A blood-based signature of cerebral spinal fluid Aβ_{1-42} status

Benjamin Goudey[1,2,3,*], Bowen J Fung[1,4,*], Christine Schieber[1], for the Alzheimer's Disease Metabolomics Consortium[†4], for the Alzheimer's Disease Neuroimaging Initiative[‡5], and Noel G Faux[1,6,+]

[1]IBM Australia - Research, Carlton, Victoria, Australia
[4]School of Psychological Sciences, University of Melbourne, Victoria, Australia
[2]Centre for Epidemiology and Biostatistics, The University of Melbourne, Parkville, Australia
[3]Department of Computing and Information System, The University of Melbourne, Parkville, Australia
[4]A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.
[5]A complete listing of ADMC investigators can be found at: https://sites.duke.edu/adnimetab/about-us/the-team/
[6]The Florey Institute for Neuroscience and Mental Health, The University of Melbourne, Parkville, Australia
[+]Corresponding author: noel.faux@au1.ibm.com
[*]These authors contributed equally to this work

ABSTRACT

It is increasingly recognized that Alzheimer's disease (AD) exists before dementia is present and that shifts in amyloid beta occur long before clinical symptoms can be detected. Early detection of these molecular changes is a key aspect for the success of interventions aimed at slowing down rates of cognitive decline. Recent evidence indicates that of the two established methods for measuring amyloid, a decrease in cerebral spinal fluid (CSF) amyloid β_{1-42} (Aβ_{1-42}) may be an earlier indicator of Alzheimer's disease risk than measures of amyloid obtained from Positron Emission Topography (PET). However, CSF collection is highly invasive and expensive. In contrast, blood collection is routinely performed, minimally invasive and cheap. In this work, we develop a blood-based signature that can provide a cheap and minimally invasive estimation of an individual's CSF amyloid status using a machine learning approach. We show that a Random Forest model derived from plasma analytes can accurately predict subjects as having abnormal (low) CSF Aβ_{1-42} levels indicative of AD risk (0.84 AUC, 0.78 sensitivity, and 0.73 specificity). Refinement of the modeling indicates that only *APOEε4* carrier status and four analytes are required to achieve a high level of accuracy. Furthermore, we show across an independent validation cohort that individuals with predicted abnormal CSF Aβ_{1-42} levels transitioned to an AD diagnosis over 120 months significantly faster than those predicted with normal CSF Aβ_{1-42} levels and that the resulting model also performs reasonably across PET Aβ_{1-42} status.
This is the first study to show that a machine learning approach, using plasma protein levels, age and *APOEε4* carrier status, is able to predict CSF Aβ_{1-42} status, the earliest risk indicator for AD, with high accuracy.

1 Introduction

Alzheimer's disease (AD) is a terminal neurodegenerative disease that has historically been diagnosed based on "clinically significant" cognitive decline of an individual and exclusion of other conditions. However, it is increasingly recognized that AD is a decades-long neurodegenerative process, with shifts in amyloid β_{1-42} (Aβ_{1-42}) providing the first indicators of disease development, long before "Alzheimer's dementia" (significant cognitive decline) is clinically apparent[1–5].

There is currently no cure or disease-modifying therapy for this terminal illness despite hundreds of clinical trials being conducted since 2002[6,7]. It is hypothesized that the high failure rate of AD trials is in part due to the trials targeting AD patients with significant cognitive impairment, who are therefore in the late stages of the disease and likely have suffered a level of brain tissue loss that cannot be compensated for. Compounding this is the discovery that many patients enrolled in clinical trials, up to 20% in one study[8], were retrospectively found to have normal levels of amyloid and hence did not have AD[9]. Given

[†]Data used in preparation of this article were generated by the Alzheimer's Disease Metabolomics Consortium (ADMC). As such, the investigators within the ADMC provided data but did not participate in analysis or writing of this report.

[‡]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report.

these findings, there is a great interest in amyloid screening for clinical trial enrichment in order to recruit individuals at the earliest stages of AD, where intervention is thought to have the greatest chance of success[10]. It would also ensure that included individuals are amyloid positive (i.e have abnormal levels of amyloid), a necessary precondition for the development of AD. This selective screening is a precursor for the longer-term goal of population screening for AD[11].

There are currently two established methods to measure an individual's amyloid burden: either in vivo in the form of reduced levels of Aβ_{1-42} in the cerebral spinal fluid (CSF) or increased uptake of radioactive tracers that bind selectively to the Aβ fibrillary aggregates by PET imaging. Unfortunately, existing methodologies for measuring an individual's amyloid levels suffer drawbacks that limit their utility for screening. Lumbar punctures are highly invasive, with this factor alone limiting the applicability of CSF biomarkers for screening. While PET scans are less invasive, they are far more expensive and access to PET scanning facilities is limited in some regions. Despite this, nearly all current trials related to amyloid now require positive amyloid imaging at baseline to ensure accurate diagnosis, a cost-intensive process[7].

Despite their invasiveness, recent studies have found evidence that changes in CSF Aβ_{1-42} may indicate AD risk long before these same changes are reflected in PET Aβ imaging[12–14]. Palmqvist et al[13] have shown shown compelling evidence that changes in CSF Aβ_{1-42} occur up to a decade before the same signal is found by PET Aβ imaging. These results indicate that CSF may be a more suitable measure for early detection, whereas Aβ PET contributes independent information that is more related to disease progression and downstream pathology.

To bypass the invasiveness of CSF collection, there is a strong interest in finding blood-based markers that yield the same information about amyloid status as would be obtained from CSF. There have been a number of studies which have shown that a blood protein signature can be found that reflects AD brain pathology as measured by PET[15–24]. Of particular interest is the recent study by Nakamura et al. (2018) [25], whereby levels of Aβ_{1-40}, Aβ_{1-42} and APP$_{669-771}$ in plasma, measured using specialised immunoprecipitation (IP) coupled with Matrix Assisted Laser Desorption/Ionization (MALDI) time-of -flight (TOF) mass spectrometry (MS) (henceforth referred to as IP-MALDI-TOF-MS), were shown in combination to have strong performance (> 0.94 area under the receiver operating characteristic curve (AUC)) in predicting PET Aβ_{1-42} status across two cohorts. This combination of biomarkers was also found to be predictive of CSF Aβ_{1-42} status, with an AUC of 0.876 on a smaller subset of patients (n=46). The novel method employed by Nakamura et al. (2018) [25], IP-MALDI-TOF-MS, is still in its infancy and it is unclear how easily this will be translated into a clinical setting. Thus, there is still strong interest in finding blood markers for CSF Aβ_{1-42} using alternative approaches that rely on more established technologies.

Here, we evaluated the ability of proteomic and metabolomic data to predict the levels of CSF Aβ_{1-42} using a Random Forest (RF) approach and explore which types of measurements lead to the strongest predictive performance. We then determine the minimal set of features required to achieve comparable predictive performance. Finally, we evaluate the robustness and utility of these predictive models across a held-out validation cohort of individuals with mild cognitive impairment (MCI), demonstrating that subjects with predicted abnormal CSF Aβ_{1-42} levels showed a faster rate of cognitive decline (measured by transition to AD) than those with predicted normal CSF Aβ_{1-42} levels.

