validation cohort (expression + imaging)</th>
<th>AD–Control validation cohort (expression)</th>
<th>AD–Control validation cohort (expression + imaging)</th>
<th>AD–Control validation cohort (expression)</th>
<th>AD–Control validation cohort (expression + imaging)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD–Control training cohort (expression)</td>
<td>128/156 (% correct classified)</td>
<td>68/78 (% correct classified)</td>
<td>40/48 (% correct classified)</td>
<td>21/26 (% correct classified)</td>
<td>60/78 (% correct classified)</td>
<td>19/26 (% correct classified)</td>
</tr>
<tr>
<td>AD–Control validation cohort (expression + imaging)</td>
<td>36/46 (% correct classified)</td>
<td>19/23 (% correct classified)</td>
<td>18/23 (% correct classified)</td>
<td>17/23 (% correct classified)</td>
<td>76.9 %</td>
<td>79.7 %</td>
</tr>
<tr>
<td>AD–Control validation cohort (expression)</td>
<td>39/46 (% correct classified)</td>
<td>20/23 (% correct classified)</td>
<td>18/23 (% correct classified)</td>
<td>17/23 (% correct classified)</td>
<td>76.9 %</td>
<td>79.7 %</td>
</tr>
<tr>
<td>AD–Control validation cohort (expression + imaging)</td>
<td>32/46 (% correct classified)</td>
<td>16/23 (% correct classified)</td>
<td>16/23 (% correct classified)</td>
<td>16/23 (% correct classified)</td>
<td>76.9 %</td>
<td>79.7 %</td>
</tr>
</tbody>
</table>

(B) Performance of AD-Control classifier in MCI samples

<table>
<thead>
<tr>
<th>Classifier</th>
<th>MCI classified as AD-like (% correct classified)</th>
<th>MCI classified as AD-like (% correct classified)</th>
<th>MCI classified as AD-like (% correct classified)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCI cohort (expression)</td>
<td>60/168 (% correct classified)</td>
<td>32/41 (% correct classified)</td>
<td>32/41 (% correct classified)</td>
</tr>
<tr>
<td>MCI cohort (expression + imaging)</td>
<td>57/92 (% correct classified)</td>
<td>15/19 (% correct classified)</td>
<td>15/19 (% correct classified)</td>
</tr>
<tr>
<td>MCI cohort (expression + imaging)</td>
<td>45/92 (% correct classified)</td>
<td>12/19 (% correct classified)</td>
<td>12/19 (% correct classified)</td>
</tr>
<tr>
<td>MCI cohort (expression + imaging)</td>
<td>64/92 (% correct classified)</td>
<td>16/19 (% correct classified)</td>
<td>16/19 (% correct classified)</td>
</tr>
</tbody>
</table>

For bootstrapped data, the mean value and 95% confidence interval (95% CI), measured over 200 models, is shown.
Fig. 2. Bootstrap results for AD-Control classifier in classifying AD and Control samples. The graphs demonstrate the mean accuracy of the classifier measured from bootstrapping (short dashed line) and the accuracy measured using the “out of bag” classifier in the validation cohort (long dashed line). Results are shown for (A) expression classifier alone, (B) expression and imaging classifiers combined, (C) imaging classifier alone, and (D) expression classifier alone in the subset of individuals with sMRI measures.

by t-tests in the original training cohort. Using this method, the accuracy of the expression classifiers was slightly higher than we had estimated, with a mean accuracy of 78.2% (95% CI, 77.4–79.0%), a sensitivity of 79.6% (95% CI, 78.4–80.8%), and a specificity of 76.8% (95% CI, 75.7–77.9%) (Fig. 2A; Table 3A).

