A Novel Blood Test for the Early Detection of Alzheimer’s Disease

Phil. D. Ryea,1,2, Birgitte Boomstra Booij1,3, Gisle Gravea, Torbjørn Lindahl4, Lena Kristiansen5, Hilde-Marie Andersen6, Peter O. Hornaksveen7, Harald A. Nygaard8, Mala Naik9, Dagne Hopreklstadf, Peter Wetterbergg, Christer Nilssonh, Dag Aarslandi, Praveen Sharmaa and Anders Lönneborga

1DiaGenic ASA, Oslo, Norway
2Memory Clinic, Sanderud Hospital, Ottestad, Norway
3NKS Olaviken Hospital for Old Age Psychiatry, Erdal, Norway
4Section for Geriatric Medicine, Department of Public Health and Primary Health Care, University of Bergen, Bergen, Norway
5Geriatric Section, Haraldsplass Deaconess Hospital, Bergen, Norway
6Geriatric Clinic, Stavanger University Hospital, Stavanger, Norway
7Medical Division, Department of Geriatrics, Memory Clinic, Ullevål University Hospital, Oslo, Norway
8Department of Clinical Sciences, Lund University Hospital, Lund, Sweden
9Division of Mental Health Services, and Institute of Clinical Medicine, Stavanger University Hospital, and Akershus University Hospital, University of Oslo, Lørenskog, Norway

Accepted 8 September 2010

Abstract. Despite a variety of testing approaches, it is often difficult to make an accurate diagnosis of Alzheimer’s disease (AD), especially at an early stage of the disease. Diagnosis is based on clinical criteria as well as exclusion of other causes of dementia but a definitive diagnosis can only be made at autopsy. We have investigated the diagnostic value of a 96-gene expression array for detection of early AD. Gene expression analysis was performed on blood RNA from a cohort of 203 probable AD and 209 cognitively healthy age matched controls. A disease classification algorithm was developed on samples from 208 individuals (AD = 103; controls = 105) and was validated in two steps using an independent initial test set (n = 74; AD = 32; controls = 42) and another second test set (n = 130; AD = 68; controls = 62). In the initial analysis, diagnostic accuracy was 71.6 ± 10.3%, with sensitivity 71.9 ± 15.6% and specificity 71.4 ± 13.7%. Essentially the same level of agreement was achieved in the two independent test sets. High agreement (24/30; 80%) between algorithm prediction and subjects with available cerebrospinal fluid biomarker was found. Assuming a clinical accuracy of 80%, calculations indicate that the agreement with underlying true pathology is in the range 85%-90%. These findings suggest that the gene expression blood test can aid in the diagnosis of mild to moderate AD, but further studies are needed to confirm these findings.

Keywords: Alzheimer’s disease, biomarker, blood, diagnostic test, gene expression, RNA

Supplementary data available online: http://www.j-alz.com/issues/23/vol23-1.html#supplementarydata05

INTRODUCTION

Globally, the number of people with dementia is estimated to increase from 24.3 million in 2005 to 81.1 million by 2040 [1]. Considering that worldwide societal costs in 2009 was estimated to US$...
422 billion [2] the costs for the society likely will pass US$ 1000 billion in 2040. Since Alzheimer’s disease (AD) is the most frequent cause of dementia (around 70%), these figures represent a strong driving force for the development of new drugs and diagnostic tools for the treatment and patient management of AD.

Current diagnostic criteria of AD are based on clinical inclusion and exclusion criteria [3]. The diagnostic accuracy is ranging from 65 to 96%, and the specificity of AD versus other dementias is even lower [4]. There is evidence that the AD process develops several years before clinical symptoms, and early intervention with disease-modifying therapies is likely to become available. On this basis, there is a need to develop biomarkers to increase the diagnostic accuracy of early AD, and promising tools such as structural MRI, amyloid imaging using PET, and CSF-based proteins exist [5–7]. However, these may not be sufficiently accurate (structural MRI), costly or not generally available (PET), or unpleasant and less user-friendly (CSF). From this perspective, potential blood-based markers are being explored. Although there is no commercial blood based diagnostic biomarker test for AD available today, several independent studies have indicated that a peripheral blood based test could be used for diagnostic profiling in neurological diseases [8–16].

A convenient blood test to detect AD would be a valuable addition to the clinician’s diagnostic strategy for subjects with memory complaints presenting at clinics.

