A Decade of Blood Biomarkers for Alzheimer’s Disease Research: An Evolving Field, Improving Study Designs, and the Challenge of Replication

Liu Shi, Alison L. Baird, Sarah Westwood, Abdul Hye, Richard Dobson, Madhav Thambisetty and Simon Lovestone

Department of Psychiatry, University of Oxford, Oxford, UK
King’s College London, Institute of Psychiatry, Psychology and Neuroscience, and NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at South London and Maudsley NHS Foundation, London, UK
Unit of Clinical and Translational Neuroscience, Laboratory of Behavioral Neuroscience, National Institute on Aging, Baltimore, MD, USA

Accepted 11 September 2017

Abstract. Blood-based biomarkers represent a less invasive and potentially cheaper approach for aiding Alzheimer’s disease (AD) detection compared with cerebrospinal fluid and some neuroimaging biomarkers. Acknowledging that many in the field have made great progress, here we review some of the work that our group has pursued to identify and validate blood-based proteomic biomarkers through both case control and AD pathology endophenotype-based approaches. Our focus is primarily to identify a minimally invasive and hopefully cost-effective blood-based biomarker to reduce screen failure in clinical trials where participants have prodromal or even pre-clinical disease. We summarize some of the key findings and approaches taken in these biomarker studies, while addressing the main challenges, including that of limited replication in the field, and discuss opportunities for biomarker development.

Keywords: Alzheimer’s disease, blood proteomic biomarkers, endophenotype, replication, validation

INTRODUCTION

The health and economic impact of Alzheimer’s disease (AD) is enormous and with a projected increase in numbers of people living with dementia from over 75 million today to over 135 million by 2050. It is obvious why finding a therapeutic intervention has become a major strategic goal for science in many countries. However, despite substantial advances in our understanding of mechanisms of disease and progress in pre-clinical and early drug discovery, a series of major failures at phase III has dented confidence. The scientific field of neurodegeneration research has become confused; is it our understanding of disease that is wrong or our development of drugs that is inadequate? Another view is that despite the obvious need for widening our development of therapeutic approaches, something that was...
biomarkers of Alzheimer’s Disease, we focus on our findings, describe our evolving approach, and address some of the main challenges of the field, along with the suggestions to improve the reproducibility of biomarker development; acknowledging of course that this work is just one part of a rapidly evolving body of work from laboratories around the world seeking biomarkers in readily accessible tissues for clinical trials and for clinical practice utility.

EARLY DISCOVERY OF BLOOD-BASED BIOMARKERS OF ALZHEIMER’S DISEASE

It is important to note that there is no blood-based biomarker that has got beyond first base, the discovery phase. It remains entirely possible that no biomarker from blood will ever progress beyond rigorous replication studies and enter the even more taxing qualification and validation phases. Indeed, when we initiated our first study using proteomics in blood [15], our starting point was to address the null hypothesis—that there would be no protein signal or signature in blood that reflected disease process. After all, why would there be such a signature when a supposedly impenetrable barrier separates brain and blood? However, in this first study, we were able to reject the null hypothesis and this, together with the work of Wyss-Coray et al. published soon after [16], persuaded us that there is a signal in blood; but the question remains, can that signal be translated to a replicable and useful biomarker?

Case-Control Approach

Like almost all other biomarker studies, we started our work using a case-control study design, and this remains the predominant study type in the literature, although as we describe below, we are increasingly moving to other approaches. In Table 1, we summarize the main case-control studies conducted within our group. In our first study [15], we used two-dimensional gel electrophoresis (2DGE) coupled with mass spectrometry (MS) comparing people with established disease to a group of healthy elderly. For the discovery stage, we tested 100 subjects and we found nine proteins showing an increase and four a decrease in AD. Moreover, we were able to identify cases from controls with a sensitivity of 56% and a specificity of 80% based on 2DGE image analysis alone, using machine learning, making a reasonably powerful case that a signature of disease was present...
in blood. We then replicated two of the predominant proteins contributing to this signature including complement factor H (CFH) and α-2-macroglobulin (α2M) in reasonably large independent cohorts (511 subjects). The combination of CFH and α2M gave a sensitivity of 62% and a specificity of 60% to discriminate AD from non-AD controls. To the best of our knowledge, this is the first large study to use an agnostic proteomics approach to seek biomarkers of AD in peripheral tissue.

Ten years later, both CFH and α2M remain of considerable interest, both as biomarkers and as possible participants in disease process. When we published these findings, we noted that the gene encoding CFH was strongly associated with age-related macular degeneration, the only other Aβ-associated disease, thus suggesting that complement biology might be a critical factor in AD pathogenesis. This has been amply substantiated by subsequent genome-wide association studies where some of the more strongly associated susceptibility genes are complement factors and where pathway analysis identifies complement as one of only three or four biological processes in disease etiology [17–19]. The finding of altered levels of CFH in blood in people with AD was replicated in some but not other independent studies but was also, intriguingly, found in blood of transgenic mouse models of amyloidopathy [20–23]. The finding in mice [21] is particularly interesting as one possible confound of the human case-control studies is that people with AD are systemically unwell, have altered diet and a multiplicity of other environmental changes that could in themselves alter the blood proteome. None of these is relevant to a mouse model where the only disease process is confined to a relatively small part of the brain and the mice, disappointingly for those intending to develop complete models of disease, remain in robust health. If there is a change in blood from these animals, and especially one that as in the case of CFH is measurable very early in the model development (3–6 months) then the signal must have been induced in brain and transferred to blood. How this occurs should be a matter of very considerable interest. These findings in mice suggest a role in the disease process and while the mechanism underlying such a role is not known, we find that in neurons from mice lacking their equivalent of the AD susceptibility gene CR1, the levels of CFH and phosphorylated tau are significantly, and substantially, decreased [24]. Similarly, α2M, identified in this first proteomic study, has been largely replicated [23, 25–27], associated with other indicators of AD [28] and altered in blood from APP/PS1 transgenic mice [21]. The recently reported associations between serum α2M levels and tau phosphorylation states in the brain as well as its potential role as a sex-specific inflammatory marker in preclinical AD have been especially interesting [29].