2 Methods

2.1 Overview of cohort and measurements

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a large, multicenter, longitudinal neuroimaging study, launched in 2004 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and non-profit organizations.

ADNI is a longitudinal study of older adults, designed to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

2.2 Data preparation

We examined 566 individuals in the ADNI cohort who had measures at baseline of age, *APOE*ε4 carrier status, 193 protein levels (including homocysteine, $A\beta_{1-40}$, and $A\beta_{1-42}$) and a further 190 proteins measured on a Rules Based Medicine (RBM) platform) and 186 LC-MS/MS metabolites and lipids. After applying previously documented quality control procedures (Supplementary Methods) and removing analytes with more than 15% missingness, 149 proteins and 138 metabolites remained. No samples were removed from the analysis as missingness levels were less than 5%. Any remaining missing data points were imputed using an unsupervised RF approach[26]. After quality control, a total of 566 individuals, each with measures for 289 analytes, age and APOEε4 carrier status, were present in the ADNI cohort (Table 1).

2.3 Training and Validation cohorts

The 566 individuals in this study ranged from 54.4-89.6 years of age and could be categorized by their AD clinical diagnosis as cognitively normal (CN; n = 58), amnestic MCI (n = 396) or, probable AD (n = 112). A breakdown of the demographics of the

Dataset	Training			Validation		
Diagnosis	**CN**	**MCI**	**AD**	**CN**	**MCI**	**AD**
Number of participants (n)	58	198	102	0	198	10
Age (mean, (sd))	75.11 (5.77)	74.35 (7.48)	74.90 (7.91)	-	75.13 (7.32)	73.73 (10.04)
Gender; female (n, (%))	28 (48.28)	65 (32.82)	43 (42.16)	-	75 (37.88)	4 (40.00)
Years of education (mean, (sd))	15.67 (2.78)	15.81 (2.99)	15.16 (3.29)	-	15.48 (3.09)	14.70 (2.06)
***APOEε4* carriers (n, (%))**	5 (8.62)	106 (53.54)	71 (69.61)	-	105 (53.03)	5 (50.00)
CSF $A\beta_{1-42}$ abnormal (n, (%))	0 (0)	146 (73.7)	94 (92.16)	-	-	-
PET imaging (n, (%))	28 (48.28)	71 (35.86)	9 (8.82)	-	108 (65.55)	-

Table 1. Demographic characteristics of the ADNI data set used separated into training and validation cohorts, corresponding to individuals with and without CSF measures respectively. Columns in each cohort provide a further breakdown into individuals that are cognitively normal (CN), have mild cognitive impairment (MCI) or Alzheimer's disease (AD). The units of each cell are shown in parentheses in the row names and commonly include number of patients (n) or mean of a given quantity (mean). If a secondary measure (percent (%) or standard deviation (sd)) is also present, it is listed in brackets next to the primary measure.

566 individuals by baseline diagnosis and CSF availability is shown in Table 1.

This cohort was split into a training and a validation cohort with 356 and 210 individuals with and without measures of CSF $A\beta_{1-42}$, respectively, where CSF was measured using the Luminex xMAP platform and Innogenetics INNO-BIA AlzBio3 immunoassay. The training set was used to build predictive models and evaluate their performance directly using the measured $A\beta_{1-42}$ levels while the validation cohort was not used to build the predictive model, but was instead used to evaluate the generalizability and utility of the model's predictions. For each cohort, we also considered a subset of individuals for whom $A\beta_{1-42}$ status from PET was available at least one timepoint (not just at baseline), either using Pittsburgh compound B (PiB) or Florbetapir (AV45) tracer, for further validation of our modelling. Further demographic information for these cohorts can be found in Supplementary Tables 1 to 3.

2.4 Binary and regression modelling tasks

The primary aim of this work was to produce a model that predicts if an individual's CSF $A\beta_{1-42}$ levels are below the recognized clinical threshold of 192pg/ml for the platform used, indicating an abnormal CSF $A\beta_{1-42}$ level, and hence AD risk. Given the continuous CSF $A\beta_{1-42}$ measures in the ADNI cohort, two approaches were considered: a 'regression' task, learning the continuous CSF $A\beta_{1-42}$ levels and thresholding these post-prediction and a 'binary' task, learning the dichotomized CSF $A\beta_{1-42}$ status based on clinical thresholds directly. While both tasks result in a binary classifier, they face different trade-offs. The regression task makes use of the full information in the CSF levels but needs to learn a suitable threshold to convert its continuous predictions into suitable binary labels whereas the binary task only learns from the dichotomized CSF levels. Given these trade-offs, we have investigated both modelling approaches throughout this work.

2.5 Statistical modeling

We made use of Random Forests (RF) as the modeling approach to predict CSF $A\beta_{1-42}$ levels for both the binary and regression tasks. RFs are a widely-used ensemble method that have a number of advantages for the small sample size and disparate types of features observed in the ADNI dataset. RFs are invariant to the scale of the observed features and make few assumptions about the distributions of observed data allowing them to be applied to multiple data modalities easily.

All analysis in this work made use of the RF implementation in the R package *ranger*[27]. Each forest contained 2000 individual trees, each making use of a random selection of $p^{3/4}$ features, where p was the total number of variables used in a given model. These parameter choices were based on recommendations provided in Ishwaran *et al.* (2011)[28]. All other parameters in the *ranger* implementation were set to their default values.

To get an estimate of the performance of our models, we have made use of a nested cross-validation (CV) framework, whereby an inner CV was used to determine model parameters, and the outer CV was used to gain an estimate of the model's performance on unseen data[29]. In this study, we used 3 repetitions of 3 fold cross validation for the inner CV loop and 10 repetitions of 10 fold CV for the outer loop.

As the RF used pre-determined parameter values, only a single parameter had to be determined; the threshold on the continuous regression predictions necessary to generate binary labels. This threshold was selected based on performance in the

inner CV loop, using the R package *OptimalCutoffs*[30] to evaluate six potential cutoff metrics (Supplementary Methods) and selecting the method which maximized the accuracy over all of the test folds from the inner cross validation loops. The best performing cutoff criteria was then used in the current iteration of the outer cross-validation loop and the accuracy, sensitivity and specificity derived from this threshold was recorded for that fold. While this approach means that a different method could be used to derive the regression threshold for each fold in the outer cross-validation, the resulting estimate of performance is unbiased and hence is likely to be more representative of performance on unseen data compared with selecting a threshold based on the entire set of training data.

2.6 Measures of model performance

Model performance was summarized by the mean and standard deviation of area under the Receiver Operating Characteristic (ROC) curve (AUC), accuracy, sensitivity and specificity from the testing performance across the different cross-validation runs. R^2 values were also calculated for the regression task. Increases in AUC between models were tested for significance using a one-tailed Wilcoxon signed-rank test. Receiver operating curves were constructed by aggregating all of the test predictions from the outer cross-validation.