Structural MRI is able to accurately detect subtle changes in the thickness or volume of particular brain areas that occur during AD pathogenesis and is believed to be associated with the onset of neuronal dysfunction and cell loss. We first investigated the power of sMRI to classify the AD and control subjects using 68 local cortical thickness and 41 brain volume MRI measures collected in a subset of our cohort around the time of blood sampling [34, 36, 59, 60] (Supplementary Table 1). As six patients used in our expression validation cohort did not have neuroimaging measures, we initially re-evaluated our 203 probes on this subset of 46 patients. Accuracy, sensitivity, specificity, PPV, and NPV were all 69.6%, indicating a small drop in performance in this subgroup. However, when bootstrapping, the mean accuracy achieved was 76.0% accuracy (95% CI, 75.1–76.9%), with a sensitivity of 76.4% (95% CI, 75.1–77.7%) and specificity of 75.4% (95% CI, 74.1–76.7%) (Fig. 2D; Table 3A), values similar to that observed in the larger expression cohort (Fig. 2A; Table 3A) and demonstrating that bootstrapping is important for assessing the true variation in accuracy within a population. The sMRI measures correctly classified 39 people out of 46 tested (84.8% accuracy), corresponding to 20 out of 23 AD subjects correctly included (87.0% sensitivity), and 19 out of 23 control subjects correctly excluded (82.6% specificity) (Table 3A). The mean accuracy from bootstrapping was lower (82.0%,
95% CI, 81.3–82.7%), due to fewer AD subjects being correctly classified over the 200 models (80.0% mean sensitivity; 95% CI, 78.9–81.1%), although specificity was slightly improved (84.0% specificity; 95% CI, 82.9–85.1%) (Fig. 2C; Table 3A). The Kappa coefficient was 0.31 (p-value <0.05) demonstrating good agreement between the imaging and expression classifiers.

Combining the expression and imaging measures resulted in a small increase in mean accuracy following bootstrapping (84.0; 95% CI, 83.3–84.7%) (Fig. 2B; Table 3A). This was due to a small increase in the number of AD subjects being correctly classified (84.4% mean sensitivity; 95% CI, 83.4–85.4%), than achieved with imaging or expression alone.

Performance of diagnostic classifier in MCI subjects

To evaluate whether our diagnostic blood gene expression classifier could be an early marker of cognitive dysfunction, we tested it in 118 MCI subjects to see how many subjects would be classified as “AD-like” rather than as controls, and whether there was any preference for subjects who subsequently received an AD diagnosis within 2 years (MCl) compared to those who had not received an AD diagnosis (MClc), but nevertheless were likely to have an AD-endpoint. It categorized the majority of MCI subjects as “AD-like” (90/118 subjects, 76.3%) rather than as controls, which corresponded to 32/41 MClc (78.0%) and 58/77 MCInc (75.3%) (Table 3B). When bootstrapped, the mean accuracy was higher (86.0; 95% CI, 84.6–87.4%), with a similar proportion of MCInc and MClc being classified as AD-like (mean 87.2 and 85.5%, respectively) (Fig. 3A; Table 3B).

Although there were only a limited number of MCIc subjects with sMRI scans available for analysis (only 19 subjects with imaging compared to 41 available with expression), we assessed the performance of our diagnostic expression classifier in the MCI and MCInc groups separately and compared this to the performance of sMRI based models in 92 individuals for which sMRI and expression data were both available. Of the 92 people with MCI, 64 were classified as AD-like (69.6%) using the expression classifier, corresponding to 48/73 MCIc (65.8%) and 16/19 MCIc (84.2%). When the sample cohort was bootstrapped, the mean accuracy was 70.4% (95% CI, 69.8–71.0%). Compared to the whole expression cohort, accuracy in this subpopulation was lower and while the mean sensitivity in classifying MCIc was high (89.0%; 95% CI, 88.4–89.7%), the mean sensitivity to classify MCInc as AD-like was far lower than in the whole population analyzed (65.6%; 95% CI, 65.0–66.2%) (Fig. 3D; Table 3B).

sMRI alone appeared less able to classify MCInc as “AD-like” (63.2%) compared to the expression classifier (84.2%), classifying only 12 out of 19 patients as AD-like, suggesting imaging is not as sensitive at identifying potential prodromal AD subjects compared to the expression classifier, even when these individuals were within two years of receiving an AD diagnosis. In the MCInc group, the imaging classifier was no better than chance (33/73; 45.2%) at classifying individuals as “AD-like” or control, whereas the expression classifier identified more “AD-like” individuals in the MCInc population (65.8%) (Table 3B).

DISCUSSION

We have shown that there may be potential in using blood expression as a diagnostic classifier for AD. Specifically, we have identified 48 genes (50 probes) which can distinguish between AD and normal elderly control subjects with an accuracy of 75% in a validation cohort. This performance is consistent with previously reported accuracies using blood gene expression measures [31–33] and is similar to other single or multivariate classifiers incorporating measures from CSF and/or neuroimaging [64–68].