Based on a disease classifier developed on a whole genome expression screen in blood, we have previously selected 1239 gene probes that are able to predict AD [17]. Further selection of gene probes was achieved by ranking and retaining predictive value when using RT-PCR instead of hybridization based platform using novel independent sets of samples from AD patients and age-matched cognitively healthy controls, thus reducing the number of probes to 96 in total. The gene identities represented in the 96-gene array used in this study are shown in supplemental information Table 1 (available online: http://www.j-alz.com/issues/23/vol23-1.html #supplementarydata05).

The aim of the present study was to develop a model based on the method of partial least squares regression (PLS-regression) using the expression in blood from an array of 96 genes and investigate its diagnostic accuracy in the detection of early AD.

METHODS

Participants

Patients with probable AD and cognitively healthy, age- and gender-matched controls were recruited from 13 clinical centers located in Norway and Sweden from January 2007 until January 2009. The procedures including blood sample collection were approved by local ethics committees and written informed consent was obtained from each subject included. The clinical evaluations were performed by clinicians with experience in diagnosing dementia. A clinical interview and medical examination were performed, and physical diseases and drug use were recorded. The evaluation included an overall cognitive screening test (Mini-Mental State Examination (MMSE)) [18] and a set of standardized cognitive tests assessing verbal and non-verbal memory, mental speed and executive function (shifting attention): The Clock Test [19, 20], Kendrick OLT test [21], 10-word test [22, 23], and Trail Making test A and B [24]. Also, the clinical dementia rating (CDR) scale [25] was scored based on a detailed clinical interview with the patient and a caregiver. Imaging (either MRI or CT) and routine blood tests were performed to exclude causes of dementia other than AD. In a subgroup of 28 patients and 2 healthy controls lumbar puncture was performed and CSF was analyzed for AD biomarkers: the Innotest products (Innogenetics, Gent, Belgium) for detection of amyloid-β, Tau and P-Tau in CSF were used according to the manufacturer’s instructions.

Subsequently, a designated group of clinicians from the recruitment centers evaluated all available patient data to establish a consensus diagnosis of AD according to ICD-10 and the National Institute of Neurological and Communication Disorders and Stroke/Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) [3].

Inclusion of patients into the AD group was based on: 1) a diagnosis of probable AD according to ICD-10 and NINCDS-ADRDA; 2) an MMSE score equal to or lower than 27; and 3) a CDR scale score equal to or greater than 1 and less than or equal to 2. Cognitively healthy control subjects were recruited from spouses of patients or from community-centers for the elderly to the study based on: 1) an MMSE score greater than 27 and accurate clock test; and 2) age 50 years or greater; and 3) no history suggesting brain disease or cognitive decline.

The study flow diagram in Fig. 1 provides further details on sample use. Maintaining an equal number of
Fig. 1. Study flow diagram. Recruitment period: May 2006–January 2009.
samples and a balance in gender and mean age between AD and control within each test set, the selected eligible samples were distributed randomly into a calibration, an initial validation and an extended validation set, prior to Real Time PCR analyses, the characteristics of the three groups are shown in Table 1. The two validation tests were separated in time. In total, blood samples from 412 subjects were included, and the distribution of AD and control subjects for the calibration and two validation test sets are shown in Table 1. Patient data for the calibration and the two validation test sets are provided in Supplementary Tables 2-4.

Sample collection

Venous blood samples (2.5 mL) were drawn into PAXgene™ tubes (Becton & Dickinson, Qiagen Inc., Valencia, CA) according to the manufacturer’s instructions. Tubes were incubated at room temperature overnight prior to freezing and storage at −70°C or below and subsequent transport on dry ice to the testing laboratory. RNA was extracted from all samples within 6 months of blood draw.

Gene probes

Based on a disease classifier developed on a whole genome screen comprising in total 126 clinically diagnosed AD patients, 98 age-matched cognitively healthy controls, 28 young controls, 28 Parkinson’s disease (PD) patients, and 10 subjects with mild cognitive impairment (MCI), a total of 1239 gene probes were originally selected [17]. Further selection on novel independent sets of samples from AD patients and age-matched cognitively healthy controls and further evaluation of the original set of gene probes emanated in a final selection of 96 probes in total including controls (results not shown). The selection was essentially done in two steps: an initial selection based on ranking the 1239 predictive genes based on predictive value and a second selection of genes retaining predictive value when using RT-PCR instead of hybridization based platform. The known gene identities represented in the 96-gene array used in this study are shown in supplemental information Table 1.