2.7 Evaluating importance of different modalities

The input variables were separated into three classes: a commonly used baseline model (B) including age and *APOEε*4 carrier status; Proteomics (P), which included the 146 analytes measured on the RBM panel as well as homocysteine and plasma $A\beta_{1-40}$ and $A\beta_{1-42}$; Metabolomics (M), including 138 metabolites and lipids.

Four separate random forests were created using different subsets of these features to determine which were most useful for modeling CSF $A\beta_{1-42}$. We denoted these models by the combination of features they included; for example 'BPM' refers to a model built using all three classes of features. The best performing model was selected for all subsequent analysis.

2.8 Discovery of smallest set of markers needed for strong predictive performance

After evaluating the impact of the different modalities, we determine the minimum set of individual analytes necessary to achieve high predictive performance. This was done by treating the number of included features as a parameter to be determined in our nested CV framework. Within each fold of the inner cross validation loop, we used a recursive feature elimination approach, ranking features according to their *Variable Importance*, the difference in the prediction error on the out-of-bag data when a given feature was permuted and unpermuted[31], and removed the lowest ranking features in a stepwise fashion. The AUC of the resulting RF was recorded, and the procedure was repeated over increasingly smaller subsets of features until no features were left to be removed. After the inner CV loop finishes, we determine the number of features that achieved the optimal trade-off between model complexity and performance by selecting the smallest subset of features that achieved within 4% of the maximal observed AUC. A model using this subset of features was then trained on all training folds of the outer CV loop and evaluated on the test fold. Again, by determining the number of features to include within our nested cross-validation framework, we are able to determine an unbiased estimate of the model's expected performance over unseen data.

2.9 Survival Analysis

Survival analysis was conducted to determine if the rate of conversion from MCI to AD was different between those with predicted low and normal CSF $A\beta_{1-42}$ levels, enabling us to determine if our predictions lead to useful clinical outcomes in the validation cohort.

Four separate analyses were performed: using the measured CSF status on the training set (n=198), and using the predicted CSF status from either the B, BP and BP_{fs}, the feature selected model (model with the smallest set of features), in the validation set (n=198). For each analysis, we have examined the hazard ratios using Cox regression and used log-rank tests to compare the survival distribution of low/normal CSF $A\beta_{1-42}$ stratifications in the three analyses, as well to compare equivalence between the actual and predicted stratifications.

2.10 Validation performance over PET $A\beta_{1-42}$ status

In order to further validate our model, we have examined the ability of our model to differentiate PET $A\beta_{1-42}$ abnormal and normal status. While it is known that $A\beta_{1-42}$ status from PET can differ from that observed in CSF, measurements from the two modalities are correlated and should be very similar for individuals who are not close to the cutoff indicating pathology. This provides a further indication of our model's ability to determine $A\beta_{1-42}$ status in individuals in the validation cohort, where CSF measurements are not available.

Given that only a limited number of individuals had associated measures of PET imaging at baseline (n=18 and 27 for training and validation cohorts respectively), we have made use of the earliest PET image available, leaving us with 108 and 68 individuals in the training and validation cohort to evaluate. The threshold for abnormality was defined as a SUVR of 1.5 and

Feature set	Regression					Binary			
	AUC	R^2	Acc	Sens	Spec	AUC	Acc	Sens	Spec
BPM	0.830 (0.08)	0.274 (0.11)	**0.771 (0.07)**	0.784 (0.09)	**0.748 (0.13)**	**0.826 (0.07)**	**0.766 (0.06)**	**0.884 (0.06)**	0.536 (0.14)
BM	0.788 (0.08)	0.209 (0.13)	0.744 (0.08)	0.779 (0.09)	0.679 (0.16)	0.784 (0.08)	0.737 (0.08)	0.855 (0.08)	0.507 (0.15)
BP	**0.839 (0.07)**	**0.293 (0.11)**	**0.765 (0.08)**	0.782 (0.10)	**0.731 (0.13)**	**0.834 (0.07)**	**0.769 (0.07)**	**0.882 (0.08)**	0.552 (0.14)
B	0.795 (0.08)	**0.275 (0.14)**	0.746 (0.06)	0.751 (0.08)	**0.747 (0.15)**	0.800 (0.07)	0.745 (0.07)	0.731 (0.09)	**0.780 (0.11)**
BP_{fs}	0.812 (0.08)	0.270 (0.14)	0.749 (0.07)	**0.807 (0.10)**	0.638 (0.15)				

Table 2. Mean and standard deviation (in parentheses) of performance metrics (area under the receiver operator curve, AUC; accuracy, Acc; sensitivity, Sens; specificity, Spec and R^2 for the regression models) for the different Random Forest models using different feature sets across all cross-validation folds. Left and right halves are for the Regression and Binary tasks respectively. Bold faced text on AUCs indicates the best performing model or those that are statistically equivalent (via a Wilcox rank signed test, with a Bonferroni-corrected significance threshold of 0.05/5=0.01). Features sets describe combinations of (B) baseline model (age and *APOEε*4 carrier status), (P) Proteomics, (M) Metabolomics.

1.11 for PET images using PiB and AV45 tracers respectively. The mean number of years past baseline that a scan was taken was 3.07 and 2.97 years for training and validation cohorts respectively.

The use of imaging at non-baseline times assumes that differences between the baseline and time that the image was taken is relatively small (which may be reasonable assuming a slow rate of $A\beta_{1-42}$ accumulation), and that few individuals are close to the defined threshold for abnormality. If this assumptions do not hold, it is likely to worsen predictive performance, making this analysis somewhat conservative.

3 Results

3.1 Models utilizing protein levels accurately predict CSF positivity

We evaluated the ability of blood-based biomarkers to predict CSF $A\beta_{1-42}$ normal/abnormal status using RFs trained using different subsets of input variables, treating the modelling of CSF $A\beta_{1-42}$ as either a regression or binary task. Summaries of the performance metrics from the resulting models are shown in Table 2 with their corresponding ROC curves shown in Figure 1.

We observed strong overall predictive performance for both the regression and binary tasks within our cross-validation framework. All sets of features outperformed the base model of age and *APOEε*4 carrier status with BP based models leading to the highest AUC in both tasks of 0.84 and 0.83 for the regression and binary tasks respectively. Standard deviation for the AUC was relatively high (7-8%), likely due to the noise inherent in both the analytes being used for prediction as well as in the CSF $A\beta_{1-42}$ measurements.

The BP models resulted in a mean R^2 of 0.29 for the regression task. The automatically derived threshold for this regression RF yielded a mean accuracy of 0.77, with a sensitivity of 0.78 and a specificity of 0.73. Across the 100 cross-validation runs, the chosen threshold ranged from 164pg/ml to 194pg/ml with a median of 185pg/ml.