Contrary to our expectations then, this study using agnostic proteomics not only suggested a signal in blood in AD but identified components of that signal that have been widely, albeit not unequivocally replicated and powerfully linked to etiopathogenesis through genomics, through empirical studies in vitro, and through in vivo models. Rather than abandoning blood-based biomarkers and turning to CSF, we were instead emboldened to advance these studies and began to do so using other, more advanced, untargeted proteomic technologies. For example, using isobaric tags (tandem mass tags) labelling to enable multiplexing mass spectrometry (LC-MS/MS), we identified gelsolin as a possible marker of AD and then replicated the finding in independent samples using western blot analysis [30]. Plasma gelsolin was associated with disease severity, although its accuracy to discriminate AD from non-demented controls was low (area under the curve, AUC = 0.64). Others have also found association between levels of lower levels of gelsolin and AD using both an isobaric tag/MS, untargeted approach [31] and targeted ELISA [32].

Mass spectrometry based discovery tends to identify predominantly highly abundant proteins, even when combined with depletion strategies, resulting in putative biomarkers such as many we discuss in this article, including gelsolin, being associated with multiple unrelated diseases. As an alternative high dimensionality detection technology not biased toward abundant proteins, we turned to the aptamer capture, SOMAscan array, developed by SomaLogic to test 1001 proteins in blood samples from 691 subjects [33]. We found that prostate-specific antigen complexed to α1-antichymotrypsin and pancreatic prohormone were significantly associated with AD diagnosis. Furthermore, a panel of 13 proteins (placenta growth factor, Interleukin-17, Fibronectin FN1.4, Fibronectin, Secretory leukocyte protease, Fibronectin FN1.3, Epithelial cell kinase, Prolactin, C-C motif chemokine, Calcium/calmodulin-dependent protein kinase type II subunit α, Seprase, Pancreatic prohormone and Coagulation factor XI) predicted AD with an AUC of 0.70 with several of these analytes replicating previous studies, including, for example, pancreatic prohormone [34] and prolactin [35].
### Table 1
List of case-control studies of blood-based protein biomarkers of AD conducted by our group

<table>
<thead>
<tr>
<th>Reference</th>
<th>Cohorts</th>
<th>Markers</th>
<th>Methods</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hye et al., [15]</td>
<td>Discovery: ART cohort: 50 AD, 50 CTL</td>
<td>9 proteins increase and 4 decrease in AD</td>
<td>2DGE LC-MS/MS</td>
<td>Sensitivity = 56%, specificity = 80%</td>
</tr>
<tr>
<td></td>
<td>Replication: ART, MND and Institute of Neurology Huntington’s disease study cohorts: 111 AD and 400 non-AD controls</td>
<td>CFH and α2M</td>
<td>Western blot</td>
<td>Sensitivity = 62%, specificity = 60%</td>
</tr>
<tr>
<td>Guntert et al., [30]</td>
<td>Discovery: ART cohort: 30 AD and 15 CTL</td>
<td>gelsolin and C1 inhibitor protein gelsolin</td>
<td>TMT/MS</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>Replication: ART cohort: 60 AD and 35 CTL</td>
<td></td>
<td>Western blot</td>
<td>Sensitivity = 39%, specificity = 80%, AUC = 0.63</td>
</tr>
<tr>
<td>Sattlecker et al., [33]</td>
<td>DCR and AddNeuroMed cohorts: 331 AD, 211 CTL and 149 MCI</td>
<td>1001 proteins probed, 13 proteins identified</td>
<td>SOMAscan</td>
<td>Sensitivity = 0.67, specificity = 0.64, AUC = 0.70</td>
</tr>
<tr>
<td>Kiddle et al., [45]</td>
<td>ART, DCR and AddNeuroMed cohorts: 286 AD, 182 MCI and 209 CTL</td>
<td>96 proteins probed, 13 proteins identified</td>
<td>SOMAscan</td>
<td>Sensitivity = 0.83, specificity = 0.66</td>
</tr>
<tr>
<td>Greco et al., [37]</td>
<td>AddNeuroMed cohort: 78 AD, 80 MCI and 82 CTL</td>
<td>25 proteins nominated by text mining and Intelligence Network analytics from ‘all’ biological datasets</td>
<td>In silico nomination and in vitro verification by Western blot</td>
<td>PLAUR association with disease $p &lt; 0.001$ and ChAt association with brain atrophy $p &lt; 0.01$</td>
</tr>
<tr>
<td>Hakobyan et al., [36]</td>
<td>AddNeuroMed and DCR cohorts: 106 AD, 186 CTL and 189 MCI</td>
<td>5 complement proteins and 4 activation products</td>
<td>MSD platform</td>
<td>-combination of clusterin and ApoE status discriminate AD from controls with an AUC of 0.78; -combination of clusterin, factor I and terminal complement complex predict MCI conversion to AD with an AUC of 0.85</td>
</tr>
</tbody>
</table>

AD, Alzheimer’s disease; ART, Alzheimer’s Research Trust; DCR, Maudsley and King’s Healthcare Partners Dementia Case Register; MND, KCL motor neuron disease study; 2DGE, 2D gel electrophoresis; LC-MS/MS, liquid chromatography tandem mass spectrometry; CFH, complement factor H; α2M, α2-macroglobulin; CTL, cognitively healthy elderly controls; TMT, Tandem Mass Tags; AUC, area under the curve; MCI, mild cognitive impairment; PLAUR, Urokinase-type plasminogen activator receptor; ChAt, choline acetyltransferase; MSD, mesoscale discovery.

