Journal of Alzheimer's Disease xx (20xx) x–xx
DOI 10.3233/JAD-150570
IOS Press


# Accurate Prediction of Conversion to Alzheimer's Disease using Imaging, Genetic, and Neuropsychological Biomarkers

Juergen Dukart[a,b,*], Fabio Sambataro[a], Alessandro Bertolino[a,c] and for the Alzheimer's Disease Neuroimaging Initiative[1]
[a]*F. Hoffmann-La Roche, Roche Innovation Centre Basel, Basel, Switzerland*
[b]*Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[c]*Department of Basic Medical Science, Neuroscience and Sense Organs, University of Bari, Bari, Italy*

Accepted 2 October 2015

**Abstract**. A variety of imaging, neuropsychological, and genetic biomarkers have been suggested as potential biomarkers for the identification of mild cognitive impairment (MCI) in patients who later develop Alzheimer's disease (AD). Here, we systematically evaluated the most promising combinations of these biomarkers regarding discrimination between stable and converter MCI and reflection of disease staging. Alzheimer's Disease Neuroimaging Initiative data of AD ($n = 144$), controls ($n = 112$), stable ($n = 265$) and converter ($n = 177$) MCI, for which apolipoprotein E status, neuropsychological evaluation, and structural, glucose, and amyloid imaging were available, were included in this study. Naïve Bayes classifiers were built on AD and controls data for all possible combinations of these biomarkers, with and without stratification by amyloid status. All classifiers were then applied to the MCI cohorts. We obtained an accuracy of 76% for discrimination between converter and stable MCI with glucose positron emission tomography as a single biomarker. This accuracy increased to about 87% when including further imaging modalities and genetic information. We also identified several biomarker combinations as strong predictors of time to conversion. Use of amyloid validated training data resulted in increased sensitivities and decreased specificities for differentiation between stable and converter MCI when amyloid was included as a biomarker but not for other classifier combinations. Our results indicate that fully independent classifiers built only on AD and controls data and combining imaging, genetic, and/or neuropsychological biomarkers can more reliably discriminate between stable and converter MCI than single modality classifiers. Several biomarker combinations are identified as strongly predictive for the time to conversion to AD.

Keywords: Florbetapir, [$^{18}$F]fluorodeoxyglucose positron emission tomography, mild cognitive impairment, structural magnetic resonance imaging



## INTRODUCTION

Alzheimer's disease (AD) is a complex disorder of deteriorating cognition with multiple known neuropathological mechanisms which include amyloid-β (Aβ) and tau deposition and neurodegeneration. Numerous genetic and nongenetic risk factors of this

[1]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. 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.

*Correspondence to: Juergen Dukart, PhD, Biomarkers & Clinical Imaging, NORD DTA, F. Hoffmann-La Roche, Grenzacherstrasse 170, 4070 Basel, Switzerland. Tel.: +41 61 68 77091; Fax: +41 61 68 79848; E-mail: juergen.dukart@gmail.com.

ISSN 1387-2877/16/$35.00 

neuropathology such as apolipoprotein E (APOE) genotype, neuropsychological measures, and *in vivo* measures of atrophy, glucose utilization, and amyloid depositions have been identified in studies on AD [1–6]. Considering several of these biomarkers have been shown to be a promising way for improving diagnostic accuracy, researchers are now integrating them into the revised diagnostic criteria for AD [7, 8]. However, the understanding on which biomarkers provide an additive value when combined with others is rather limited. This applies even more for the early stages of the disease.