Similar AUC and accuracy could be observed for learning the dichotomized CSF labels directly (e.g AUC 0.83, accuracy 0.77 for the BP model). For the binary task, a slight drop in both AUC, as well as an altered trade-off between sensitivity and specificity, was observed across all different feature sets compared to the regression task. Given this, we chose to focus on the regression model for much of the follow-up analysis.

While all sets of features outperformed the base model, models that made use of the protein level measurements consistently achieved the strongest predictive performance, whereas metabolites appeared to be of limited utility. In both the regression and binary tasks, models containing metabolites and proteomic data (BPM) achieved equivalent or worse AUCs than models containing only the proteomic data (BP). Furthermore, we observed that the use of the base features and metabolites alone (BM) lead to decreased performance compared to the baseline model, indicating that the set of measured metabolites may have contributed little predictive information or may have been too noisy to be useful for predicting CSF status. These findings are in contrast to the utility of metabolites in predicting PET $A\beta_{1-42}$ positivity[22].

While the results presented in this section include clinical diagnosed AD individuals, who are almost all CSF $A\beta_{1-42}$ positive, it is worth considering only 'pre-clinical' individuals as this may be more relevant for selective screening in drug trials. Evaluating our model's performance on CN and MCI individuals only, we find that similarly strong predictive performance can be obtained (Supplementary Table 4, Supplementary Figure 1, AUC 0.80, Acc 0.77 for the BP model) supporting our primary findings that plasma protein levels can be utilized to predict amyloid pathology status.

To ensure that our imputation procedure did not bias our results, we also built similar models using only complete cases after applying more stringent QC (removal of plasma analytes where more than 1% of measurements were missing), obtaining similar AUCs of 0.81 for the regression and binary tasks (Supplementary Figure 2).

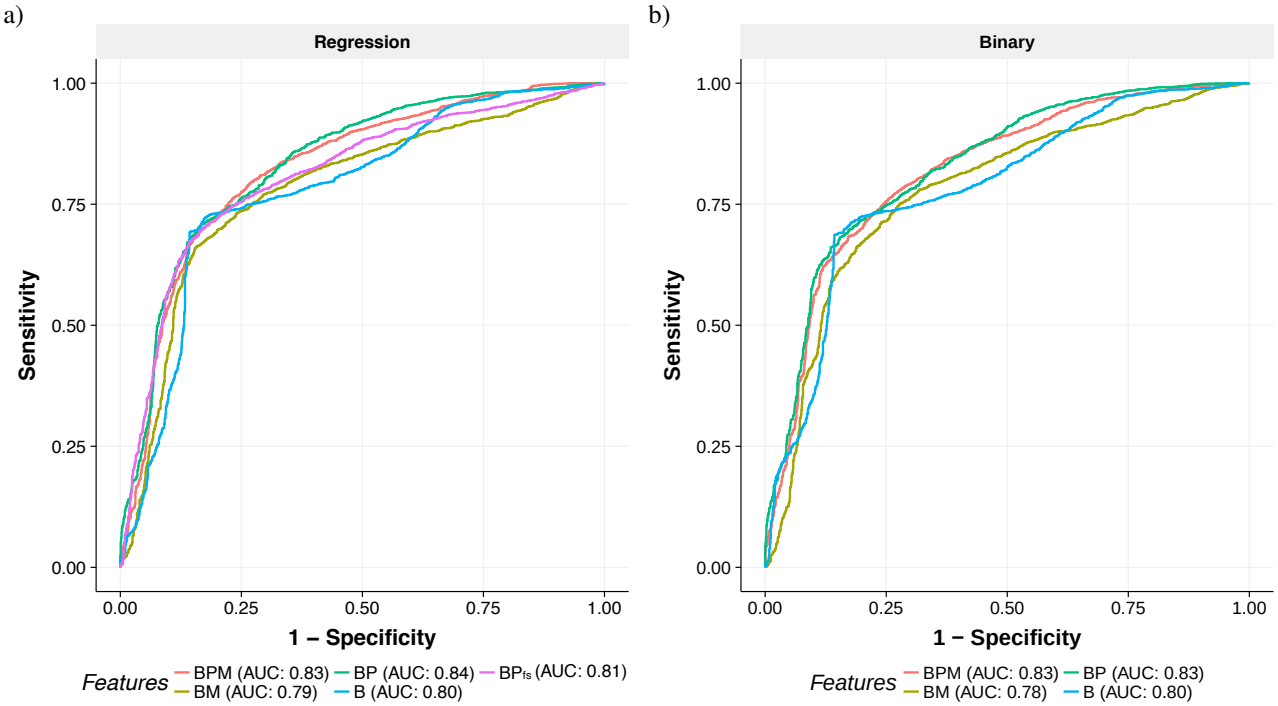

Figure 1. ROC curves comparing different sets of features to determine predictive value for the a) regression, and b) binary tasks. Different colours of lines correspond to different feature sets,(B) baseline model (age and $APOE\varepsilon4$ carrier status), (P) Proteomics, (M) Metabolomics, with corresponding AUCs indicating in the legend under each plot. The model BP_{fs} in subplot a) indicates the performance of the RF using feature selection.

3.2 Strong predictive performance is maintained using only four proteins

The models described so far used all (> 140) available features in this dataset. In practice, measuring hundreds of analytes is costly, negating a key advantage of using blood biomarkers for screening. Given this, we have applied feature selection to the BP regression RF, the most parsimonious model while achieving high predictive performance, to identify the smallest number of features that still achieved high predictive performance. Within cross-validation, we find that the average performance of this feature selection approach, denoted BP_{fs}, yields an AUC, sensitivity and specificity of 0.81, 0.81 and 0.63. The number of features selected in the model ranges from 2 to 15, with a median of 5 features included.

When applying this feature selection procedure to the entire set of training data, we identified a subset of four plasma analytes as well as APOE4 genotype status critical for model performance: Chromogranin-A (CGA), $\text{A}\beta_{1-42}$ (AB42), Eotaxin 3, and Apolipoprotein E (APOE). This combination of protein levels, together with $APOE\varepsilon4$ is denoted as BP5. Figure 2 indicates how each variable influences the model predictions after we have accounted for the influence of all other variables. As expected, the strongest relationship with CSF $\text{A}\beta_{1-42}$ is with $APOE\varepsilon4$ carrier status, where being a carrier ($APOE\varepsilon4 = 1$) leads to a low predicted $\text{A}\beta_{1-42}$ level. While the relationships between the proteins and CSF $\text{A}\beta_{1-42}$ are non-linear (expected given the nature of RFs), the overall correlation with CSF $\text{A}\beta_{1-42}$ is positive for CGA, Plasma $\text{A}\beta_{1-42}$, and APOE protein levels and negative for Eotaxin 3.