RNA extraction and cDNA synthesis

Total RNA was extracted from blood samples using PAXgene™ Blood RNA kit (Qiagen Inc., Valencia, CA) according to the manufacturer’s instructions. Total RNA was stored at −70°C or below until analysis. The RNA was assessed for quality using the NanoDrop ND-1000 spectrophotometer (NanoDrop, Wilmington, DE) and the Agilent 2100 BioAnalyzer (Agilent, Santa Clara, CA), with sample acceptance limits RIN ≥ 7.3, 28S/18S ≥ 1.0, A260/A230 ≥ 1.0, A260/A280 ≥ 1.8; RNA concentration ≥ 15 ng/μL. cDNA was prepared in batches using the High-Capacity cDNA Reverse Transcriptase kit (Applied Biosystems, Foster City, CA) according to manufacturer’s instructions.

Real time PCR

Real Time PCR was performed using the universal PCR Master Mix reagent and the ABI Prism 7900HT Fast System (Applied Biosystems, Foster City, CA). The 96-well format low density array Microfluidic cards (MFCs) were purchased from Applied Biosystems with DiaGenic’s proprietary 96-gene array set for Alzheimer’s disease, ADtect® (DiaGenic ASA, Oslo, Norway). The 96-gene array was selected from a previ-
ously performed whole genome screen study [17]. The probes used in this array are represented by the genes shown in supplemental information Table 1. The layout of assays on the MFC enabled 4 individual samples to be run simultaneously on each MFC. Samples were run on MFCs according to a randomization scheme such that 2 AD subjects and 2 non-AD subjects were run on each MFC. All MFCs were prepared and run on an ABI Prism 7900HT Fast System at an accredited central laboratory (DNAvision, Belgium) according to the methods previously outlined and the Applied Biosystems SDS software v2.2.2.

**Data analysis**

All data was analyzed at DiaGenic (DiaGenic ASA, Oslo, Norway). In the final calibration model, data from the primary analysis set was transformed taking the base 10 logarithm, and missing values k-Nearest Neighbor (k-NN) imputed using standardized values and ten nearest neighboring values [26]. Transformed and imputed values were then normalized by mean normalization.

For each assay the mean observed raw quantification cycle (C_q) value was calculated and missing values were imputed using k-NN. Partial Least Square Regression (PLSR) and Leave One Out Cross-Validation (LOOCV) were used to build the calibration model and to estimate the classification efficacies. A number of calibration models were fitted and evaluated: for each model the LOOCV estimates of sensitivity, specificity, accuracy and area under the receiver operating characteristic curve (AUC) were used as a measure of the models performance. When the model is used to score other samples values outside this range are defined as missing. The disease algorithm was developed using PLSR on the calibration set (n = 208) (Table 1).

Furthermore, ten of the eleven blood collection sites used in the validations had not previously been used. The run of the samples on MFCs for the initial validation test and the second and expanded validation test were separated in time with 15 months. Imputed, log-transformed and normalized data were predicted using the final calibration model. For each sample a test score (classification) was assigned. The data was compared to the clinical diagnosis and the accuracy, specificity, sensitivity, AUC and a positive likelihood ratio was calculated.

The estimation of accuracy in the calibration and validation analyses assumes 100% correct classification [27–29], which is unlikely in the current study [5, 30, 31]. Therefore, a simulation model was generated together with a maximum likelihood estimate (MLE) to determine the expected accuracy (a more detailed description of the method can be found in the supplemental material).

**RESULTS**

The algorithm in the PLS model predicted the class of 149/208 samples (accuracy, 71.6 ± 6.1%) in agreement with clinical diagnosis, including 74/103 AD samples (sensitivity, 71.8 ± 8.7%) and 75/105 cognitively healthy controls (specificity, 71.4 ± 8.6%). The Positive Likelihood Ratio (PLR) was 2.51 and the AUC was 0.77 (Table 2).

The imputed, log-transformed and normalized data from the two validation test sets were predicted using the model described. The classification agreement with clinical diagnosis shows almost identical performance characteristics to those observed for the calibration analysis. In the initial validation test the algorithm predicted the class of 53/74 samples in agreement with clinical diagnosis (accuracy 71.6 ± 10.3%), including 23/32 AD samples (sensitivity, 71.9 ± 15.6%) and 30/42 cognitively healthy controls (specificity, 71.4 ± 13.7%). In the extended validation