A large proportion of patients with amnestic mild cognitive impairment (MCI) are now considered to represent an early AD stage [7, 9]. A series of studies have been performed with the aim of increasing the diagnostic accuracy in MCI. Whilst most studies have focused on single biomarkers [10–17], multiple studies have also applied machine learning algorithms to compare biomarkers and their combinations with the intent of capturing different aspects of the complex pathophysiology of AD [18–27]. A consistent finding across the multimodal studies is increased accuracy ranging between 60 and 90 % for discrimination between stable mild cognitive impairment (sMCI) and MCI converting to AD (cMCI) when information from different biomarkers is combined. However, none of these studies systematically evaluated the additive value of all of these biomarkers and their combinations in the same MCI population. Furthermore, applications of extensive parameter optimization procedures to increase cross-validation performance might have led to an overestimate of accuracies achievable for new data – a problem that is commonly referred to as overfitting. Accuracies reported when performing a strict separation of training and testing data, which is considered as the gold standard of machine learning, are typically lower, ranging below 80% [20, 27, 28]. A further aspect that has been commonly shown is that classifiers trained on AD and healthy controls can be applied to reliably discriminate between cMCI and sMCI. Another common limitation of most of the above mentioned studies is the use of non-histopathologically validated training cohorts to establish the classifiers. The known limited accuracy of clinical diagnoses may lead to the inclusion of other dementias in the AD groups or of preclinical AD as healthy controls [29, 30]. Both could reduce the capability of the classifiers to discriminate between new AD and control cases. While there are still no sufficiently large histopathologically confirmed datasets available for most of the biomarkers, novel amyloid positron emission tomography tracers provide a close *in vivo* approximation of the corresponding AD histopathology [31]. Thus, using this information to identify AD and control training cases may further increase accuracies reported for different biomarkers.

A further aspect neglected in previous studies is the sensitivity of identified biomarkers to disease staging. Earlier studies have mostly focused on the categorical question of conversion versus non-conversion, without evaluating if the identified biomarkers also reflect disease staging as indicated, for example by the time to conversion to AD (TTC). This aspect might yet be essential to monitor progression in clinical trials focusing on early disease stages and because potential treatment is considered to be more beneficial for patients when loss of function is not yet strongly advanced.

Given that genetic risk, deterioration of cognition, Aβ deposition, and brain structural and functional biomarkers contribute to the diagnosis of AD, we systematically evaluated the potential of combinations of these factors to accurately stratify the MCI population according to risk of conversion to AD and disease staging. We hypothesize that a combination of biomarkers covering several genetic, behavioral, and neuropathological factors will provide higher sensitivity for early AD detection and disease staging as compared to best performing single modality biomarker. Further, we hypothesize that the use of only amyloid negative healthy controls and amyloid positive AD for training the classifiers will further improve the discrimination accuracies for cMCI and sMCI.

## METHODS

### *Subjects*

All available ADNI1, ADNI-GO and ADNI2 (ADNI: Alzheimer's Disease Neuroimaging Initiative) data as of December 2013, of AD, healthy control subjects (HC), amnestic sMCI and cMCI having APOE genotype and neuropsychological evaluation were included in the study. Additionally, an imaging subcohort was identified from these data for which each of the following imaging biomarkers was available for at least one of the time points: Structural magnetic resonance imaging (sMRI), [$^{18}$F]fluorodeoxyglucose positron emission tomography (FDG-PET) and/or [$^{18}$F]AV45-PET (florbetapir) (Table 1). To avoid biases in accuracies due to use of different amyloid compounds, we restricted our analyses to AV45-PET as a tracer with greater availability in the ADNI database

[32]. For sMCI, an inclusion criterion of at least 2 y of follow-up was applied to ensure stability of the diagnosis over time. For cMCI, all three imaging modalities had to be available prior to or at conversion to AD. The final dataset for APOE and neuropsychology included data of 144 AD, 112 NL, 177 cMCI, and 265 sMCI, with overall 958 observations (number of subjects times number of visits) for MCI and 750 observations for AD and HC (Table 1, Supplementary Material 1).

Diagnosis of AD was based on National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria [33]. Imaging and genetic biomarkers evaluated in our study were not part of criteria used by the ADNI to establish diagnostic labels of MCI or AD. The study was conducted according to the Declaration of Helsinki. Written informed consent was obtained from all participants before protocol-specific procedures were performed.

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been 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 mild cognitive impairment (MCI) and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