The studies we describe in Table 1 all use essentially agnostic methods to discover novel biomarkers—proteomic technologies not predicated on any a priori hypotheses. Sometimes called unbiased proteomics, the analogy with truly unbiased genomics is not quite right as proteomic technologies are some way off from being able to measure all proteins in complex matrices such as blood and even further away from measuring all characteristics of proteins including post-translational modification. Moreover, the current technologies are all, in some way, biased; MS toward highly abundant proteins and the SOMAscan technology toward proteins in certain classes such as inflammation. The relative inattention paid to post-translational modification in particular is likely to be important as we discuss in the case of clusterin below. An alternative approach to using these agnostic technologies is to take a targeted approach, measuring proteins nominated because of a possible role in disease pathogenesis or some other reason. As an example, we recently performed a systematic study of proteins involved in the complement pathway, measuring five complement proteins and four activation products in AD, MCI, and controls [36]. Overall, we tested 471 subjects and results showed that only one complement protein (clusterin) differed significantly between control and AD groups but that the combination of clusterin and ApoE status was
able to discriminate AD from controls with an AUC of 0.78. Moreover, along with clusterin, two other markers (factor I and terminal complement complex) were significantly different between MCI individuals who had converted to dementia one year later compared to non-converters. The combination of the three biomarkers was able to predict MCI conversion to AD with an AUC of 0.85.

In another radically different approach to targeted biomarker discovery, we employed true Big Data analytics to nominate proteins as putative markers [37]. We used text mining to interrogate the vast publically available literature and other datasets, including structured and unstructured databases such as Medline, AD research forums, gene expression, and proteomics and genomics databases, to establish an Intelligence Network based on a shared vocabulary or terms. We then interrogated this Intelligence Network using pre-defined axioms. These were essentially descriptors of what an ideal biomarker might look like; for example, ‘a protein involved in AD pathogenesis that is expressed in regions of the brain vulnerable to AD and is not normally present in plasma’. With this workflow, we identified a set of putative biomarkers, interestingly, some of which such as clusterin and transthyretin (TTR) were independently nominated as biomarkers of AD through agnostic proteomics. As a proof of concept, we chose to explore further two entirely unexpected and novel proteins (PLAUR and ChA1) and in a large series of subjects found the former to be strongly associated with AD and the latter to have associations with brain volume on imaging. This study perhaps signposts a truly innovative and exciting approach to targeted biomarker discovery that might be developed in future studies; exploiting as it does the vast amount of information in biological data sources and not being constrained by the limitations of protein detection technologies today.

In summary, the case-control studies we, and of course many other groups, have performed in the past decade have compellingly rejected the null hypothesis that there is no protein biomarker signature in blood. We have used a range of untargeted and targeted proteomics, and even big data informatics, to nominate protein markers and those we have identified have subsequently frequently been found to have a role in disease pathogenesis. However, it is equally clear that the size of the biomarker effect is modest at best and although some of the biomarkers we nominated have been widely replicated, this replication is not found in all such studies nor for all nominated biomarkers. Clearly, another approach was needed.

Endophenotype-based approach

Although these case-control studies were promising, there is an obvious problem in that a biomarker of established disease, although a starting point for many studies, is not the outcome we are looking for. Diagnosis of established dementia is not difficult; previously we showed both research assistants using a simple interview and automated algorithms both make highly accurate diagnoses [38]. Where diagnosis is hard is in the very early stages of disease. Indeed, in order to conduct clinical trials in preclinical disease then a biomarker identifying potential participants in the absence of symptoms is the real target for research. And yet in case-control studies, these individuals are not in the ‘case’ but in the ‘control’ group. As an alternative approach, we designed studies using ‘endophenotype discovery’. In this, subjects are allocated not to disease category but to a grouping or to a continuous variable based on some other measure of disease such as brain atrophy measured by structural MRI, rate of cognitive decline, Aβ plaque burden measured by PET, and CSF biomarkers (Aβ and tau). Many of these changes are detectable before dementia onset. Aβ plaque burden, increased levels of tau in CSF, and brain atrophy are detected 15 years before expected symptom onset while global cognitive impairment is detected 5 years before symptom onset [39, 40], suggesting that such a study design could be effective in detecting blood-based biomarkers for preclinical disease.

In the first study to use such an approach, we performed gel-based proteomics (2DGE and LC-MS/MS) in two independent groups of subjects with mild AD with analysis by the endophenotypes of cognitive decline and hippocampal atrophy [41]. We found that clusterin was associated with both endophenotypes and then showed that the level of clusterin was associated with cognition and other phenotypes in a large cohort of 689 subjects (464 AD, 115 MCI, and 110 controls) by ELISA and in a third independent cohort that clusterin levels were associated with brain amyloid burden determined by PET in cognitively unaffected individuals. Two other findings are worth noting from this work. First we showed that in transgenic mice carrying mutant human APP/PS1 genes, clusterin was increased in blood at about the same time as plaques are produced in brain, a finding identical to that recently reported by Wang et al. [21].
Secondly, the data we presented in the paper of the 2DGE results shows clusterin protein to be present in multiple different regions on the gel. This suggests different forms of protein and as clusterin is known to be highly glycosylated we subsequently developed an assay for this post-translational modification and demonstrated that a glycosylation site on the beta subunit of clusterin (beta64N) is significantly reduced in subjects with relatively more cerebral atrophy [42]. This is another example, alongside tau, where an analysis of post-translational modification may add to the value of a protein biomarker.

Soon after we completed this study, two major genome-wide association studies [18, 19] both reported that the gene encoding clusterin was the strongest association with AD after APOE, and the association between clusterin protein levels in blood and some aspect of disease was repeated by later studies from our own group [33, 43–45] as well as those from others [46, 47]. In addition to clusterin, this study [41] also found that six proteins (complement C3, γ-fibrinogen, serum albumin, complement factor-I, α-1-microglobulin, and serum amyloid-P) were associated with hippocampal atrophy and four proteins (complement component C4a, complement C8, ApoA1, and TTR) could discriminate fast from slow progressing AD groups.

**Blood-based biomarkers of brain atrophy and rate of cognitive decline (Table 2)**

Again emboldened by a striking result, we then embarked on a series of blood-based endophenotype studies, some of which are ongoing. First we replicated six of the proteins nominated in Thambisetty et al. [41], in an independent cohort of AD (n = 79), MCI (n = 88), and control (n = 95) subjects using ELISA and western blot [48]. Results showed that five (complement components C3 and C3a, complement factor-I, γ-fibrinogen, and α-1-microglobulin), together with age and sex, could explain more than 35% of variance in whole brain volume in AD patients, suggesting that these proteins are likely to be strong predictors of AD pathology. We then confirmed the association of TTR with rate of cognitive decline [49]. Briefly, we firstly tested the level of TTR by western blot in 140 subjects, and then replicated it in a larger independent cohort (270 subjects) by ELISA. Results confirmed that decreased TTR levels were associated with rapid cognitive decline and severe cognitive impairment. Furthermore, we compared these two immune-based techniques and a good correlation was observed between western blot and ELISA ($r^2 = 0.65$).