3.3 Validation of clinical utility

To demonstrate the utility of our modeling on unseen data, we conducted a survival analysis over the validation cohort (n=198), evaluating the probability of baseline MCI individuals transitioning to AD diagnosis over 120 months, stratified by predicted CSF $A\beta_{1-42}$ status from either the B, BP or reduced BP5 model. These survival distributions could then be compared to those of the real $\text{A}\beta_{1-42}$ status observed over the training cohort. Given the demographic similarity of the two cohorts, we would expect to see strong similarities in rates of conversion.

From Figure 3, we observed that in all cases, the predicted low CSF $\text{A}\beta_{1-42}$ group transitioned to AD significantly faster than the $\text{A}\beta_{1-42}$ normal group. Comparing the predictions from the BP and BP5 models on the validation cohort to the actual

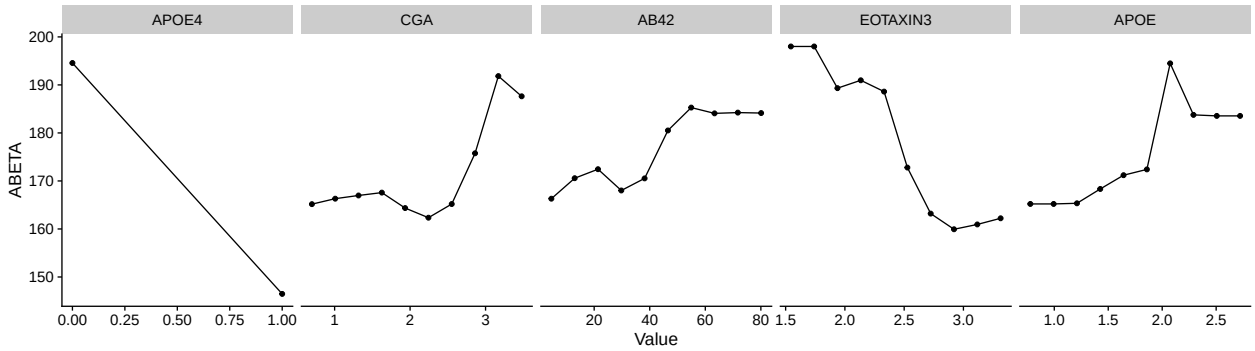

Figure 2. Partial dependency plots of the five features selected from the full BP model using a recursive feature elimination approach. Each subplot shows how the variation of a specific feature, assuming the other four features are fixed, impacts that predicted level of CSF Aβ_{1-42}.

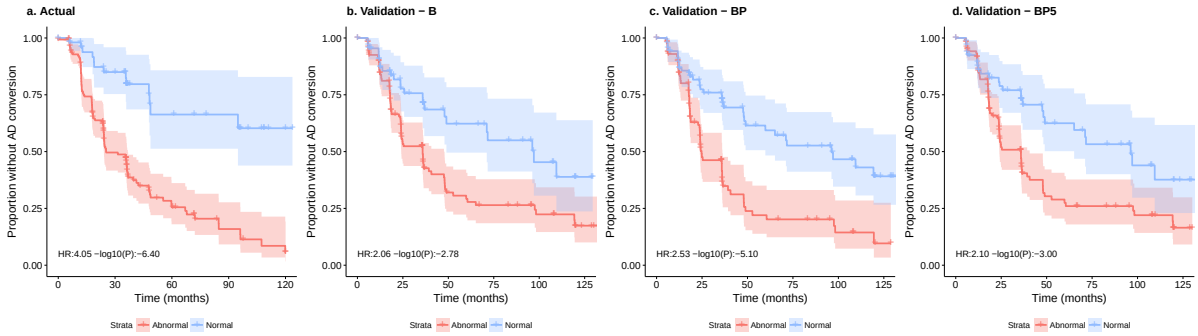

Figure 3. Kaplan-Meier curves for a) the training cohort stratified by actual CSF Aβ_{1-42} status, and the validation cohort stratified by predicted CSF Aβ_{1-42} status from the b) B model, c) BP model and d) BP5 model. The transparent ribbons along the curves represent the 95% confidence intervals. Hazards ratios and 95% confidence intervals for the abnormal group compared to the normal are shown in the bottom left of each subplot. In all cases, the low CSF Aβ_{1-42} group transitioned to AD diagnosis significantly faster then the normal group ($p = 3.97\times10^{-7}, 7.89\times10^{-6}, 9.96\times10^{-4}, 1.65\times10^{-3}$ for the four plots left-right). CN individuals were not included in this analysis given that there were no CN individuals present in the validation cohort.

CSF Aβ_{1-42} status on the training cohort, we find that there is no significant difference between the survival distributions for either the normal (log-rank test $p = 0.19, 0.2, 0.21$) or abnormal (log-rank test $p = 0.97, 0.31, 0.23$) survival distributions, respectively, reflecting the overlapping confidence intervals of the hazard ratios. However, it can be observed that due to differences in the thresholding of the Aβ_{1-42} levels, fewer individuals are deemed as CSF Aβ_{1-42} 'normal' in the actual data (n=53), compared with any of the three models applied to the validation datasets (n=95, 73, and 71 for BP, BP5, and B models respectively), highlighting the well-recognized issues of defining standardized cutoff values across studies[32]. The significant differences in conversion rates between the predicted normal/low strata, especially from the more parsimonious BP5 model, together with their similarity to the survival distributions of the actual CSF measures, provide strong evidence that our blood-based biomarkers have similar utility to actual CSF Aβ_{1-42} levels, given the similar properties of the two cohorts (Table 1).

3.4 Concordance with PET Aβ_{1-42} status

To further validate and quantify our model's performance, we have explored the relationship between the predicted CSF Aβ_{1-42} scores and PET imaging status. While there are difference between CSF and PET Aβ_{1-42} status (such as indications that CSF Aβ_{1-42} pathology can be detected before PET Aβ_{1-42}[13], the two biomarkers are correlated, with Aβ_{1-42} status for the first PET image only differing from the baseline CSF Aβ_{1-42} status in seven individuals. As such, evaluating our model against the PET Aβ_{1-42} status should provide a conservative estimate for the AUC on the validation cohort, despite the lack of CSF measures.

