Methodology for Discovery of Alzheimer’s Disease
Blood-Based Biomarkers

Olivier C. Maes,1 Hyman M. Schipper,1,2,3 Howard M. Chertkow,1,2,3 and Eugenia Wang4,5

1Centre for Neurotranslational Research, Bloomfield Centre for Research in Aging, Lady Davis Institute for Medical Research, Montréal, Québec, Canada.
2Department of Neurology and Neurosurgery and 3Department of Medicine (Geriatrics), McGill University, Montréal, Québec, Canada.
4Gheens Center on Aging and 5Department of Biochemistry and Molecular Biology, School of Medicine, University of Louisville, Kentucky.

Alzheimer’s disease (AD) is a degenerative brain disorder. The disease also affects peripheral tissue such as peripheral blood mononuclear cells (PBMCs). Delineating biochemical alterations in AD blood constituents may possibly allow the identification of accessible footprints that reflect degenerative processes within the central nervous system. Here, we describe an integrated methodology for the generation of a blood-based molecular bio-repository, including the collection of clinical and demographic data for downstream stringent sample selection and stratification for the study of molecular signatures in AD. We report the simultaneous extraction of high quality and yield of DNA, RNA, and protein from PBMCs of individuals with sporadic AD, mild cognitive impairment, and normal elderly controls. We describe experimental designs and present examples for the discovery of underlying etiopathogenetic networks in sporadic AD. We suggest that PBMC-associated biomarkers may provide insights into the pathogenesis of AD and be used to monitor disease diagnosis and progression.

Key Words: Bio-repository—Biomarker—Peripheral blood mononuclear cell—Sporadic Alzheimer’s disease—Mild cognitive impairment.

ALZHEIMER’S disease (AD) is a major cause of cognitive decline and death among the elderly population, and its incidence is rapidly growing within the aging population (1). Over the last century, considerable progress has been made in the understanding of familial AD, with the identification of inheritable autosomal dominant mutations in amyloid precursor protein (APP), presenilin-1, or presenilin-2 genes; reviewed in (2). Sporadic AD accounts for 93% of cases with incompletely delineated complex genetic and environmental risk factors (2–4). In fact, various dysfunctions may contribute to the pathogenesis of sporadic AD, such as aberrant iron deposition, oxidative stress, mitochondrial insufficiency (3,5), calcium homeostasis (6), neuropeptidergic responses (7), cerebrovascular ischemia (8), and altered glucose and insulin metabolism (9).

Intermediate between normal elderly cognition and dementia, an emerging diagnostic category known as mild cognitive impairment (MCI) has been recognized (5,10). Although this diagnostic group consists of cases having different presymptomatic neuropathologies, a subgroup referred to as amnestic MCI with predominant memory impairment constitutes a high risk for, or represents the preclinical stage of, AD (11). Amnestic MCI shares common genetic and environmental risk factors with sporadic AD, including, for example, the susceptibility alleles, apolipoprotein E e4 (APOEe4), and glutathione S-transferase M3 (GSTM3) polymorphism (12).

The various forms of AD lead to common neuronal dysfunctions and are characterized by extracellular amyloid-β (Aβ) deposition (senile plaques) and intracellular hyperphosphorylation of tau (neurofibrillary tangles) in the brain (13). Various species of Aβ derived from the sequential cleavage of APP have been identified, including the most common Aβ1-40 and Aβ1-42 peptide residues (14). Aβ deposition is also found on blood vessels (15) and is detected in cerebrospinal fluid (CSF) and peripheral blood plasma (16). Its levels in the CSF together with phosphorylated tau may prognosticate for AD conversion in MCI (17). Unfortunately, lumbar puncture for CSF examination is relatively invasive and unsuitable for mass screening of the aging population. Readily accessible biomarkers from a noninvasive blood specimen would greatly facilitate the management of this condition (17,18).

Blood-based biomarkers have been investigated for a considerable time (19) and support the notion that peripheral alterations occur in AD pathology and may not only originate from central nervous system (CNS) signaling (20,21) but also leave systemic “footprints.” Thus, various components of blood may provide an ideal “window” into the CNS concerning the etiopathogenesis of neurological disorders such as AD (19). In fact, a selection of altered plasma proteins in AD has been shown to be predictive biomarkers for disease classification (21).