As well as using this endophenotype approach to identify novel biomarkers, we have also employed it to test previously nominated biomarkers. Using the SOMAscan array, we found that nine proteins previously nominated in case-control studies (α-1-antitrypsin, Complement C3, Pancreatic pro-hormone, Granulocyte colony-stimulating factor, Insulin-like growth factor-binding protein 2, Cluster, Complement C6, Inter-alpha-trypsin inhibitor heavy chain H4, and C-C motif chemokine 18) were also associated with at least one AD-related phenotype; clusterin with cognitive decline and seven others with brain atrophy [45]. Although promising, some of these apparent associations are in the opposite direction to that predicted from literature studies emphasizing again the problem of replication in this field. The most consistent findings though are that proteins of inflammatory cascades are altered in disease [50–53] and so we then analyzed a panel of 27 cytokines in a cohort of 351 subjects and compared them with structural MRI measures and rate of cognitive decline rate [54]. Results showed that five inflammatory proteins (IL-1ra, IL-6, IL-10, TNF-α, and IL-13) were associated with brain atrophy and six (IL-4, IL-10, G-CSF, IL-2, IFN-γ, and PDGF) were associated with fast cognitive decline within one year of follow-up.

More recently we have tested protein changes associated with rate of change and progression to dementia [55] in 235 subjects including 69 controls, 37 ‘stable’ MCI patients, 39 patients with MCI converting to AD within a year, and 90 AD patients. Results showed that twelve proteins were found to significantly associate with the rate of progression. They include Complement C2, Serum amyloid A-1 protein, Complement C9, Mannose-binding protein C, Serum amyloid P-component, α2-Antiplasmin, CHK1 (Serine/threonine-protein kinase Chk1), Interleukin-17A, Eukaryotic translation initiation factor 5A–1, Hemopexin, CDC37 (C-C motif chemokine 19), and Complement factor H-related protein 5.

Since these various studies appear to nominate a moderately large set of proteins associated with decline, whether measured directly, or indirectly through neuroimaging, we then set out to see if we could assemble a panel of proteins that would predict decline. Briefly, we measured twenty-six previously identified candidate proteins in 1,148 subjects selected from three independent centers [44].
Table 2
List of studies of blood-based protein biomarkers of AD by endophenotype approach including brain atrophy and rate of cognitive decline

<table>
<thead>
<tr>
<th>Reference</th>
<th>Cohorts</th>
<th>Markers</th>
<th>Methods</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Endophenotype: Brain atrophy measured by structural MRI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thambisetty et al., [41]</td>
<td>ART and AddNeuroMed cohorts: Discovery: 27 AD, 17 MCI, 22 AD rapid decline, 29 AD slow decline</td>
<td>Complement C3, FGG, serum albumin, CFI, α-1-microglobulin, and SAP</td>
<td>2DGE and LC-MS/MS</td>
<td>-six proteins were associated with hippocampal atrophy</td>
</tr>
<tr>
<td></td>
<td>ART and AddNeuroMed cohorts: Validation: 464 AD, 115 MCI and 110 CTL</td>
<td>Clusterin</td>
<td>ELISA</td>
<td>-clusterin was associated with brain atrophy</td>
</tr>
<tr>
<td>Thambisetty et al., [48]</td>
<td>AddNeuroMed cohort: 79 AD, 88 MCI and 95 CTL</td>
<td>Complement C3 and C3a, FGG, CFI, and α-1-microglobulin based on [41]</td>
<td>-ELISA for C3, C3a, and α-1-microglobulin; Western blot for CFI and FGG</td>
<td>-the combination of five biomarkers with age and sex could explain more than 35% of variance in whole brain volume in AD patients</td>
</tr>
<tr>
<td>Leung et al., [54]</td>
<td>AddNeuroMed cohort: 117 AD, 122 MCI and 112 CTL</td>
<td>IL-1ra, IL-6, IL-10, TNF-α, and IL-13</td>
<td>Luminex</td>
<td>-five proteins were associated with brain atrophy</td>
</tr>
<tr>
<td>Kiddle et al., 2014 [45]</td>
<td>AddNeuroMed cohort: 98 AD, 81 MCI and 95 CTL</td>
<td>96 proteins probed</td>
<td>SOMAscan</td>
<td>-α-1-antitrypsin, Complement C3, Pancreatic prohormone, Granulocyte colony-stimulating factor, Complement C6, Inter-alpha-trypsin inhibitor heavy chain H4 and C-C motif chemokine 18 were associated with brain atrophy</td>
</tr>
<tr>
<td>Sattlecker et al., [33]</td>
<td>DCR and AddNeurMed cohorts: 331 AD, 211 CTL and 149 MCI</td>
<td>1001 proteins probed</td>
<td>SOMAscan</td>
<td>-fetuin B and PPY were associated with brain atrophy</td>
</tr>
<tr>
<td>Hye et al., [44]</td>
<td>AddNeuroMed, DCR and GenADA cohorts: 476 AD, 169 MCI, 51 MCIc and 452 CTL</td>
<td>26 proteins</td>
<td>xMAP assay</td>
<td>-clusterin, RANTES, NSE, and TTR were associated with cortical atrophy in the MCI group; -α-1-antitrypsin, NSE, ApoC3, ApoA1, ApoE, and BDNF were associated with brain atrophy in AD group.</td>
</tr>
<tr>
<td><strong>Endophenotype: rate of cognitive decline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thambisetty et al., [41]</td>
<td>ART and AddNeuroMed cohorts: Discovery: 27 AD, 17 MCI, 22 AD rapid decline, 29 AD slow decline</td>
<td>Complement C4a, complement C8, ApoA1, and TTR</td>
<td>2DGE and LC-MS/MS</td>
<td>-four proteins were associated with cognitive decline</td>
</tr>
<tr>
<td></td>
<td>ART and AddNeuroMed cohorts: Validation: 464 AD, 115 MCI and 110 CTL</td>
<td>Clusterin</td>
<td>ELISA</td>
<td>-clusterin was associated with cognitive decline</td>
</tr>
</tbody>
</table>