The resulting ROC curves in Figure 4 provide further evidence that the BP and BP5 models are able to predict Aβ_{1-42}

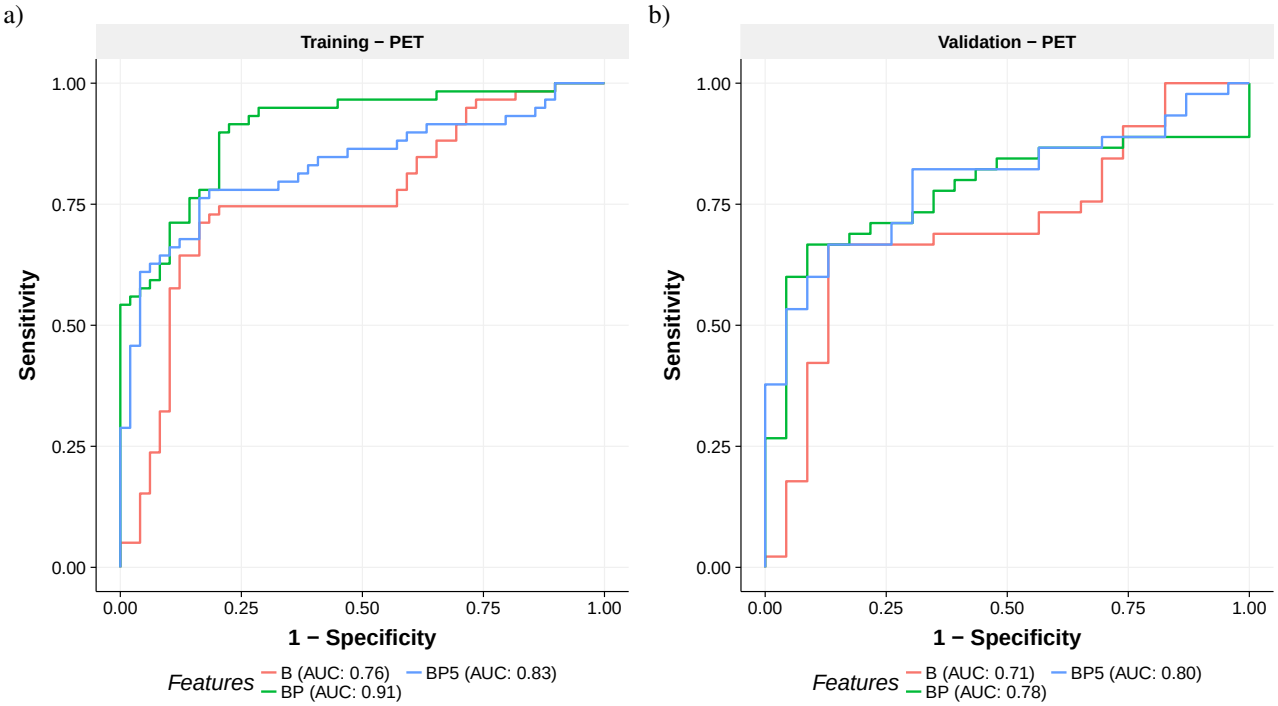

Figure 4. ROC curves comparing different sets of features to determine PET-based Aβ_{1-42} status on the a) training and b) validation cohorts. Different colours of lines correspond to different feature sets with corresponding AUCs indicating in the legend under each plot. Results in the training cohort are more useful as a measure of similarity between the tasks of predicting CSF and PET Aβ_{1-42} status given that this was the data used to train the CSF model and hence the AUC are upwardly biased, especially for more complex models (e.g. BP).

status, with AUCs on the validation of 0.78 and 0.8 for the BP and BP5 models respectively. These results are similar to those for CSF status from the training data, with a small expected drop due to the inherent differences between the two tasks. Interestingly, we observe stronger performance for the reduced BP5 model compared to the full BP model, with both again significantly improved compared to the baseline model of age and *APOE*ε4 status.

4 Discussion

The most positive results from AD trials to date have been found in patients with early forms of the disease, leading to an increasing awareness that treatments are likely to be most successful if applied at the earliest stages of AD[10]. While some AD clinical trials are enriching pre-symptomatic AD individuals with PET screening, recent findings that shifts in CSF amyloid can be observed up to a decade before those from PET may indicate that CSF positive individuals are more suitable for clinical trial enrichment [33]. Direct measurement of CSF biomarkers is too invasive to be used in such a screening test[34]. However, the development of a minimally-invasive, low-cost solution that provides the same type of information would overcome these limitations.

This current study evaluates the utility of a blood-based signature of CSF Aβ_{1-42} status using a Random Forest approach. We demonstrated that CSF Aβ_{1-42} normal/abnormal status using age, *APOE*ε*4* carrier status and protein levels can be predicted with a high AUC, sensitivity and specificity of 0.84, 0.78 and 0.73 respectively. Compared to the base model (age and *APOE*ε4 genotype) the inclusion of the plasma analytes improved the performance (AUC) by 6%. To make the model more suitable for clinical application, we identified four plasma analytes which, together with *APOE*ε*4* carrier status, still achieved a high AUC, sensitivity and specificity of 0.81, 0.81 and 0.64 respectively. These predictive models were then validated on a separate cohort of individuals to demonstrate that MCI subjects with predicted abnormal CSF Aβ_{1-42} (low) levels transitioned to an AD diagnosis at a significantly higher rate than those predicted with normal CSF Aβ_{1-42} levels. Furthermore, these rates were similar to those observed in a demographically similar cohort of MCIs using actual CSF Aβ_{1-42} levels. This is a strong

validation of our modelling as the blood-based biomarkers for CSF Aβ_{1-42} status is only useful if they can replicate the behavior of the actual Aβ_{1-42} status for clinically relevant endpoints for individuals that were not used to build the predictive model. Strong predictive power of PET Aβ_{1-42} status on the validation cohort provides further evidence for the generalizability and robustness of our modeling.

A number of studies have previously investigated the use of blood analytes to predict the burden of amyloid in the neocortex, as measured by PET [15,16,18–20,22,23]. Some of these studies showed similar performance metrics to those reported in this work ($>$ 0.80 AUCs[15,23–25] or $>$ 0.78 accuracy[17]), indicating that prediction of PET and CSF $A\beta_{1-42}$ status are of similar difficulty. PET Aβ is directly related to brain fibrillar amyloid, whereas CSF amyloid is a marker of soluble Aβ_{1-42} and they may, therefore, give different insights into AD progression and mechanisms. For example, CSF Aβ_{1-42} has been shown to be associated with *APOEε4* whereas PET has been shown to have a greater association with tau[35]. Thus, the development of a blood-based screening test for CSF Aβ_{1-42} levels is a complementary approach to existing blood-based biomarkers of PET amyloid status.

Of the above studies, the study by Nakamura et. al.[25] not only showed a very high AUC in a discover and independent validation dataset for PET Aβ_{1-42} status (AUC 0.94 and 0.96 respectively), but also showed strong performance for predicting abnormalities in CSF Aβ_{1-42} levels (AUC 0.88%), in a small cohort (n=45) of their validation set. While these results are promising, the automation of the novel technique used (IP-MALDI-TOF-MS), and hence transfer to a clinician setting, is non-trivial, motivating the search for complementary approaches. The protein signature presented in this study, based on a multiplex immunoassay, is likely to require a far shorter timeframe for clinical translation given the high level of automation that already exists for multiplex immunoassays, and that biomarkers from such platform have already been used in commercially available diagnostic tests that have been approved by the FDA

The use of metabolites appeared to be of limited utility for predicting CSF Aβ_{1-42}. In both the regression and binary tasks, models containing metabolites achieved equivalent or worse AUCs than models without.These findings can be contrast with the utility of metabolites in predicting PET Aβ_{1-42} positivity[22] and their association with AD more broadly[36]. Alternative methods for integrating this source of data[37], may be required in order to find robust associations with CSF Aβ_{1-42} status.