We previously described protocols for isolating blood components for biogerontological studies, including the processing of peripheral blood mononuclear cells (PBMCs) for biochemical extraction of DNA, RNA, and protein and the establishment of lymphoblastoid cell lines from frozen whole blood (22–24). Here, in addition to the extraction of
the three major biologic specimens (ie, DNA, RNA, and protein), we further process the RNA samples to small RNA fractions containing microRNAs (miRNAs). In general, miRNAs have emerged as powerful molecular silencers, regulating gene expression by either degrading the messages or inhibiting translation (25). These 18- to 22-nucleotide-long miRNAs are noted to be involved in various biologic functions and associated with human diseases including AD (26, 27).

In this article, we describe the flow path for neurological and translational research in AD using blood samples, from detailed clinical informatics to biomarker discovery using freely available bioinformatics tools (Figure 1). This scheme includes the steps of participant recruitment, clinical data registry, molecular specimen banking, sample selection, and experimental protocols. We provide examples of biomarker discovery from our previous case–control studies using PBMC samples from our bio-repository, such as the use of DNA for single nucleotide polymorphisms (SNPs) association studies and RNA for messenger RNA (mRNA) or miRNA expression profiling in sporadic AD (12, 27, 28). The goal of establishing a noninvasive blood-based bio-repository is to develop a comprehensive repertoire of lead miRNA, messenger, and protein biomarkers. This integrated approach will facilitate the delineation of underlying miRNA regulatory networks associated with AD.

A comprehensive repertoire of biomarkers from elderly populations with varying degrees of memory impairment is crucial for understanding dysfunctions associated with early stages of sporadic AD. Moreover, information on the various risk factors as well as longitudinal sampling of volunteers will aid in the discovery of specific disease onset biomarkers. Importantly, blood-based biomarkers may then facilitate prognostication in preclinical stages of the disease and hopefully provide leads for the treatment of this devastating disease at the presymptomatic stage.

**Methodology**

**Participants**

The study is first approved by a research ethics committee of the institutions concerned. For the AD research described here, approval was obtained from the Research Ethics Committee of the Sir Mortimer B. Davis Jewish General Hospital (JGH, Montreal, Quebec). Written informed consent is obtained from all volunteering participants, patients, and primary caregivers evaluated, as in our case at the McGill–JGH Memory Clinic. Normal elderly controls (NEC) participants include caregivers and family members of afflicted patients because they share environmental, social, and genetic backgrounds. NEC participants are also recruited through publicity in the Family Practice Clinics at hospitals and in the community (ie, elder care homes, community centers, etc.). Control participants with other neurological and AD risk–associated diseases (29) are also recruited from various clinics (see Supplementary Table S1).

Neuroimaging (ie, computed tomography and magnetic resonance imaging) is performed on all AD and MCI patients in order to exclude patients with other dementing conditions, such as normal pressure hydrocephalus. AD and MCI patients also undergo formal neuropsychological testing, as previously described (30). In this study, AD is diagnosed at the JGH–McGill University Memory Clinic according to National Institute of Neurological and Communicative Disorders and Stroke—Alzheimer’s Disease and Related Disorders Association criteria (31). Only cases ascertained as probable AD on clinical and radiological grounds are included. MCI meet the criteria defined...
in the working group of Winblad and colleagues (32) and are classified “0.5” according to the Washington University clinical dementia rating scale (33), as described (10). NEC participants score within 1 SD of age- and education-standardized means on a series of memory and attention tests (30).

The Mini-Mental State Examination (MMSE) (34) is administered to all participants and patients. Participants also complete a health questionnaire that addresses first-order family history of AD, age, ethnicity, education, health status, medications, nutritional status, intake of antioxidant vitamins, and alcohol and tobacco consumption (see Supplementary material). The information is compiled into a clinical and demographic database and is required for stringent sample selection in downstream experimental designs.

Blood Samples and Extractions

Whole blood is collected by phlebotomy in EDTA vacutainers (23,24,28). The use of EDTA vacutainers (Becton Dickinson, Franklin Lakes, NJ) is recommended for high recovery of PBMCs and for subsequent extraction efficacy of nucleic acids and protein specimens. Blood is drawn between 9 and 11 AM, to minimize circadian variation in samples, and processed within 2 hours to prevent alteration or degradation. For each sample, an identification number is established; information on the date, times of bleed and processing, volume of whole blood and plasma, plasma color, and turbidity are recorded in the bio-repository database. Although the participants are not asked to fast before giving blood, information on what they had for breakfast is recorded.