Table 1
Clinical and demographic characteristics for training and testing data

| All data | Training data | | Testing data | | Statistical test |
|---|---|---|---|---|---|
| | HC | AD | sMCI | cMCI | (test value,df,p) |
| n | 112 | 144 | 265 | 177 | – |
| N observations | 471 | 279 | 657 | 301 | |
| Age (mean ± SD [range], y) | 74.4 ± 5.2 [62–96] | 75.4 ± 8.1 [55–92] | 74.6 ± 7.5 [48–89] | 75 ± 7.1 [55–89] | ANOVA (0.5, 3, 0.650) |
| Gender (male/female) | 55/57 | 84/60 | 170/95 | 107/70 | $\chi^2$ (7.6, 3, 0.056) |
| Education (mean ± SD, [range] y) | 16.4 ± 2.6 [10–20] | 15.2 ± 3.0 [6–20] | 15.8 ± 3.0 [7–20] | 15.7 ± 2.8 [6–20] | ANOVA (3.9, 3, 0.009) |
| MMSE (mean ± SD) | 29 ± 1.2 | 23.4 ± 2* | 27.7 ± 1.8 | 26.7 ± 1.7* | ANOVA (263.8, 3,<0.001) |
| GDS (mean ± SD) | 0.7 ± 1.0 | 1.7 ± 1.3* | 1.6 ± 1.3 | 1.6 ± 1.5 | ANOVA (16.7, 3,<0.001) |
| ADAS (mean ± SD) | 11.4 ± 4.7 | 34.3 ± 8.9* | 19.8 ± 7.0 | 24.9 ± 6.7* | ANOVA (263.8, 3,<0.001) |
| RAVLT | 5.9 ± 2.4 | 1.9 ± 1.8* | 4.3 ± 2.6 | 2.7 ± 2.2* | ANOVA (78.5, 3,<0.001) |
| RAVLT im | 44.0 ± 8.5 | 22.6 ± 7.0* | 34.3 ± 10.2 | 28.3 ± 7.5* | ANOVA (143.7, 3,<0.001) |
| FAQ | 0.1 ± 0.5 | 13.2 ± 6.8* | 2.7 ± 3.6 | 5.3 ± 4.7* | ANOVA (227.4,<0.001) |
| TTC (mean ± SD, y) | – | – | – | 2 ± 1.4 | – |
| **Imaging sub-cohort** | **Training data** | | **Testing data** | | **Statistical test** |
| | HC | AD | sMCI | cMCI | (test value, df, p) |
| N | 83 | 36 | 135 | 29 | – |
| N observations (sMRI/FDG/AV) | 359/208/112 | 63/37/36 | 354/223/164 | 73/53/31 | – |
| Age (mean ± SD [range], y) | 74.3 ± 5.1 [65–90] | 76 ± 8.3 [56–91] | 73.3 ± 7.5 [48–88] | 73 ± 8.1 [55–85] | ANOVA (1.6, 3, 0.185) |
| Gender (male/female) | 42/41 | 21/15 | 84/51 | 18/11 | $\chi^2$ (3.0, 3, 0.384) |
| Education (mean ± SD [range], y) | 16.5 ± 2.7 [10–20] | 15.2 ± 2.6 [9–20] | 15.7 ± 2.8 [8–20] | 16.3 ± 2.6 [9–20] | ANOVA (2.9, 3, 0.034) |
| MMSE (mean ± SD) | 29 ± 1.2 | 22.9 ± 2.1* | 28.1 ± 1.6 | 27 ± 1.7* | ANOVA (125.3, 3,<0.001) |
| GDS (mean ± SD) | 0.7 ± 1.0 | 1.6 ± 1.2* | 1.6 ± 1.2 | 1.7 ± 1.9 | ANOVA (11.4, 3,<0.001) |
| ADAS (mean ± SD) | 11.7 ± 4.5 | 37.3 ± 10.0* | 18.6 ± 6.9 | 23.3 ± 9.1* | ANOVA (113.7, 3,<0.001) |
| RAVLT | 5.6 ± 2.3 | 1.9 ± 1.8* | 4.9 ± 2.7 | 3.1 ± 2.2* | ANOVA (23.3, 3,<0.001) |
| RAVLT im | 43.9 ± 9.2 | 20.6 ± 6.1* | 36.6 ± 10.3 | 31.8 ± 8.1 | ANOVA (53.5, 3,<0.001) |
| FAQ | 0.1 ± 0.6 | 14.6 ± 7.3* | 2.3 ± 3.2 | 6.6 ± 5.0* | ANOVA (137.4, 3,<0.001) |
| TTC (mean ± SD, y) | – | – | – | 2.4 ± 2.1 | – |
| % amyloid positive | 23% | 92% | 36% | 83% | $\chi^2$ (69.1, 3,<0.001) |

AD, Alzheimer's disease; ADAS, Alzheimer's Disease Assessment Scale; AV, florbetapir positron emission tomography; df, degrees of freedom; FAQ, Functional Activities Questionnaire; FDG, fluorodeoxyglucose positron emission tomography; GDS, Geriatric Depression Scale; HC, healthy control subjects; cMCI, mild cognitive impairment converters to AD; GDS, Geriatric Depression Scale; MMSE, Mini Mental State Examination; RAVLT, Rey Auditory Verbal Learning Test delayed recall; RAVLT im, RAVLT immediate recall; sMCI, stable MCI; sMRI, structural magnetic resonance imaging; SD, standard deviation; TTC, time to conversion to Alzheimer's disease. *indicates significant differences in *post-hoc t*-tests relative to HC (for AD) and sMCI (for cMCI).