The subset of features found to be critical for our BP5 Random Forest model included *APOEε4* genotype, Chromogranin-A (CGA), Eotaxin 3, Aβ_{1-42} (AB42), C-reactive protein (CRP) and plasma Apolipoprotein E levels (APOE). Several of these identified proteins have known associations with Alzheimer's disease.Unsurprisingly, the levels of plasma APOE are associated with CSF amyloid levels. *APOEε4* is the strongest genetic risk factor for AD. APOE is involved in the clearance of Aβ_{1-42}[38–40] and there is a strong relationship between *APOEε4* genotype and APOE plasma levels, where *APOEε4* carriers have lower plasma levels[41,42]. Plasma Aβ_{1-42} showed a positive relationship in our model for CSF Aβ_{1-42}, in line with a prior observation[43]. This is interesting, as the link between alterations of Aβ_{1-42} levels in the blood and the progression of the disease is still controversial and studies assessing the Aβ_{1-42} concentration in blood of AD patients have produced conflicting results[43–49]. Chromogranin A (CGA) is associated with synaptic function, and has traditionally been used as an indicator of neuroendocrine tumors[50]. More recent work has shown that CGA has a degree of co-localisation with amyloid plaques in the brain[51,52]. However, levels of CGA in the CSF and blood serum do not appear to be correlated[53], and serum CGA has not previously been linked to AD. Eotaxin 3, also known as C-C chemokine ligand 26 (CCL26), plays an important role in the innate immune system and has been found to be dysregulated in AD patients[54]. CSF Eotaxin 3 has been shown to be significantly elevated in patients with prodromal AD, however, Eotaxin 3 levels in plasma or the CSF do not correlate with rates of disease progression[54,55]

This study has several limitations. The training and validation cohorts are both part of the ADNI study and thus all measures were conducted on the same platforms. Hence further cross-cohort and cross-platform replication is required. This remains an ongoing issue within the development of all AD biomarkers relating to early screening and requires significant future investment[56]. Furthermore, the current cohort is neuropathology biased, i.e. 84% of the cohort have MCI or AD, and thus likely to have neuronal damage, potentially confounding the analysis of CSF Aβ_{1-42} status. Finally, it needs to be noted that there are other medical conditions that are known to affect CSF Aβ_{1-42} levels and it is unclear whether these affect any of the patients in our cohort.

The early identification of AD disease is paramount and a major global focus as the success of disease-modifying or preventative therapies in AD may depend on detecting the earliest signs of abnormal amyloid-beta load. The differences between CSF Aβ_{1-42} and PET Aβ_{1-42} in preclinical stages of AD could also have implications in clinical trial settings. Blood-based biomarkers of amyloid can serve as the first step in a multistage screening procedure, similar to those that have been clinically-implemented in cancer, cardiovascular disease, and infectious diseases[56]. In-conjunction with biomarkers for neocortical amyloid burden, the CSF Aβ_{1-42} biomarkers presented in this work may help yield a cheap, non-invasive tool for both improving clinical trials targeting amyloid and population screening.

5 Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. This work was supported by IBM. We would like to thank Dr. Matthew Downton and Ms. Anna Trigos for helpful feedback on the manuscript.

6 Conflict of Interest

The authors declare no conflict of interest.

7 Author Contributions Statement

B.G, B.F, and N.F. designed the study, B.G. and B.F. analysed the data and ran all experiments, B.G. made the figures, B.G., C.S. and N.F. wrote the manuscript. All authors interpreted the data and critically revised the manuscript.

References

1. Villemagne, V. L. *et al.* Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *The Lancet. Neurol.* **12**, 357–367 (2013).

2. Berti, V. *et al.* Rethinking on the concept of biomarkers in preclinical alzheimer's disease. *Neurol. Sci.* **37**, 663–672 (2016).

3. Dubois, B. *et al.* Revising the definition of Alzheimer's disease: a new lexicon. *The Lancet. Neurol.* (2010).

4. Dubois, B. *et al.* Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia* **12**, 292–323 (2016).

5. Alzheimers Association. 2017 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **13**, 325–373 (2017).

6. Cummings, J. L., Morstorf, T. & Zhong, K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimer's research & therapy* **6**, 37 (2014).

7. Cummings, J., Lee, G., Mortsdorf, T., Ritter, A. & Zhong, K. Alzheimer's disease drug development pipeline: 2017. *Alzheimer's & Dementia* **3**, 367–384 (2017).

8. Salloway, S. *et al.* Two phase 3 trials of bapineuzumab in mild-to-moderate alzheimer's disease. *New Engl. J. Medicine* **370**, 322–333 (2014).

9. Sevigny, J. *et al.* Amyloid pet screening for enrichment of early-stage alzheimer disease clinical trials: Experience in a phase 1b clinical trial. *Alzheimer Dis. & Assoc. Disord.* **30**, 1–7 (2016).

10. Godyń, J., Jończyk, J., Panek, D. & Malawska, B. Therapeutic strategies for alzheimer's disease in clinical trials. *Pharmacol. Reports* **68**, 127–138 (2016).

11. Scheltens, P. *et al.* Alzheimer's disease. *The Lancet* **388**, 505–517 (2016).

12. Blennow, K., Mattsson, N., Schöll, M., Hansson, O. & Zetterberg, H. Amyloid biomarkers in alzheimer's disease. *Trends Pharmacol. Sci.* **36**, 297–309 (2015).

13. Palmqvist, S., Mattsson, N., Hansson, O. & Neuroimaging, A. D. Cerebrospinal fluid analysis detects cerebral amyloid-beta accumulation earlier than positron emission tomography. *Brain* **139**, 1226–1236 (2016).

14. Rabinovici, G. D. Amyloid biomarkers: pushing the limits of early detection. *Brain* **139**, 1008–1010 (2016).

15. Burnham, S. C. *et al.* A blood-based predictor for neocortical Aβ burden in Alzheimer's disease: results from the AIBL study. *Mol. Psychiatry* **19**, 519–526 (2014).

16. Burnham, S. C. *et al.* Predicting Alzheimer disease from a blood-based biomarker profile: A 54-month follow-up. *Neurol.* **87**, 1093–1101 (2016).

17. Voyle, N. *et al.* Blood Protein Markers of Neocortical Amyloid-β Burden: A Candidate Study Using SOMAscan Technology. *J. Alzheimers Dis.* **46**, 947 (2015).

18. Ashton, N. J. *et al.* Blood protein predictors of brain amyloid for enrichment in clinical trials? *Alzheimer's & Dementia* **1**, 48–60 (2015).

19. Kiddle, S. J. *et al.* Plasma Based Markers of [C-11] PiB-PET Brain Amyloid Burden. *PLoS One* **7** (2012).

20. Westwood, S. *et al.* Blood-Based Biomarker Candidates of Cerebral Amyloid Using PiB PET in Non-Demented Elderly. *J. Alzheimers Dis.* **52**, 561–572 (2016).