About 32 mL of whole blood is recommended and must be processed at room temperature to avoid cell leaching. Briefly, 2 mL of sterile whole blood are first aliquoted with 10% dimethyl sulfoxide (DMSO) and stored at −139°C for future cell culture analysis (22). Blood containing DMSO for establishment of lymphoblast cultures is stable for at least 2 years. The remaining 30 mL of blood are separated into plasma and cells by centrifugation at 500 g for 10 minutes. Plasma is transferred to a tube on ice, and a protease inhibitor cocktail tablet (Complet, EDTA-free; Roche diagnostics, Indianapolis, IN) is allowed to dissolve.

The PBMCs are immediately isolated on Ficoll-Paque Plus (GE Healthcare, Piscataway, NJ) following manufacturer’s instructions. The PBMCs are then lysed in Trizol (Invitrogen, Carlsbad, CA) at a ratio of 2/15 mL of initial whole-blood volume processed. The remaining red blood cells (RBCs) are then washed three times with phosphate-buffered saline (PBS;1×) to remove the Ficoll-Paque Plus. Half of the RBCs is aliquoted to microcentrifuge tubes, whereas the other half is lysed in three parts of Trizol. All samples are stored at −80°C until further processing.

Trizol samples of PBMCs are thawed at room temperature (5 minutes), and phase separation is achieved by adding 200 µL of chloroform per milliliter of Trizol, following manufacturer’s instructions. Extractions of RNA, DNA, and protein from the three phases are then performed as described (23,24,28). Briefly, the RNA from the aqueous phase is directly purified and treated with DNase using miRNeasy columns (Qiagen, Valencia, CA), according to manufacturer’s instructions. For miRNA studies, an aliquot of the total RNA sample is taken for small RNA enrichment using commercial purification kits or by salt precipitation, as previously described (35). To prevent degradation and assure long storage life of the RNA samples, working area, equipment, and solutions must be RNase free, as recommended by the commercial source, Ambion (Applied Biosystems, Foster City, CA). Next, the interphase containing the DNA is extracted and dissolved in a 10 mM tris-Cl (pH 7.4) and 1 mM EDTA buffer (23). Finally, the proteins are extracted from the phenol phase into a denaturing sodium dodecyl sulfate (SDS) buffer by dialysis at 4°C. To ensure the stability of the protein samples, 10 mM N-2-hydroxyethylpiperazine-N’-2-ethanesulfonic acid (HEPES, pH 7.9), 0.1 mM EDTA, and 10% glycerol may be added to the final 0.1% SDS dialysis buffer.

All samples are transferred to sterile DNase-free and RNase-free microcentrifuge tubes, immediately frozen in dry ice, and stored at −80°C (or −139°C for RNA samples). Aliquots of all molecular species are also prepared for quantification and quality control, including separation by gel electrophoresis, and to avoid freeze thaw cycles. RNA quality is determined by the RNA 6000 Pico LabChip kit, using the Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA). Nucleic acid concentrations are determined spectrophotometrically by absorbance at 260 nm. Protein concentrations are measured using the DC protein assay (Bio-Rad Laboratories, Hercules, CA). The data for each molecular species are then recorded in the bio-repository database along with their sample identification number.

mRNA and miRNA Expression Profiling

Different commercial microarray platforms are available for gene and miRNA expression profiling, and the methods vary according to the technology chosen. For our example, we analyzed mRNA expression using the NIA Human MGC complementary DNA (cDNA) microarray, which contains approximately 6,424 unique full-length cDNA genes (36). Briefly, mRNA is reverse transcribed into radioactively labeled cDNA using α32P-2’-deoxycytidine 5’-triphosphate, (α32P-dCTP) (12,28). The transcripts hybridized to the chip are then quantified by densitometry using Array-Pro Analyzer 4.5 software (Media Cybernetics, MD). Our NEC and Alzheimer PBMC gene expression data sets are available at the Gene Expression Omnibus (GEO) repository, under GEO accession series GSE4226, GSE4227, GSE4229, and platform GPL1211 (www.ncbi.nlm.nih.gov/).
We also analyzed miRNA expression using the custom-made human MMChip (37), which at the time, consisted of 462 human antisense DNA sequences of mature miRNAs obtained from miRBase; http://microrna.sanger.ac.uk/, version 7.0. Enriched small RNA samples are labeled with digoxigenin (DIG) at the 3‘ end, using the DIG Oligonucleotide Tailing Kit, 2nd Generation (Roche Diagnostics), following manufacturer’s instructions. The miRNA hybridized to the chip are then revealed by DIG immunodetection and quantified by densitometry using Array-Pro Analyzer 4.5 (27).