(Continued)
We found that clusterin, Regulated on Activation, Normal T Cell Expressed and Secreted (RANTES), neuron-specific enolase (NSE), and TTR were associated with cortical atrophy in the MCI group and α-1-antitrypsin, NSE, ApoC3, ApoA1, ApoE, and brain derived neurotrophic factor were associated with brain atrophy in AD group. Furthermore, ApoE, CFH, neural cell adhesion molecule, Aβ40, A-1-acid glycoprotein, and clusterin were all correlated with cognitive decline. Among these biomarkers, most replicated our previous findings, e.g., clusterin, TTR, α-1-antitrypsin, ApoE, and CFH. We then reanalyzed the data set for predictive value and showed a panel of 10 protein biomarkers (TTR, clusterin, cystatin C, A-1-acid glycoprotein, ICAM1, complement component C4, PEDF, α-1-antitrypsin, RANTES, and ApoC3) along with APOE genotype could predict the conversion from MCI to AD with an accuracy of 87%. Clearly this finding requires replication and it is important to emphasize the predictive value applies, by definition, to individuals with MCI and not people with no symptoms. Even if replicated there would be no value of such a panel in predicting dementia in a general population sample and therefore of no
blood-based biomarkers of brain amyloid burden measured by PET (Table 3)

The optimal selection marker for use in clinical trials would identify individuals in prodromal or preclinical phase with AD pathology that are more likely to progress over the period of the trial, typically 1-2 years. Clinical trials are moving toward a selection strategy that includes people with MCI who also show significant amyloid pathology using PET imaging; a combination that is the best approximation to the optimal selection marker today. However, identifying such individuals is problematical and the screen failure rate at PET imaging stage is high. This is costly, exposes large numbers of people to PET scans, poses a considerable logistics challenge, and represents a significant hurdle to effective trials recruitment. If a relatively facile (easily obtained, fast to analyze) biomarker could be found that would reduce this screen failure, then the process of clinical trials recruitment would be much enhanced. With this in mind, we set out to employ the endophenotype design to identify a biomarker that would contribute to the identification of people most likely to be harboring amyloid pathology (neocortical amyloid burden; NAB). Note that such a biomarker need only have relatively modest positive predictive value to have a substantial impact on cost and speed of recruitment to clinical trials.

Using samples from the Baltimore Longitudinal Study of Aging collected from non-demented older individuals we employed 2DGE/MS proteomics on plasma collected 10 years prior to 11\textsuperscript{C}-PiB PET scans and within ± 1 year of the scan \cite{56}. We found 6 proteins (ApoE, haptoglobin, plasminogen, complement C3, albumin, and IgG) that could discriminate subjects with high amyloid burden from those with low burden. Given the association between the APOE gene and AD risk, we selected this protein for subsequent studies, confirming the strong association between plasma ApoE concentration and Aβ burden in the medial temporal lobe.

In another study \cite{57}, we also used an untargeted approach to discover biomarkers but this study included AD, MCI, and cognitively healthy subjects, who were dichotomized by high or low NAB. Seventeen candidate biomarkers were identified and then replicated by using immunoassays (ELISA) in both the same and an independent cohort. The technical replication in the same cohort validated three proteins including α2M, fibrinogen γ-chain (FGG), and factor H-related protein 1 (FHR-1), while an opposite trend was observed for FHR-1 between discovery and validation phase. The independent replication in different cohorts found that only FGG was significantly associated with high Aβ burden, predicting high NAB with a sensitivity of 59% and specificity of 78%.

We then used the same approach (2DGE and LC-MS/MS) to conduct a longitudinal study testing samples over a 12-year period from non-demented older individuals \cite{58}. Briefly, plasma proteins were assessed for their relationship with NAB at three time points: 12 years before amyloid PET imaging, 6 years before imaging, and concurrent to imaging. Proteins that were consistently associated with NAB at all three time points were chosen as candidates and further studied in independent cohorts. For the discovery study, we found seven proteins associated with NAB consistently across all three time points: α2M, apolipoprotein A-I (APO-A1), complement C3, complement C4B, haptoglobin, Ig kappa chain C region, and serum albumin. In independent cohorts, we replicated five of these (α2M, serum albumin, APO-A1, C3, and haptoglobin). This study demonstrated that blood-based biomarkers remain stable and could reflect amyloid burden throughout the course of disease, from the pre-clinical phase through to established clinical syndromes. Moreover, this study replicated several biomarkers from our previous studies, e.g., α2M \cite{57} and complement C3 \cite{56}.

In addition to discovery, we have also conducted several replication studies. For example, in \cite{59}, we tested the association of 146 plasma analytes with NAB in 71 subjects. Results showed sixteen proteins were found to associate with NAB, including two proteins (APOE and complement C3) reported in our previous study \cite{56}. Among these biomarkers, some were also found to associate with other AD related phenotypes. For example, leptin was also associated with CSF Aβ\textsubscript{1-42}, leptin, α-1-antitrypsin, and cortisol were related to MRI features, and α-1-antitrypsin, complement C3, and fibrinogen were associated with cognitive scores, replicating our previous results \cite{41, 44, 45, 48}. Given that some proteins were highly correlated with each other, we therefore chose a panel of thirteen biomarkers including c-peptide, fibrinogen, α-1-antitrypsin, PPy, complement C3, vitronectin, cortisol, AXL receptor kinase, interleukin-3, interleukin-13, matrix
Table 3
List of studies of blood-based protein biomarkers of brain amyloid burden measured by PET