No marker would be expected to achieve perfect accuracy when tested in a population of people clinically diagnosed with AD as AD is a very heterogeneous disease with symptoms similar to other dementias and is therefore not always correctly diagnosed by a set of clinical symptoms alone. Clinical diagnosis generally achieves 80–90% accuracy against a postmortem AD diagnosis [54]. Furthermore, some of the pathologically relevant changes detected by a marker may occur before symptoms emerge in apparently normal people, leading to their apparent misclassification. Amyloid deposition is not only possible to detect in people with MCI, but is also present in 20–40% of cognitively normal elderly people, up to 10 years before AD diagnosis [11, 14, 15, 20, 69–75]. Some of the normal elderly subjects in our study displayed blood expression changes characteristic of people with AD. Although speculative, this may not be misclassification by the marker, but could instead represent subtle peripheral changes occurring in pre-clinical disease.

Of the 39 genes included in the classifier with known function, eight are associated with mitochondrial...
activity and oxidative phosphorylation including subunits of the electron transport chain or constituents of mitochondrial ribosomes and 12 are involved in translation. Changes in the expression of genes associated with these processes in AD blood was recently described in detail by us [76] and genes involved in these processes are enriched in other AD gene expression classifiers [32, 33]. These processes are also significantly altered in the brains of people with AD [77–80], which may represent a common response to the same disease-associated signal(s).

There are a number of limitations to this study, namely the sample size for validation and the lack of additional cohorts to further test the classifier. By performing power calculations using the MVPower R package, we have shown that our sample size has 80% power with a conservative effect size estimate of 0.2 and 100% with an effect size of 0.7 as derived from our data. We have addressed the second issue by applying a robust bootstrapping approach, considered by the community to be the most appropriate approach for evaluating a classifier when a second cohort is unavailable [61]. When new datasets become available to us we will be able to test the performance of our classifier in additional samples. We are confident that the findings will replicate, as the predictive genes are known to be AD-related and these processes have been previously identified in AD classifier studies, as described above.

In order to further explore the performance of the blood expression classifier we developed, we analyzed measures of cortical thickness and local
brain atrophy determined by MRI where available in the same subjects. Measures of brain atrophy correlate with the severity of cognitive impair ment and rates of neuronal and synaptic loss in AD [60, 81, 82]. Our best neuroimaging classifier achieved an accuracy of 85%, similar to what has been reported previously [34, 83] and only slightly better than the blood expression classifier. Overall these results add weight to the diagnostic validity of using the blood expression classifier we have developed for diagnosing AD independently of information provided by other methods such as structural MRI measures. However, further evaluation is needed to validate its performance in independent samples.

While a diagnostic marker to back up a clinical diagnosis of AD is of value, an arguably more important goal is to improve the likelihood of identifying people with prodromal AD as early as possible in the course of their disease, which will reduce variability in clinical trials and allow recruitment and eventually treatments to begin as close as possible to disease onset before irreversible brain changes occur. This is an enormous challenge as there are no obvious transition points to AD. Markers are likely to be far more sensitive at picking up early disease-relevant changes than phenotypic measures. In practice, it is likely that multimodal markers will be used in combination to build up a profile of disease likelihood for each individual. One of the earliest indications of AD pathology that have been reported is decreased CSF Aβ1-42 levels and increased brain Aβ plaque deposition, which can be observed in pre-clinical disease in some individuals. The connection between CSF Aβ1-42 and AD pathology is well established, with CSF Aβ1-42 correlating with post-mortem plaque [14], tangle number [84], aggregate deposition evaluated by PIB-PET and FDDNP-PET [20], and atrophy measured by CT [85].

Pathophysiological changes emerging in the course of AD are believed to be associated with a dynamic temporal sequence of changes in the brain and other biological fluids [12, 86]. Our data adds weight to the body of evidence suggesting there may be potential for markers capturing peripheral blood changes occurring in early prodromal AD, as a high percentage of people with MCI who were expected to have a likely AD-endpoint displayed changes sufficiently different from normal to classify them as “AD-like” rather than as controls. The AD classifier appeared to make only a small distinction in favor of classifying MCI subjects who were close to receiving a diagnosis of AD as AD-like than those who did not receive an AD diagnosis over the same short follow-up period of two years. This is in contrast to the performance of the imaging marker that classified far fewer MCI subjects overall as AD-like compared to the gene expression classifier, even those subjects who subsequently received an AD diagnosis within two years. This may suggest measures of structural brain changes have less sensitivity early in disease compared to the expression marker in blood we describe, but this will be uncertain until the MCI non-converters have been followed for a longer period. Changes detected by sMRI appear to occur at a later stage of disease than changes in Aβ and tau [11, 12, 87–89], and it will therefore be important in the future to evaluate our blood expression classifier alongside measures of CSF or brain Aβ to see if the high classification of AD among the MCI subjects reflects early detection of prodromal disease and to map the dynamic onset of changes we observe.