Array-Pro Analyzer 4.5 (Media Cybernetics, Bethesda, MD) is used to normalize the data of individual chips for their subsequent comparison (data mining). The data may also be transformed to the logarithm base 2 or normalized using a reference gene; reviewed in (38). It is recommended that untransformed data be analyzed first as this yields results that reflect true variation. There are several commercial software packages for microarray data mining. Here, we describe the use of excellent freely available software.

Our untransformed microarray data were analyzed using significance analysis of microarrays (SAM) and prediction analysis for microarrays (PAM) software (Stanford University, CA), as described (39,40). SAM provides significant gene lists by T-statistics comparisons between two groups. Other SAM features, such as principle component analysis, allow the comparisons of more than two diagnostic groups or variables. The feature “Block” (block effect) in SAM is used to account for variations due to experimental batches as well as demographic variables such as gender. Specific genes with significant altered expression between AD and NEC counterparts can also be identified as predictive clinical biomarkers using PAM (see Results and Discussion section).

Additional Kolmogorov–Smirnov statistics generated by gene set enrichment analysis (GSEA) software (41) provides global analysis of the microarray data and produces heat maps of enriched gene expression. Finally, higher order analysis by hierarchical clustering analysis is performed using GenePattern, a software that provides more than 100 additional analytical tools for data mining (42).

**Microarray Validation by Quantitative Real-Time Polymerase Chain Reaction**

A few genes with significantly altered expression in the microarray data are selected for validation using a complementary platform such as relative quantitative real-time polymerase chain reaction, as described (12,28). Beta-actin is generally used as the reference gene for gene expression validation (28). For miRNAs expression, the U24 small nucleolar RNA or 5S ribosomal RNA are used as reference genes (27).

Fold changes are estimated by the \( Ct \) comparative method, using the average cycle-threshold (\( Ct \)) values of the validated gene after subtraction by the average \( Ct \) value of the reference gene, from at least three or more individuals of each gender in each diagnostic group (12,27,28). The estimated fold changes calculated by the delta \( Ct \) method do not have to be of the same order of magnitude as those estimated by the microarray platform (43); however, their trend (ie, up- or downregulation) between groups compared should be the same for validation.

**Proteomic Profiling**

For proteomic profiling, protein samples from each diagnostic group may be compared by two-dimensional (2D) gel electrophoresis. Samples are either analyzed separately or pooled within each diagnostic group (44). For 2D gel comparisons, the detection of protein spot modifications between participants requires additional 2D gel analysis software. The protein spots are then analyzed by mass spectrometry, such as matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF), and may be identified by using freely available software, (i.e. ProFound; http://prowl.rockefeller.edu (45)).

Alternatively, the peptides of individual samples are quantified by differential labeling using iTRAQ technology (Applied Biosystems), coupled with a liquid chromatography–mass spectrometry (LC-MS) system (QSTAR XL, Applied Biosystems). The resulting mass spectrometry and iTRAQ data are used to identify and quantify proteins with the assistance of software such as ProQUANT; Applied Biosystems (46). The LC-MS peptide intensity data can also be analyzed by unsupervised clustering methods such as global proteomics, as described (47).

As mentioned earlier, SAM, PAM, and GenePattern software can also be applied to the proteomics data. In all the previous proteomic approaches, the lead protein biomarkers should be validated by immunoblot assays. The levels of protein are revealed by a colorimetric assay and quantified by densitometry, and the results normalized using constitutively expressed proteins, that is, beta-actin (46). Fold changes estimated by iTRAQ, for example, do not have to be of the same order of magnitude as those obtained by immunoblot, especially if the latter are normalized with the levels of a reference protein.

**Genotyping**

For AD research, APOE alleles e2, 3, and 4 should be determined for all DNA samples, as described (12,48). Known inheritable autosomal dominant mutations are determined in research concerned with familial AD. In sporadic AD, other SNP and mutations as well as their haplotypes in various genes (49) are determined according to objectives of the research. There are several commercial genotyping array platforms available. In our case–control study, we determined the rs7483 SNP in GSTM3 in Caucasians by the single-strand conformation polymorphism method (12).
Ethnicity, APOEε4 status, as well as other demographic variables such as gender may be used for population stratifications. In fact, the odds ratio of developing AD for a particular allele may differ between genders or between noncarriers and carriers of other alleles, such as APOE4. Importantly, alleles or mutations should be analyzed within the same ethnic group because their frequencies vary in different populations.