<table>
<thead>
<tr>
<th>Reference</th>
<th>Cohorts</th>
<th>Markers</th>
<th>Methods</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thambisetty et al., [56]</td>
<td>Discovery: BLSA-NI cohort: 57 non-demented elderly</td>
<td>ApoE, haptoglobin, plasminogen, complement-C3, albumin, and IgG</td>
<td>2DGE and LC-MS/MS</td>
<td>-six proteins discriminate individuals with high Aβ burden from those low</td>
</tr>
<tr>
<td></td>
<td>Validation: BLSA-NI cohort: 42 non-demented elderly</td>
<td>ApoE</td>
<td>ELISA</td>
<td>-strong association between ApoE and Aβ burden in the medial temporal lobe</td>
</tr>
<tr>
<td>Ashton et al., [57]</td>
<td>Discovery: AIBL cohort: 6 AD, 23 MCI and 50 CTL</td>
<td>17 proteins identified</td>
<td>TMT/ LC-MS/MS</td>
<td>-17 proteins were associated with NAB</td>
</tr>
<tr>
<td></td>
<td>Validation: -AIBL cohort: 6 AD, 23 MCI and 50CTL; -UCSF memory and aging center cohort: 25 AD and 54 non-AD dementia controls</td>
<td>-17 proteins for AIBL cohorts;</td>
<td>ELISA</td>
<td>-in AIBL cohorts, α2M and FGG were significantly associated with NAB; the combination of FGG levels and age could predict NAB with a sensitivity of 59% and specificity of 78%</td>
</tr>
<tr>
<td>Westwood et al., [58]</td>
<td>Discovery: BLSA-NI cohort: 54 non demented elderly</td>
<td>α2M, APO-A1, complement C3, complement C4B, haptoglobin, Ig kappa chain C region, and serum albumin</td>
<td>2DGE and LC-MS/MS</td>
<td>-seven proteins were associated with NAB consistently across all three time points</td>
</tr>
<tr>
<td></td>
<td>Validation: AIBL cohort: 6 AD, 22 MCI and 48 CTL</td>
<td>α2M, serum albumin, APO-A1, complement C3 and haptoglobin</td>
<td>TMT/ LC-MS/MS</td>
<td>-five proteins were associated with NAB</td>
</tr>
<tr>
<td>Kiddle et al., [59]</td>
<td>ADNI cohort: 16 AD, 52 MCI and 3 CTL</td>
<td>146 proteins probed, 16 proteins identified</td>
<td>Human Discovery Multi-Analyte Profile (MAP) and Luminex</td>
<td>-a panel of 13 biomarkers (c-peptide, FGG, α-1-antitrypsin, PPY, complement C3, vitronectin, cortisol, AXL receptor kinase, IL-3, IL-13, matrix metalloproteinase-9 total, APOE and IgE) couple with co-variate factors account for &gt;30% of variance of brain amyloid burden</td>
</tr>
<tr>
<td>Voyle et al., [71]</td>
<td>AIBL cohort: 78 AD and 120 CTL</td>
<td>41 proteins probed, 2 proteins identified</td>
<td>SOMAscan</td>
<td>-PPY and IgM were associated with NAB</td>
</tr>
</tbody>
</table>

PET, positron emission tomography; AD, Alzheimer’s disease; ApoE, apolipoprotein E; AIBL, Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing; ADNI, Alzheimer’s Disease Neuroimaging Initiative; UCSF, University of California San Francisco; TMT, Tandem Mass Tags; 2DGE, 2D gel electrophoresis; LC-MS/MS, liquid chromatography tandem mass spectrometry; ELISA, enzyme-linked immunosorbent assay; α2M, α2-macroglobulin; FGG, fibrinogen γ-chain; PPY, pancreatic polypeptide; MCI, mild cognitive impairment; NAB, neocortical amyloid burden; BLSA-NI, Baltimore Longitudinal Study of Aging Neuroimaging Sub-study; APO-A1, apolipoprotein A1; IL, interleukin; CTL, cognitively healthy elderly controls.

metalloproteinase-9 total, APOE, and IgE. This panel, coupled with covariates, could account for more than 30% of variance of brain amyloid burden, much higher than the covariate factors alone (4-13%). Furthermore, this panel could predict Aβ positive individuals with a sensitivity of 92% and a specificity of 55%, suggesting that plasma proteins could reflect the levels of Aβ burden in the brain. If repeated in other studies, this panel could substantially facilitate clinical trials by contributing to rapid and effective selection of research participants most likely to have NAB and hence reducing screen failure rates, reducing cost and time of trial start-up and reducing exposure of potential participants to PET imaging.
Table 4

<table>
<thead>
<tr>
<th>Reference</th>
<th>Cohorts</th>
<th>Markers</th>
<th>Methods</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kiddle et al., [59]</td>
<td>ADNI cohort: 16 AD, 52 MCI and 3 controls</td>
<td>146 proteins probed</td>
<td>Human Discovery</td>
<td>-Leptin was significantly associated with CSF Aβ1-42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Multi-Analyte Profile (MAP)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>and Luminox</td>
<td></td>
</tr>
<tr>
<td>Westwood et al., [79]</td>
<td>METSIM study cohort: 58 cognitively healthy men including 30 insulin non-resistant and 28 insulin resistant subjects</td>
<td>FCN2, FGG, CFHR1 and ApoA1</td>
<td>FCN2, FGG and CFHR1 proteins were measured by ELISA, ApoA1 was measured by Luminox xMAP assay</td>
<td>-FCN2 was significantly negatively correlated with CSF Aβ</td>
</tr>
</tbody>
</table>

CSF, cerebrospinal fluid; AD, Alzheimer’s disease; ADNI, Alzheimer’s Disease Neuroimaging Initiative; METSIM, Metabolic Syndrome in Men; MCI, mild cognitive impairment; FCN2, ficolin-2; FGG, fibrinogen γ-chain; CFHR1, complement factor H-related 1; APO-A1, apolipoprotein A1; LC-MS/MS, Liquid chromatography tandem mass spectrometry; ELISA, enzyme-linked immunosorbent assay; FGB; fibrinogen β chain; ApoC-IV, apolipoprotein C-IV; PET, positron emission tomography; SMC, subjective memory complaints; MCIc, stable MCI patients; MCIc, MCI converting to AD.