Due to small sample numbers and therefore insufficient statistical power when we divided the MCI group in to those who subsequently received an AD diagnosis within two years and those who did not, we were unable to develop a classifier able to distinguish these two groups. However, the majority of MCI subjects had peripheral blood gene expression changes in common with AD patients including those who did and did not receive an AD diagnosis within two years. This is a finding we have previously observed in a study of cell based proteins in blood comparing AD and MCI subjects [23] and is similar to the profile of reduced Aβ1-42 observed in the CSF in some control and many MCI subjects, which is believed to represent a pathological process that significantly departs from normality [37]. Two factors are likely to contribute to this observation; first, around half of the subjects in the MCI cohort were selected for analysis because we knew they had subsequently converted to a probable AD diagnosis. Second, our MCI sample population are drawn from specialist memory clinics and are therefore more likely to represent people with established cognitive problems resulting from prodromal AD than might be expected from a cohort drawn from a community population, as we previously reported from a meta-analysis of longitudinal studies of MCI [90].

Alternatively, our blood expression marker may represent an endophenotype shared by a group of diseases with related etiology. Our two year follow-up period is insufficient to confirm which of the normal elderly control or MCI subjects have prodromal AD and will therefore eventually receive a clinical diagnosis of AD. Pathological confirmation of an AD diagnosis will take longer still. We know of no existing studies with blood collected for RNA purposes where postmortem diag-
nosis confirming AD would allow such analyses. While confirming the specificity of our marker is an important goal, the existence of an endophenotype, defined here as a disease-associated expression phenotype even though not necessarily specific to or present in all individuals still has clinical relevance. Such a marker can assist in the development of drugs targeting relevant and common disease endophenotypes and does not necessitate unique disease association to achieve efficacy in a relevant disease population. It could also be useful as part of a staged diagnostic approach, providing a cost effective method of identifying individuals with a greater likelihood of developing disease who could subsequently be referred for more expensive diagnostic tests.

Blood is emerging as a very promising tissue in which to achieve a non-invasive and relatively inexpensive assay for detection of AD. In addition to changes in blood gene expression [31–33, 91], plasma protein panels have been described which accurately distinguish AD from normal elderly control people and/or MCI subjects who convert versus those who remain stable, or for predicting disease progression [92]. There are also many reports of individual proteins with altered abundance in AD plasma suggesting blood is a realistic tissue in which to identify markers of AD [23, 93–102]. However, there are still technical issues to overcome as not all protein measures reliably change across studies [103]. In the future, it will be important to explore the performance of different combinations of measures to achieve the most accurate biomarker for diagnosis and compare blood markers with more established biomarkers linked to pathophysiology and establish relative temporal patterns of change during disease progression across different biomarker modalities.

In conclusion, we have shown that peripheral blood shows promise as an AD-associated diagnostic blood gene expression marker that may be useful very early in pre-clinical disease. It performed similarly to using neuroimaging measures alone in AD and normal elderly control samples. Future work will evaluate this blood expression marker in subjects from which CSF measures of Aβ1-42 and/or PET amyloid have been measured in order to further explore the specificity and timing of these changes. The existence of such a marker in an accessible tissue such as blood would contribute significantly toward efforts to identify and treat people with AD very early in their disease before significant neuronal loss has occurred. For example, CSF Aβ1-42 and tau measurements to enrich for likely AD are predicted to reduce sample size by 67% and costs by 60% in a clinical trial compared to a trial with unselected MCI subjects [104]. Evaluation of the marker in related diseases will establish whether the changes we see are specific to AD or are shared across similar diseases. A lack of disease specificity would not diminish the value of a marker able to capture a common disease-associated endophenotype for which a specific treatment with real clinical value could be developed.

ACKNOWLEDGMENTS

We thank Rutina Leung, Megan Pritchard, Belinda Martin, Kathryn Lord, Nicola Dunlop, and Catherine Tunnard for sample collection and neuropsychiatric assessments of the London cohort. We thank all participants and their families. This work was supported by InnoMed (Innovative Medicines in Europe), an Integrated Project funded by the European Union of the Sixth Framework program priority FP6-2004-LIFESCIHEALTH-5, Alzheimer’s Research UK, The John and Lucille van Geest Foundation, the NIHR Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and Institute of Psychiatry Kings College London, and NIA/NIH RC1 grant 1RC1AG035610 (to GC). Authors’ disclosures available online (http://www.j-alz.com/disclosures/view.php?id=1502).

REFERENCES