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects, but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow-up duration of each group is specified in the protocols for ADNI-1, ADNI-2, and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

*Demographic and neuropsychological measures*

Between-group differences in gender across all groups were evaluated using a chi-square test for independent samples. Analyses of variance ($p < 0.05$) and subsequent *post-hoc t*-tests ($p < 0.05$ Bonferroni corrected for multiple comparisons) were applied to evaluate differences in age, education, and neuropsychological scores. The following 6 neuropsychological scores were included into the classification analysis based on their availability for most of the subjects: Mini Mental State Examination (MMSE [34]), Geriatric Depression Scale (GDS [35]), Alzheimer's Disease Assessment Scale (ADAS [36]), Rey Auditory Verbal Learning Test – immediate and delayed recall (RAVLT immediate and RAVLT [37, 38]) and Functional Activities Questionnaire (FAQ [39]) (Table 1).

*Imaging data*

The MRI dataset included standard T1-weighted images obtained with different 1.5T and 3T scanner types using a three-dimensional magnetization-prepared rapid gradient-echo sequence varying in repetition time and echo time with in-plane resolution of 1.25*1.25 mm and 1.2 mm slice thickness. If both 1.5 and 3T data were available for the same time points only the 3T data were used. Overall, there was a significantly higher proportion of 3T data in the AD group in the training data ($p < 0.001$). There was no significant difference in distribution of scanner types between cMCI and sMCI ($p > 0.1$). All images were corrected for distortions and B1-field non-uniformities as described on the ADNI website (http://adni.loni.usc.edu/).

FDG-PET and AV45-PET data were downloaded at the most advanced pre-processing stage (excluding smoothing) provided by ADNI. In brief, the pre-processing provided by ADNI included a within subject co-registration and averaging of all PET frames from the same time-point, interpolation to 1.5 mm cubic voxels and global mean intensity normalization. Detailed description of this pre-processing pipeline can be found on the ADNI website (http://adni.loni.usc.edu/methods/pet-analysis/pre-processing/) listed under point 3. Though other intensity normalization procedures have been shown to be more sensitive for differentiation of AD and HC subjects using FDG-PET [40–42], the choice of an optimum reference region is less clear for AV45-PET. To avoid systematic differences in pre-processing between the two PET modalities, we restricted our analyses to global mean intensity normalization. Similarly, correction of PET data for partial volume effects using structural MRIs acquired at the same time points can also improve their sensitivity for AD detection [43–45]. However, appropriate correction for these effects would require structural data acquired at the same time points. Due to a relatively low availability of AV45 and sMRI data for the same time points, applying this correction would have resulted in exclusion of a significant proportion of available PET data. To avoid this data loss, correction for partial volume effects was not applied in this study.

*Pre-processing of imaging data*

Pre-processing of all imaging data was performed in SPM8 (Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/) implemented in Matlab 7.12 (MathWorks, Inc, Sherborn, MA, USA). The pre-processing pipeline consisted of co-registration of all imaging modalities for the same visits of each subject, segmentation of sMRI data using NewSegment, spatial normalization using diffeomorphic image registration (DARTEL) [46] with subsequent affine registration into the Montreal Neurological Institute (MNI) space and smoothing with a Gaussian kernel of 8 mm FWHM. To reduce computational time DARTEL template was computed from a random representative subsample of 300 scans. The obtained grey matter images were additionally modulated to preserve the total amount of signal from each region. All further analyses were restricted to a mask obtained by applying a probability threshold of 0.2 to

the first and the last DARTEL templates co-registered to the MNI space [47]. To reduce the dimensionality of imaging data for the Bayesian feature selection procedure described below all pre-processed images were downsampled to an isotropic resolution of 6 mm.