Statistics and Data Mining

The scores of the various cognitive tests (ie, MMSE) may be correlated with the levels of the lead biomarker using regression analysis, as described (50). Moreover, the genomic and proteomic data sets can be analyzed based on various demographic variables (ie, gender, APOE4 or SNP status, medication), using SAM and GSEA software, as well as regression or principle component analyses.

Disease classification using proteomic and genomic data sets is determined by PAM or receiver-operator characteristic (ROC) analyses. Differences in demographic variables between the diagnostic groups are determined by one-way analysis of variance (ANOVA), using a multiple comparison post hoc test. One- and two-way ANOVA, Student’s t test, chi-square test, Fisher exact test, odds ratio and 95% confidence interval, regressions, correlations, survival, and ROC curves can be easily determined using a software such as GraphPad Prism version 5 (GraphPad Software, San Diego, CA).

Functional attribution to the genomic and proteomic data sets can be performed using the publicly available Database for Annotation, Visualization and Integrated Discovery (DAVID) (51). For proteomics data, the protein symbol identification is translated to gene symbol for facilitation of data mining. Other gene identification systems can be translated using Web-based database converters such as MatchMiner (http://discover.nci.nih.gov/matchminer/). In-depth information on individual genes is available in Web databases such as SOURCE (52), which facilitates literature searches and biologic interpretation. The Gene Ontology Tree (GOTree) Machine is also used for gene pattern discovery, including the calculation of functional GO enrichments (53). Software packages such as Pathway Studio (Ariadne Inc., Rockville, MD) further assist in the discovery of underlying pathways.

For miRNA research, detailed information is available at the miRBase registry (54). miRNA target predictions are also freely obtained from several algorithms, including PicTar (55), miRanda (54), TargetScan (56), miRGen (57), MicroInspector (58), and miRNApath (59). Functional attribution for miRNAs is based in part on the functional enrichment of the inverse correlation of predicted miRNA targets with altered gene or protein expression in AD. The use of the GOTree Machine, DAVID, or Pathway Studio enables further pathway analysis.

Results and Discussion

Elaboration of an AD Bio-Repository

We describe the methodology for the development of a blood-based bio-repository from an elderly population for AD research. Patients with amnestic MCI and probable sporadic AD, as well as NEC participants, are invited to volunteer in a research study. Individuals who contributed to the blood bio-repository are recontacted for subsequent blood sampling for a period of 5 years or more for longitudinal and disease progression studies. The medical research staff assists in the recruitment and in filling out the health questionnaire (Supplementary material). This questionnaire is designed to address the potentially confounding factors in AD research in a time-efficient fashion in a clinical setting.

As proof of principle, we have initiated the task of establishing a prototype blood bio-repository in collaboration with the JGH–McGill University Memory Clinic in Montreal. At present, the recruitment at the clinic is based on volunteer participation, as well as using widespread publicity in the hospital and community. The demographics of our blood bio-repository at this time are summarized in Table 1. As expected, the average MMSE scores are significantly different between the three diagnostic groups. Average years of education in the NEC group are significantly higher than those in the MCI and AD groups (Table 1). Lower educational level is a risk factor for sporadic AD (60), and generally NEC recruited have additional postsecondary education.

Next, we describe the methodology for the isolation of various blood constituents, including plasma, RBCs, and PBMCs. Following the procedure previously described (23,24,28), we report the simultaneous extraction of RNA, DNA, and protein from PBMCs of NEC, MCI, and AD (Table 2). The same amount of DNA per volume of whole blood was extracted from PBMCs in each diagnostic group. Although a lower amount of RNA was extracted in AD blood, the amount was sufficient for gene expression analysis. We obtained similar amounts of small RNA from total RNA from NEC and AD participants (Table 3) using a differential precipitation method for miRNA enrichment (27,35).

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>n</th>
<th>Women (%)</th>
<th>Age †</th>
<th>Education †</th>
<th>MMSE † (score/30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEC</td>
<td>55</td>
<td>69.6</td>
<td>73.9±7</td>
<td>14.3±4</td>
<td>29.2±1</td>
</tr>
<tr>
<td>MCI</td>
<td>118</td>
<td>54.2</td>
<td>75.7±8</td>
<td>11.3±4**</td>
<td>27.8±2*</td>
</tr>
<tr>
<td>AD</td>
<td>81</td>
<td>60.5</td>
<td>78.9±7</td>
<td>11.7±4***</td>
<td>23.4±4***</td>
</tr>
</tbody>
</table>

Notes: AD = Alzheimer’s disease; MCI = mild cognitive impairment; MMSE = Mini-Mental State Examination; NEC = normal elderly controls.