Blood-based biomarkers of CSF Aβ (Table 4)

An analogous approach to the search for blood-based markers to predict NAB is to use CSF Aβ and tau measures as the endophenotype variable. Whether such an approach will nominate the same markers as that based on NAB will be interesting; CSF measures of pathology correlate with NAB as one would expect, but only imperfectly. They measure different things and identifying blood-based correlates of both might help to explore this difference. We have begun a series of studies using a range of protein technologies using this study design, as shown in Table 4.

In another study, we have also used MS to identify blood biomarkers associated with CSF tau/Aβ42 pathology and performed replication studies in independent cohorts by immunocapture, confirming some of these hits (Baird et al, in preparation).

WHY DO MOST BIOMARKERS FAIL REPLICA TION?

In a decade of studies seeking biomarkers in blood, we have nominated many and replicated some. In order to explore this further—to test the commonly heard dismissal of blood biomarker research that ‘nothing replicates’—we have performed perhaps the first, and certainly largest systematic replication study [45]. Briefly, we performed a systematic review and from 21 published discovery or large panel (>100 proteins) studies identified 163 candidate proteins, among which approximately 66% were reported only in only one study. Then we used the SOMAscan platform to test 94 of these 163 candidates in 677 subjects (AD, controls, and MCI). The results showed that a panel of 13 proteins (pancreatic prohormone, C-C motif chemokine 18, α-1-antitrypsin, complement C6, insulin-like growth factor-binding protein 2, angiopoietin-2, C-C motif chemokine 15, cystatin C, tumor necrosis factor receptor superfamily member 1B, β-2-microglobulin, prolactin, haptoglobin, and metalloprotease inhibitor 1) could discriminate AD from controls with a sensitivity of 0.83 and a specificity of 0.66.

While this study refutes the superficial ‘nothing replicates’ dismissal, it remains the case that in the field as a whole, most nominated biomarkers fail replication posing a substantial challenge to development [60]. The reason for such failure might be caused by the heterogeneity of the disease itself as well as the complexity of blood. Indeed, the protein levels in blood span ten orders of magnitude, making the investigation of lower abundant proteins extremely challenging. In addition to the reasons mentioned above, a number of other major factors could also lead to the failure of replication. They include pre-analytical processes, analysis of different blood fractions, use of different analytical platforms, and inappropriate statistical analysis, which will be discussed below.

Pre-analytical processing

Pre-analytical processing has a substantial impact on proteomic results although it is frequently disregarded. For example, Plebani and Carraro [61] analyzed analytical errors in an emergency laboratory and found that 68.2% of mistakes were caused by pre-analytical phases, in contrast to 13.3% and 18.5% for the intra and post-analytical phases, respectively. In another review, Bonini et al [62] found that more than 68% of the errors occur in the pre- or post-analytical phase, whereas only 13–32% could be traced back to the analytical phase.
Various pre-analytical factors affect the quality of the specimens, including blood collection, e.g., needle used, the site used for blood withdrawal, and type of collection tube; blood processing, e.g., the time from blood-draw to storage, centrifugation parameters, and container type; blood storage, e.g., sample storage temperature and duration; blood transportation, etc. Since all these factors could influence the results, at the very least such parameters should be recorded and available for study and standardization of methodologies would be desirable. Currently, an international working group led by O’Bryant has provided the initial starting point for such guidelines and standardized operating procedures [63].

Serum versus plasma

Although we refer in this review to ‘blood’, there are of course many fractions of this tissue (plasma, serum, cells) that influence the concentration of proteins. Whole blood includes red blood cells, white blood cells, platelets, and proteins. Serum is the protein rich fluid in which these cells are suspended, obtained following centrifugation, and plasma is serum collected with methods to preserve fibrinogen and clotting factors. Other compartments in blood include the platelets, erythrocytes, the aggregate of white cells referred to as ‘buffy coat’ and the individual white cell populations separable by flow cytometry. Each of these cellular and non-cellular compartments might be the source of biomarkers and indeed each has been used in biomarker studies in AD. In a recent study, Huebinger et al. [64] compared the concentration of 100 proteins in matched samples of serum and plasma from 39 AD patients, showing that 40 proteins were high correlated between blood fractions while the remaining proteins were only moderately or weakly correlated, including some of considerable interest in AD.

Sample size versus biomarkers panel

The ratio between sample size and analytes will also influence the reproducibility of biomarkers. Overfitting bias occurs when a multi-marker panel is inappropriately large with respect to the number of cases evaluated. Generally, when the ratio of samples to analytes is less than 10, it is considered of potential bias [65] suggesting that overfitting of data is a risk in most agnostic proteomic studies. The failure to replicate blood biomarker studies in AD are not unique to unbiased proteomic analyses and considerations of preanalytical processing, matrix used, sample size, and statistical analyses are equally relevant to other ‘omics’ data. The recent discovery of a panel of lipid metabolites that could predict incident AD with >90% accuracy is especially relevant in this context [66]. The ultimate failure to replicate these findings from a small discovery study in large, independent cohorts [67–69] again highlights these critical considerations [70]. Ultimately it is replication that is the acid test of any study but the biomarker field in AD research today is arguably in the same phase as genetic studies were a decade or more ago with analyte size rapidly increasing but with sample numbers not keeping up. The field of genetics was transformed by the huge collaborative studies now typically including tens of thousands of research participants. Biomarker studies will be abandoned before studies get to this size as a biomarker effect size needs to be of practical utility and hence larger than the effect size of most genetic susceptibility factors for common disease, but they do need to be substantially larger than most studies are today.

Analytical methods

Using different methods will inevitably lead to different results and this may contribute to why proteins identified by an untargeted approach (e.g., 2D-GE/MS) fail in replication using a targeted approach (e.g., ELISA, western blot). Indeed, proteins exist in many different isoforms, they are metabolized, have different biophysical states, are complexed with other proteins, have altered activities, and have a very large and variable number of post-translational modifications. Even very similar targeted analytical methods are not completely comparable. An ELISA, for example, is not actually a measure of a protein but of the binding of a capture antibody to an epitope. Change the binding agent, or the epitope and a change in the result is only to be expected.