### *Feature selection*

To ensure that the features used from the different imaging modalities for subsequent classification are not biased by differences in general characteristics of the cohorts (e.g., demographic factors or disease severity) or image pre-processing (e.g., differential smoothing or spatial normalization) used to identify these features across different studies we adopted our own feature selection approach for the current study. All feature selection steps for imaging data were performed using a subset of AD and HC data. The subset was selected from the whole training dataset and included only the earliest time points for AD and HC for which all three imaging modalities were available (AD: $n = 38$, HC: $n = 93$). This selection step was performed to ensure that exactly the same subjects were used to identify the most relevant features across the three imaging modalities. To avoid a bias towards a specific modality (e.g., using only amyloid positive AD and negative HC), all AD and HC patients in this subset were used for feature selection.

Feature selection for imaging data was performed using a Bayesian Markov Blanket approach integrated in the Causal Explorer toolbox implemented in Matlab [48, 49]. In brief, the algorithm identifies features that are relevant for Bayesian separation between AD and HC subjects at a predefined statistical threshold. The setting for continuous data with conditioning set size of 0 was used for feature identification. A full Bonferroni corrected threshold of $p < 0.05$ was applied to identify most relevant sMRI and FDG-PET features. For AV45-PET this already conservative threshold resulted in a very high number of features covering the whole brain. To reduce the AV45-PET feature set to a comparable size as observed for FDG-PET and sMRI, a Bonferroni corrected threshold of $p < 0.000001$ was applied. The feature selection procedure resulted in identification of 13 clusters for FDG-PET, 13 clusters for AV45-PET, and 29 clusters for sMRI (Fig. 1, cluster images are provided in Supplementary Material 2). Mean values extracted from each of the identified clusters were used for subsequent classification. All cluster images will be published on nitrc.org upon acceptance of this manuscript.

### *Naïve Bayes Classification*

We used a Naïve Bayes (NB) classification algorithm, as implemented in Matlab 7.12 to evaluate the predictive accuracy of different genetic, neuropsychological, and imaging biomarkers for differentiation between cMCI and sMCI. In brief, the NB approach provides a probability for each new case to belong to a particular class based on frequencies for categorical and means and standard deviations for continuous features as observed in training data. Similarly to a clinician-based decision, the NB approach is considering all biomarkers as independent evidence for assignment to one of the diagnostic classes. A strong advantage of the NB classifier as compared to most other machine learning algorithms is its capability to deal with sparse, categorical, and continuous data and the posterior probability it provides for each new case to belong to a particular class. As the NB approach does not require any extensive parameter optimization, it also reduces the risk of overfitting the classifier to the training data.

NB classifiers were first built using all available AD and HC data separately for each of the modalities (APOE genotype, neuropsychological scores, AV45-PET, FDG-PET, and sMRI). In a further analysis, NB classifiers were then built for all possible combinations of imaging biomarkers with APOE genotype and neuropsychological profiles. For all classifiers, equal prior probability was set for AD and HC classes to avoid the risk that the classifier is biased by the differential numbers of training cases per class.

The obtained NB classifiers were then applied to MCI data having the same biomarker constellations. APOE genotype was treated as a categorical variable, while all other measurements were treated as continuous. Applying the NB classifiers to the MCI data resulted in one set of predicted labels for each biomarker constellation for the MCI subset having the corresponding biomarker measures. An assignment of sMCI as HC and of cMCI as AD was considered as correct. Balanced accuracies ((sensitivity+specificity)/2), sensitivities, specificities, receiver operating characteristics (ROC) curves, and the area under the curve (AUC) were computed based on predicted labels by each NB output. To evaluate the prognostic values of each biomarker combination for cMCI and because neuropsychological information is used to establish the AD diagnosis therewith inducing circularity in the classification problem, all metrics for this group were computed separately for biomarker data acquired before and at conversion to AD. Further, to test if