† Mean and standard deviation, with differences between groups compared by the Newman–Keuls multiple comparison post hoc test. *p < .05; **p < .001 vs NEC.
Importantly, RNA integrity among NEC, MCI, and AD samples did not significantly differ, with an average 28S/18S ratio of 1.9 ± 0.1 and RNA integrity number (RIN) (61) of 9.3 ± 0.5. The stability of the samples may not last indefinitely, and their integrity should be verified periodically. Aliquots of RNA and protein samples should be prepared to avoid freeze thaw cycles, especially for longitudinal studies. We evaluated the stability of our RNA samples kept at −139°C over a period of 2 years and obtained stable RIN values in the three diagnostic groups. Precaution in sample preparation and handling as described in the Methodology section is crucial in order to avoid RNA degradation.

The amount of DNA is indicative of PBMC numbers and suggests that there is no significant difference in cell number between the three diagnostic groups (Table 2). Nevertheless, it is highly recommended that a blood cell count be made for all blood samples and the information included in the database because different cell types in the PBMC fraction may decrease with age or vary between NEC and AD (62). Moreover, despite the similar apparent quality in the three molecular species extracted, specific disease-associated oxidative damage may exist in MCI and AD samples. In fact, oxidative damage has been documented in RNA (63,64) and DNA of peripheral lymphocytes in MCI and AD (65,66), as well as in plasma proteins (44,67) relative to NEC.

Table 2. Yields of Protein, RNA, and DNA From PBMCs

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>n</th>
<th>DNA (µg/mL WB)</th>
<th>RNA (µg/mL WB)</th>
<th>Protein (µg/mL WB)</th>
<th>RNA/DNA</th>
<th>Protein/DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>38</td>
<td>12.5 ± 7</td>
<td>2.5 ± 1</td>
<td>212.3 ± 76</td>
<td>0.22 ± 0.1</td>
<td>19.7 ± 10</td>
</tr>
<tr>
<td>Men</td>
<td>17</td>
<td>10.5 ± 6</td>
<td>2.3 ± 1</td>
<td>208.7 ± 88</td>
<td>0.28 ± 0.2</td>
<td>26.2 ± 17</td>
</tr>
<tr>
<td>Total</td>
<td>55</td>
<td>12.1 ± 7</td>
<td>2.5 ± 1</td>
<td>210.6 ± 78</td>
<td>0.24 ± 0.1</td>
<td>21.5 ± 12</td>
</tr>
<tr>
<td>MCI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>63</td>
<td>9.6 ± 5</td>
<td>2.2 ± 1</td>
<td>206.6 ± 93</td>
<td>0.24 ± 0.1</td>
<td>22.5 ± 11</td>
</tr>
<tr>
<td>Men</td>
<td>54</td>
<td>11.3 ± 8</td>
<td>2.2 ± 1</td>
<td>171.8 ± 63</td>
<td>0.27 ± 0.2</td>
<td>19.4 ± 10</td>
</tr>
<tr>
<td>Total</td>
<td>117</td>
<td>11.2 ± 7</td>
<td>2.2 ± 1</td>
<td>190.2 ± 81</td>
<td>0.25 ± 0.2</td>
<td>21.1 ± 11</td>
</tr>
<tr>
<td>AD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>48</td>
<td>10.3 ± 6</td>
<td>2.0 ± 1*</td>
<td>196.5 ± 67</td>
<td>0.26 ± 0.3</td>
<td>23.6 ± 17</td>
</tr>
<tr>
<td>Men</td>
<td>31</td>
<td>8.7 ± 5</td>
<td>2.1 ± 1</td>
<td>199.6 ± 75</td>
<td>0.28 ± 0.2</td>
<td>26.4 ± 15</td>
</tr>
<tr>
<td>Total</td>
<td>79</td>
<td>10.2 ± 2</td>
<td>2.0 ± 1*</td>
<td>197.1 ± 69</td>
<td>0.27 ± 0.2</td>
<td>24.7 ± 16</td>
</tr>
</tbody>
</table>

Notes: AD = Alzheimer’s disease; MCI = mild cognitive impairment; NEC = normal elderly controls; PBMC = peripheral blood mononuclear cell.

1Expressed per milliliter of whole blood (WB) processed; mean and standard deviation, with differences between groups compared by the Newman–Keuls multiple comparison post hoc test. *p < .05 vs NEC.

2Ratio of total protein or RNA over total DNA.

DNA of peripheral lymphocytes in MCI and AD (65,66), as well as in plasma proteins (44,67) relative to NEC.