<table>
<thead>
<tr>
<th>Performance characteristics</th>
<th>MMSE score</th>
<th>Calibration (%)</th>
<th>Validation (%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7–29</td>
<td>71.6 ± 6.1</td>
<td>71.6 ± 10.3</td>
<td>71.5 ± 7.8</td>
<td>71.6 ± 4.4</td>
</tr>
<tr>
<td>20–29</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>71.4 ± 6.2</td>
</tr>
<tr>
<td>7–19</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>71.1 ± 7.7</td>
</tr>
<tr>
<td><strong>Sensitivity</strong></td>
<td>71.8 ± 8.7</td>
<td>71.9 ± 15.6</td>
<td>70.6 ± 10.8</td>
<td>71.4 ± 6.2</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71.4 ± 8.6</td>
<td>71.4 ± 13.7</td>
<td>72.6 ± 11.1</td>
<td>71.8 ± 6.1</td>
<td></td>
</tr>
<tr>
<td><strong>Positive likelihood ratio (PLR)</strong></td>
<td>2.51</td>
<td>2.52</td>
<td>2.57</td>
<td>2.53</td>
</tr>
<tr>
<td><strong>Area under the curve (AUC)</strong></td>
<td>0.77</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
</tbody>
</table>
test the algorithm predicted the class of 93/130 samples in agreement with clinical diagnosis (accuracy, 71.5 ± 7.8%), including 48/68 AD samples (sensitivity, 70.6 ± 10.8%) and 45/62 cognitively healthy controls (specificity, 72.6 ± 11.1%). The diagnostic performance is summarized in Table 2. The PLR was 2.52 and 2.57 and AUC was 0.74 for both test sets (Table 2 and Fig. 2).

Among subjects with CSF available for analysis, the classification model predicted 24 of the 28 subjects as AD, and the two healthy controls were both predicted as non-AD. Although the number of samples is limited it suggests that there may be a good agreement between the CSF biomarkers and the classification model.

No significant effects were observed for the most common co-morbidities, such as diabetes, coronary disease, depression, previous strokes, hypertension, cancer and rheumatoid arthritis and AD where predicted with a similar agreement to clinical diagnosis as subjects without these co-morbidities (results not presented). Similar findings emerged when only patients taking antidementia drugs were included, indicating that there is limited interference from these medications on the test.

Comparing prediction accuracy in AD subjects with moderate to severe (MMSE 7-19) or mild (MMSE 20-27) dementia, no significant difference emerged (Table 2).

From the MLE and diagnostic accuracy simulations it can be shown that, assuming an accuracy of 80% for the clinical diagnosis compared with definite neurologically confirmed AD pathology, the accuracy of the 96-gene array test is most likely in the range 85%–90% (see Supplementary data).

DISCUSSION

We have developed a novel 96-gene expression test to aid in the diagnosis of mild to moderate AD. In this study the test has been validated using samples from 100 AD and 104 elderly control subjects. The performance characteristics showed good agreement with clinical diagnosis reaching 71.6% and 71.5% in the two separate validations. The similar performance between the calibration test (71.6%) and the two validation tests, which were separated in time with 15 months, suggests that a large enough number of samples were included in the calibration to yield a reliable prediction model and it also suggests that test performance can be reproduced. However, more independent validation tests are needed to further confirm the tests performance.

The observed accuracy is lower than the 80% minimum suggested by the 1998 Alzheimer’s Association Working Group to be required for an AD biomarker [32]. There are several possible explanations for the somewhat lower sensitivities and specificities observed in the current study. It is possible that not all AD patients express changes that is detected with the predictive 96 gene model used in this test and that some healthy controls express the same changes due to aging, preclinical AD or other unknown reasons. Histological confirmation of disease is desirable but has not been used in this study. Since the test requires specific sample procedures, brain bank samples cannot be used, and histological confirmation would require many years, in particular since we also wanted to include patients with early and mild disease.

Although the diagnosis was based on standard clinical consensus criteria, and standardized tests were used, it is possible that non-AD patients were included. For example, we did not employ standardized screening measures for symptoms of dementia with Lewy bodies (DLB), the second most common neurodegenerative dementia after AD, representing 15%–20% of the dementia population, and we cannot exclude the possibility that some of the cases have DLB and not AD. In addition, the control samples were required to have MMSE of 28 or higher and normal clock test, and no history of cognitive decline. However, standardized neuropsychological tests were not employed, and it is thus possible that some of the control subjects have a mild cognitive impairment and thus possibly very early
AD despite the normal history. Finally, novel biomarkers such as structural and functional imaging and CSF biomarkers are promising new biomarker candidates for AD based on established pathological changes in AD. These were not systematically administered in these cohorts (only a limited subset had CSF analyzed, and structural imaging was used only to exclude other pathologies). Employing such markers to enhance the diagnostic accuracy might have contributed to an improved performance of our test, and such studies are planned. Compared to these biomarkers, a blood-based test is cheaper, more user-friendly and convenient and thus might still have a role in the diagnostic work-up of people with cognitive impairment. A reduced apparent accuracy could also be attributed to the lack of a reliable reference or ‘gold standard’, the absence of which increases the uncertainty in the observed accuracy for the test. When an imperfect standard is used to evaluate a diagnostic test, many commonly used measures of test performance are distorted. FDA has realized that this bias can create inaccurate estimates and in a recent guideline [33] they have addressed the issue of biased estimates of diagnostic performance and recommend that the terms positive percentage of agreement and negative percentage of agreement be used instead of specificity and sensitivity in cases where no gold standard is available. In agreement with these recommended terms we have instead of accuracy chosen to use the term agreement with clinical diagnosis. Nevertheless, this bias is predictable and can be accommodated using different approaches [29]. In the current study these methods were used to define a more accurate estimation for the observed test accuracy when compared with the underlying true pathology. The results indicate that the likely accuracy of the 96-gene array test would be improved by 13%–18% depending on how the accuracy of the imperfect gold standard was estimated. Although this observation clearly needs to be substantiated with additional clinical studies, it does suggest that the diagnostic accuracy for the 96-gene array test may be in the range of 85%–90% when compared with underlying true pathology.