Statistical analysis

Different complex statistical analysis also might result in discrepancy across studies. In one of our studies [71], we compared our statistical analysis model with the one used by Burnham et al. [72] and found that the model we used gave somewhat improved results. Given that there are no absolutes when it comes to choice of complex statistical approach, the only viable solution is open access to raw data as well as transparency of analytical process.
MOVING FROM DISCOVERY TO TRANSLATION AND FROM PROTEIN TO MULTI-MODAL MARKERS

Although the majority of published biomarkers have failed to replicate and perform as a reproducible classifier for AD, we do observe several proteins consistently associated with clinical AD or with multiple indicators of AD pathology in our studies (see Table 5), many of which are also replicated by other laboratories. Given the relatively small size of the studies thus far, these results are highly promising but clearly need to be replicated in larger studies.

The National Biomarker Development Alliance (NBDA) proposes that following discovery of a nominated and replicated biomarker, the subsequent phase of research is ‘translatable discovery’; the demonstration that the marker is “accurate and reproducible within the intended context of use—in other words, it has evidence-based potential for use in diagnosis, clinical decision making, or as a clinical tool (e.g., stratifying patients for trial)” and also that “the assay will ultimately perform in the real world of varying sample quality” (http://nbdbiomarkers.org/about/what-we-do/pipeline-overview/translatable-discovery). The evidence reviewed here we believe demonstrates that there are biomarkers in blood in AD that should now transition to this phase of development. With that in mind, we are now embarking on large scale Translatable Discovery studies using some of the infrastructures for collaboration established in Europe over the past few years.

The IMI-European Medical Information Framework (http://www.emif.eu) is a large public private consortium across multiple countries that enables data visibility and interoperability to facilitate research. The data sets included are both large, real-world and population datasets as well as research cohorts. We have used the infrastructure and collaborations established by EMIF to identify more than 1500 samples for biomarker studies and are currently analyzing these samples using endophenotype design and both targeted and untargeted proteomics including explicitly ‘Translatable Discovery’ of the biomarkers described in this review. As the samples come from multiple cohorts with somewhat varying sample collection protocols, they match the NBDA requirement for this phase of development for demonstrable efficacy in real-world situations. Very similar to EMIF in some respects, and reutilizing much of the informatics infrastructure, is the Dementias Platform UK (DPUK; http://www.dementiasplatform.uk/) which is also now generating both Early Discovery and Translatable Discovery biomarker programs in suitably large cohort studies.

Of other studies, the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (http://www.adni-info.org/) has in many ways led the field. First, because this study established many of the protocols that have become standard in the field, especially in imaging. Second because it has gathered and analyzed large numbers of brain scans, genetic profiles, and biomarkers in blood and CSF that have been used to develop methods to assess disease progress and potentially, effects of treatment. Thirdly, and arguably most importantly, the ADNI study has been an ‘open science’ program sharing data with the scientific community from the outset and in doing so has made substantial contribution not only to biomarker research but to the increasingly collaborative spirit of research in this area. Through a collaborative agreement with ADNI, the Australian Imaging, Biomarker & Lifestyle Flagship Study of

<table>
<thead>
<tr>
<th>Biomarkers</th>
<th>Reference number</th>
<th>Case controls studies</th>
<th>Biomarkers of brain atrophy and rate of cognitive decline</th>
<th>Biomarkers of brain amyloid burden</th>
<th>Biomarkers of CSF Aβ</th>
</tr>
</thead>
<tbody>
<tr>
<td>clusterin</td>
<td>6</td>
<td>[36]</td>
<td>[33], [45], [41], [43], [44]</td>
<td>[41]</td>
<td></td>
</tr>
<tr>
<td>complement C3</td>
<td>6</td>
<td></td>
<td>[45], [41], [48], [59]</td>
<td>[56], [58], [59]</td>
<td></td>
</tr>
<tr>
<td>FGG</td>
<td>5</td>
<td></td>
<td>[41], [48]</td>
<td>[57], [59]</td>
<td>[79]</td>
</tr>
<tr>
<td>APOE</td>
<td>4</td>
<td>[36]</td>
<td></td>
<td>[59], [56]</td>
<td></td>
</tr>
<tr>
<td>PPY</td>
<td>3</td>
<td></td>
<td></td>
<td>[59], [71]</td>
<td></td>
</tr>
<tr>
<td>α-2M</td>
<td>3</td>
<td>[15]</td>
<td></td>
<td>[57], [58]</td>
<td></td>
</tr>
<tr>
<td>α-1-antitrypsin</td>
<td>3</td>
<td>[45]</td>
<td>[45], [44], [59]</td>
<td>[59]</td>
<td></td>
</tr>
<tr>
<td>FCN2</td>
<td>3</td>
<td></td>
<td></td>
<td>[33]</td>
<td></td>
</tr>
<tr>
<td>TTR</td>
<td>3</td>
<td></td>
<td></td>
<td>[41], [44], [49]</td>
<td></td>
</tr>
</tbody>
</table>

α2M, α-2-macroglobulin; FGG, γ-fibrinogen; PPY, pancreatic polypeptide; FCN2, ficolin-2; TTR, transthyretin.
suitable for such studies and to reduce the cost of for selection markers to identify a participant group sitating long and expensive trials and hence the need for trial outcome measures; specifically biomarkers to enable proof of concept studies in preclinical disease where proof of concept is most likely to be demonstrable. In the Deep and Frequent Phenotyping study, we shall collect data for Early Discovery of such markers; utilizing a suite of technologies ranging from structural and functional imaging of brain and eye, electrophysiology including MEG and EEG, connected devices to measure gait, movement, cognition, and moreover sample collection including blood, CSF, and urine. These will all be measured in preclinical and prodromal AD and controls at frequent intervals ranging from three to six times over the course of a year. This enormously challenging, complicated but exciting study will share both samples and data with the scientific community as part of the drive to Open Science and in the belief that it is only through co-operation and collaboration that we will make progress.

ACKNOWLEDGMENTS

Funding in the authors laboratory supporting this work came from Alzheimer’s Research UK, Alzheimer’s Society, Parkinson’s UK, MRC, the European Union through FP6 and IMI programmes, and NIHR.

Authors’ disclosures available online (http://j-alz.com/manuscript-disclosures/17-0531r1).

REFERENCES