Here, the expected disease-related oxidative damage in our three PBMC constituents did not adversely affect their yields from MCI and AD samples. Nevertheless, oxidative damage may affect synthesis and stability of genes and proteins. This may explain in part the commonly reported downregulation of various biomarkers in Alzheimer brain and peripheral tissues, including previous results on protein and mRNA expression (21,28,68,69).

Experimental Design

For AD genomic and proteomic research, it is necessary to control for confounding and risk factors that are intrinsic to the disease. The elaboration of a detailed demographic database, as described previously, enables stringent sample selection within the bio-repository. Initial comparisons between NEC and MCI or AD molecular samples should be matched for gender, ethnicity, age, level of education, and health status. For example, participants with chronic metabolic and inflammatory conditions (ie, diabetes, rheumatoid arthritis, chronic active hepatitis), as well as those with symptoms of depression, high alcohol consumption, or tobacco usage, should be excluded. These criteria necessitate lengthy recruitment times to obtain ideal samples from participants in an elderly population that is generally heavily medicated.

The study should be designed for either familial or sporadic AD. Participants with possible AD family history should not be mixed with those without. In fact, there exists an unknown proportion of late-onset AD cases that harbor “less aggressive” familial autosomal dominant mutations as opposed to the known highly penetrant mutations in early-onset familial AD (2). Moreover, the presence of allele risk factors such as APOEε4 should be carefully balanced in the selection of samples between diagnostic groups in order to determine their impact on gene and protein expressions.
Also, significant differences in gender are observed in AD (12,28), and therefore, male and female participants should be analyzed separately or at least equally represented in each diagnostic group if they are to be analyzed together.

Our initial investigations on genomic profiling of PBMCs concerned the common late-onset form of sporadic AD (12,27,28). Using the data collected in the demographic database, we first selected age-matched septuagenarian NEC and AD participants without first-order AD family history. Our second criterion for individuals selected is those participants without chronic diseases including cancer, cardiovascular disorders, etc. Additional exclusion criteria used to disqualify participants were tobacco and morning alcohol consumption, depression, head injury, and low level of education. Finally, we balanced the number of men and women in the two diagnostic groups (AD vs NEC) that we initially compared (Figure 2). We envision future global proteomic profiling studies using the protein samples from the same biologic specimens and experimental designs.

In general, a minimum of seven biologic repeats per diagnostic group (individual male or female NEC and AD participants) is recommended for a pilot study in genomics, depending on the number of genes analyzed and the level of false discovery rate tolerated. We determined the sample size using publicly available bioinformatics software (ie, http://bioinformatics.mdanderson.org/MicroarraySample-Size/). The pooling of samples within each diagnostic group is less desirable, for it masks existing variation between participants and may increase false-positive rates.

Another source of variation may arise from the diversity of cell types that comprise the PBMC fraction (70). The proportion of lymphocytes, monocytes, granulocytes, and thrombocytes varies between samples or diagnostic groups and may obscure results, for example, in the study of cell-specific miRNA expression (27). These different sources of variation may result in weaker statistics or underestimation of actual fold differences between diagnostic groups. Notwithstanding these sources of variation, significant systemic manifestations of AD were identified using PBMCs.

An alternative experimental design is treating microarray expression as a quantitative trait, like neurological phenotypes or genotypes (71), for example, the risk factors APOEe4 (27,72) or SNP (12) in AD. In the latter study, DNA obtained from PBMCs was used to demonstrate a GSTM3 polymorphism association with MCI and AD, in the APOEe4-negative stratum and in women. Similarly, to the approach shown in Figure 2, differences in gene expression between noncarriers and carriers of the SNP were made within each diagnostic group and then compared between NEC and AD (12). Interestingly, the GSTM3 polymorphism was associated with a hub of transcriptional networks that may coordinate antioxidant defense with DNA repair mechanisms. This experimental design can be applied to other known risk alleles (49) in order to delineate discrete expression networks associated with other risk alleles.

### Experimental Design

<table>
<thead>
<tr>
<th>Normal Elderly Controls (NEC)</th>
<th>Alzheimer disease (AD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-8 women</td>
<td>7-8 women</td>
</tr>
<tr>
<td>7-8 men</td>
<td>7-8 men</td>
</tr>
</tbody>
</table>

1. AD vs NEC (all data set)
2. AD vs NEC women
3. AD vs NEC men

**Common expression**

Figure 2. Experimental design for expression profiling in Alzheimer’s research. Stringently selected samples matched for various demographic variables as described in the text are compared between two diagnostic groups in a case–control study. Biologic replicates of equal numbers of male and female samples in each diagnostic group are analyzed separately or together, as depicted. A minimal sample size of seven to eight individual female or male samples per diagnostic group is recommended. AD = Alzheimer’s disease; NEC = normal elderly controls.