Although CSF samples were not a pre-requisite for clinical diagnosis, data was available for 28 AD cases and two (2) controls. Twenty-four (24) of the AD subjects and both controls were correctly predicted using the 96-gene array test giving an agreement between the tests of 80%. It is known that the CSF biomarkers do not completely agree with the neuropathology of AD [34], and the number of samples with CSF is limited. Still, these results suggest a good correlation between the CSF biomarkers and the prediction based on gene expression pattern in blood. How well these biomarkers correlate need to be further explored and further studies are currently in progress to compare the diagnostic performance of the 96-gene array with the CSF biomarkers Aβ, total tau, and p-tau.

All genes included in the 96-gene array are expressed at medium to high rates and should thus not be products of the low transcription of any tissue-specific genes in non-specific cells often referred to as “illegitimate” transcription. The demonstration in our study that the Alzheimer’s condition reproducibly affects the transcriptional regulation of a wide variety of genes in peripheral blood indicates that this is a real response and that it can be used as a diagnostic biomarker for the disease.

Although the function of many of the included genes at present not is known, the 96-gene array still covers a wide range of known biological processes (Supplementary Figure 1) likely combining both local and systemic responses to the disease. Compared to what would normally be expected if there was a random distribution of biological functions among the genes, six processes were significantly over represented including nucleoside, nucleotide and nucleic acid metabolism and transport, apoptosis, and other forms of protein metabolism, cell cycle control and transcription (data not shown).

It has been shown that about 80% of genes expressed in the brain are also expressed in blood cells [35] and that at least some of the genes have comparable regulation in both tissues [35, 36]. This is also reflected in the composition of the 96 gene array test used in this study. The assays in the test cover a wide range of known functions associated with AD pathology, such as Aβ, P-tau, and mitochondrial processing as well as inflammation, calcium regulation and ubiquitin-associated protein processing (Supplementary Table 1). It is interesting to note that four of the genes included, GRB2, TARDBP, TCEB3, and TNF, are also present in the current AlzGene database [37]. An additional 5 genes have been associated with AD in the literature and a further 15 genes are associated with essential functions in brain and/or neurons (Supplementary Table 1).

Although the gene probes have been selected based on predictive value in the algorithm and not on a presumed association with AD pathology, still there are 32 of the 84 genes represented in the model that have some form of association with AD. The present results indicate that changes in gene expression occurring in AD are not restricted to the brain, but occur in peripheral blood cells as well.
A recent study identified a panel of 18 signaling proteins [38] in plasma that were able to classify AD from non-AD samples with a high diagnostic accuracy. Although it is difficult to make any direct comparisons between protein expression and gene expression studies, it is interesting to note that none of those proteins listed appear to be coded by any of the genes present in our 96-gene array. This may reflect the different approaches used in the selection of protein/gene candidates, but also highlights the complexity of the peripheral responses.

Additional validation studies are still needed to substantiate the wider application of this 96-gene array test. Nevertheless, the current findings do suggest that a gene expression test for AD based on the peripheral blood transcriptome is possible, and can also provide diagnostic information to facilitate clinical management of the disease.

ACKNOWLEDGMENTS

The contribution of samples and clinical expertise from the following investigators is greatly acknowledged: Cristina Alm (Uppsala), Ole Birkedal (Sandefjord), Ann-C.T. Backstrom (Stockholm), Dagny Hoprekstad (Stavanger), Lena M. Lundberg (Stockholm), Mala Naik (Bergen), Nina Voss Skaane (Oslo), Magda Tsolaki (Greece), Lars-Olof Wahlund (Stockholm), Peter Wetterberg (Oslo).


REFERENCES