21. Thambisetty, M. *et al.* Proteome-Based Plasma Markers of Brain Amyloid-beta Deposition in Non-Demented Older Individuals. *J. Alzheimers Dis.* **22**, 1099–1109 (2010).

22. Voyle, N. *et al.* Blood metabolite markers of neocortical amyloid-β burden: Discovery and enrichment using candidate proteins. *Transl. Psychiatry* **6**, e719 (2016).

23. Gupta, V. B. *et al.* Follow-up plasma apolipoprotein e levels in the australian imaging, biomarkers and lifestyle flagship study of ageing (aibl) cohort. *Alzheimer's research & therapy* **7**, 16 (2015).

24. Haghighi, M., Smith, A., Morgan, D., Small, B. & Huang, S. Identifying cost-effective predictive rules of amyloid-β level by integrating neuropsychological tests and plasma-based markers. *J. Alzheimers Dis.* **43**, 1261–1270 (2015).

25. Nakamura, A. *et al.* High performance plasma amyloid-β biomarkers for Alzheimer's disease. *Nat.* **554** (2018).

26. Ishwaran, H. & Kogalur, U. Randomforestsrc: Random forests for survival, regression and classification (rf-src). *R package version* **1** (2014).

27. Wright, M. N. & Ziegler, A. ranger: A fast implementation of random forests for high dimensional data in c++ and r. *J. Stat. Softw.* **77** (2017).

28. Ishwaran, H., Kogalur, U. B., Chen, X. & Minn, A. J. Random survival forests for high-dimensional data. *Stat. Analysis Data Mining: The ASA Data Sci. J.* **4**, 115–132 (2011).

29. Cawley, G. C. & Talbot, N. L. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **11**, 2079–2107 (2010).

30. López-Ratón, M., Rodríguez-Álvarez, M. X., Cadarso-Suárez, C., Gude-Sampedro, F. *et al.* Optimalcutpoints: an r package for selecting optimal cutpoints in diagnostic tests. *J Stat Softw* **61**, 1–36 (2014).

31. Breiman, L. Random forests. *Mach. learning* **45**, 5–32 (2001).

32. Blennow, K. *et al.* Clinical utility of cerebrospinal fluid biomarkers in the diagnosis of early alzheimer's disease. *Alzheimer's & dementia: journal Alzheimer's Assoc.* **11**, 58–69 (2015).

33. Palmqvist, S., Mattsson, N. & Hansson, O. Reply: Do we still need positron emission tomography for early alzheimer's disease diagnosis? *Brain* **139**, e61–e61 (2016).

34. Hane, F. T. *et al.* Recent progress in alzheimer's disease research, part 3: Diagnosis and treatment. *J. Alzheimer's Dis.* 1–21 (2017).

35. Mattsson, N. *et al.* Independent information from cerebrospinal fluid amyloid-β and florbetapir imaging in Alzheimer's disease. *Brain* **138**, 772–783 (2015).

36. Toledo, J. B. *et al.* Metabolic network failures in alzheimer's disease—a biochemical road map. *Alzheimer's & Dementia* (2017).

37. Wang, B. *et al.* Similarity network fusion for aggregating data types on a genomic scale. *Nat. Met.* **11**, 333–337 (2014).

38. Jiang, Q. *et al.* ApoE Promotes the Proteolytic Degradation of A? *Neuron* **58**, 681–693 (2008).

39. Verghese, P. B. *et al.* ApoE influences amyloid-β (Aβ) clearance despite minimal apoE/Aβ association in physiological conditions. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E1807–16 (2013).

40. Bachmeier, C. *et al.* A multifaceted role for apoE in the clearance of beta-amyloid across the blood-brain barrier. *Neurodegener. Dis.* **11**, 13–21 (2013).

41. Simon, R. *et al.* Total ApoE and ApoE4 isoform assays in an Alzheimer's disease case-control study by targeted mass spectrometry (n=669): a pilot assay for methionine-containing proteotypic peptides. *Mol. & cellular proteomics : MCP* **11**, 1389–1403 (2012).

42. Martínez-Morillo, E. *et al.* Total apolipoprotein E levels and specific isoform composition in cerebrospinal fluid and plasma from Alzheimer's disease patients and controls. *Acta Neuropathol. (Berl.)* **127**, 633–643 (2014).

43. Janelidze, S. *et al.* Plasma β-amyloid in Alzheimer's disease and vascular disease. *Sci. Rep.* **6**, 26801 (2016).

44. Lewczuk, P. *et al.* Amyloid beta peptides in plasma in early diagnosis of Alzheimer's disease: A multicenter study with multiplexing. *Exp. Neurol.* **223**, 366–370 (2010).

45. Lui, J. K. *et al.* Plasma amyloid-beta as a biomarker in Alzheimer's disease: the AIBL study of aging. *J. Alzheimers Dis.* **20**, 1233–1242 (2010).

46. Rembach, A. *et al.* Changes in plasma amyloid beta in a longitudinal study of aging and Alzheimer's disease. *Alzheimer's & Dementia* **10**, 53–61 (2014).

47. Park, J.-C. *et al.* Chemically treated plasma Aβ is a potential blood-based biomarker for screening cerebral amyloid deposition. *Alzheimer's research & therapy* **9**, 20 (2017).

48. Pesaresi, M. *et al.* Plasma levels of beta-amyloid (1-42) in Alzheimer's disease and mild cognitive impairment. *Neurobiol. Aging* **27**, 904–905 (2006).

49. Tzen, K.-Y. *et al.* Plasma Aβ but not tau is related to brain PiB retention in early Alzheimer's disease. *ACS Chem. Neurosci.* **5**, 830–836 (2014).

50. Campana, D. *et al.* Chromogranin A: is it a useful marker of neuroendocrine tumors? *J Clin Oncol* **25**, 1967–1973 (2007).

51. Lechner, T. *et al.* Chromogranin peptides in alzheimer's disease. *Exp. gerontology* **39**, 101–113 (2004).

52. Brion, J.-P., Couck, A.-M., Bruce, M., Anderton, B. & Flament-Durand, J. Synaptophysin and chromogranin a immunoreactivities in senile plaques of alzheimer's disease. *Brain research* **539**, 143–150 (1991).

53. Blennow, K. *et al.* Tau protein in cerebrospinal fluid. *Mol. Chem. Neuropathol.* **26**, 231–245 (1995).

54. Huber, A. K., Giles, D. A., Segal, B. M. & Irani, D. N. An emerging role for eotaxins in neurodegenerative disease. *Clin. Immunol.* (2016).

55. Westin, K. *et al.* Ccl2 is associated with a faster rate of cognitive decline during early stages of alzheimer's disease. *PLoS One* **7**, e30525 (2012).

56. O'Bryant, S. E. *et al.* Blood-based biomarkers in alzheimer disease: Current state of the science and a novel collaborative paradigm for advancing from discovery to clinic. *Alzheimer's & Dementia* **13**, 45–58 (2017).