### Integrated Genomic and Proteomic Platform

The approach we propose for the investigation of systemic dysfunctions in AD is the simultaneous analysis of miRNA, mRNA, and protein levels in blood cells and plasma (Figure 1). The results are integrated with at-risk alleles and miRNA target predictions, using various publicly available algorithms such as miRanda and TargetScan (73). miRNAs target mostly the 3’ untranslated region of mRNAs, and although their mode of action is diverse, they generally induce mRNA degradation or inhibit protein synthesis. Their expression is inversely correlated with the levels of their target mRNA and proteins, and together this may provide an understanding of etiopathogenetic mechanisms at play in AD.

As an example using PBMCs, we inversely correlated target predictions with miRNA and mRNA levels expressed in AD. Interestingly, we mainly detected the upregulation of miRNAs in Alzheimer PBMCs (27). This is consistent with the observation in several reports of various genes and proteins being downregulated in the peripheral blood of AD patients. The functional attributions for miRNAs identified in Alzheimer PBMCs are summarized in Figure 3. The pathways are denoted by the functional categories (of altered gene expression) and where miRNA deregulation may be correlated. This analysis is based on the identical genes displaying similarly altered expression in AD-affected brains (74). The results point to miRNA-mediated deregulation in transcriptional networks concerned with antioxidant defense, DNA repair mechanisms, and cytoskeletal and mitochondrial maintenance (Figure 3).

The implication of lead miRNAs identified in Alzheimer PBMCs with similarly altered levels in AD-affected brains remains to be determined, as reviewed in (26). Ultimately, the miRNA biomarkers and associated gene and protein dysfunctions should be validated in cell-based assays.
Consistent with our findings, however, upregulated miRNAs are reported in AD-affected brains (75,76). Downregulation of key miRNAs associated with Aβ accumulation is also reported in AD brains (77,78). Hence, the importance of miRNA deregulation in neurological diseases such as AD is paramount and may be associated with intrinsic mechanisms of aging (46,79).

In summary, we described the methodology for the establishment of a blood-based bio-repository for AD research, the parsimonious use of molecular specimens, and the need for an information-rich database for the genetic and molecular characterization of NEC, MCI, and AD participants (Figure 1). As previously reported for plasma proteins using PAM analysis (21), PBMC gene expression may also prove useful as predictive markers of sporadic AD. In fact, we obtained 90% accuracy for correct AD classification by PAM analysis (ie, 36 of 40; four NEC classified as AD) using PBMC gene expression (Table 4). Investigation of the contribution of these genes to the pathology of AD and their usefulness as prognostic biomarkers is currently underway in longitudinal studies of the MCI group at the McGill–JGH Memory Clinic. The sensitivities and specificities of any prognosticating biomarkers identified in initial surveys will then need to be validated using additional sets of blinded disease and control samples (Supplementary Table S1).

Our investigations using PBMCs hold promise for the advent of reliable peripheral biomarkers to facilitate diagnosis of AD and prognosis of individuals with MCI. Successful outcomes may satisfy an urgent unmet clinical need regarding the management of memory dysfunction in a rapidly aging population. Advances in this area should allow a more complete elucidation of the systemic manifestations of AD and provide potential surrogate biomarkers and leads for effective therapeutic interventions for AD and related dementias.

**Supplementary material.**

Supplementary material can be found at: http://biomed.gerontologyjournals.org/
Funding

Sir Mortimer B. Davis-Jewish General Hospital; Kentucky “Bucks-for-Brains” and Gheens Foundation Endowment funds to E.W.

Acknowledgments

The authors thank all the volunteering participants in our research and Shelley Solomon and Adrienne Liberman for their excellent assistance in recruitment. The authors gratefully acknowledge the contribution of the medical staff, neuropsychologists, nurses, patients, research coordinators, and secretaries of the JGH–McGill Memory Clinic. None of the authors had any disclosure concerning any actual or potential conflicts of interest.

Correspondence

Address correspondence to Eugenia Wang, PhD, Gheens Center on Aging, University of Louisville, 580 South Preston Street, Louisville, KY 40202. Email: eugenia.wang@louisville.edu

References


Received August 5, 2008
Accepted February 25, 2009
Decision Editor: Huber R. Warner, PhD