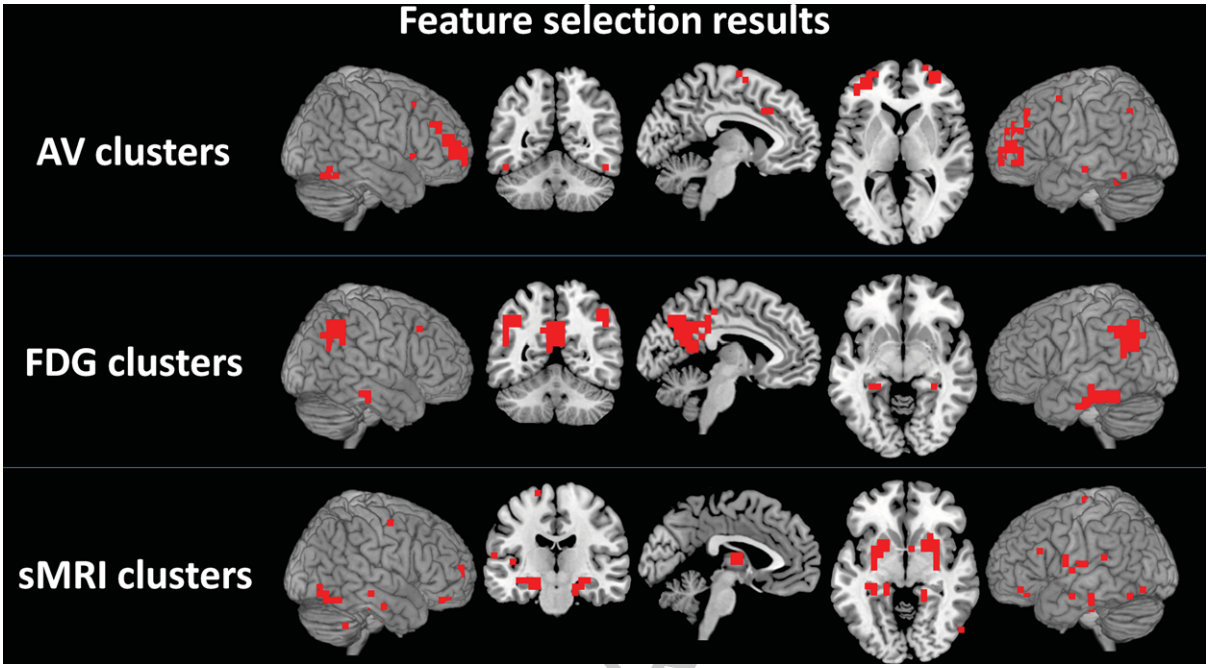


Fig. 1. Feature selection results. Clusters identified in [$^{18}$F]AV45-PET (florbetapir positron emission tomography), [$^{18}$F]FDG-PET (fluorodeoxyglucose positron emission tomography) and sMRI (structural magnetic resonance imaging) are displayed in the top, middle and bottom row, respectively.



the obtained balanced accuracies were significantly above chance, we ran permutation statistics (1000 permutations) for each biomarker constellation randomly shuffling the stable and converter MCI labels to the biomarker data and then computed balanced accuracy for each permutation. We then computed z-scores and corresponding *p*-values for the balanced accuracies obtained on real data relative to means and standard deviations obtained in randomly permuted data.

As AV45-PET was only added in ADNI2, the average time to conversion for these data was significantly lower. To control for this, we recomputed all sensitivity metrics for the testing data after matching them for TTC. Although the NB classifier is considered to be relatively robust regarding the number of training data, we aimed to exclude potential biases caused by these differences. For this reason, we repeated all training and classification procedures with the same, randomly drawn number of training cases as available for the biomarker constellation with the lowest number of cases.

Histopathological evaluation is still considered the gold standard for AD diagnosis. Thus, stratifying training data based on an *in vivo* biomarker of histopathology might improve its performance for early AD detection. Considering AV45-PET information as its *in vivo* approximation, we evaluated this possibility of using only data of amyloid positive AD and amyloid negative HC to train the classifiers. For these analyses, a previously reported threshold of 1.1 was applied to the mean AV45-PET standard uptake value ratio extracted from the selected clusters in the training dataset including only HC with a mean amyloid load below and AD patients with a mean above this threshold [50]. Applying this threshold resulted in an average exclusion of about 25% of the training data. Differences in accuracies, sensitivities, and specificities obtained using all versus amyloid thresholded data were evaluated using one-sample *t*-tests ($p < 0.05$ Bonferroni corrected for multiple comparisons) assuming no differences between the classifiers. As classification based on AV45-PET data might be differentially biased by application of an amyloid threshold for selection of training data, one-sample *t*-tests were performed separately for classifiers with and without this biomarker.

To illustrate the contribution of the APOE genotype, both the training and the testing dataset were stratified by the APOE allele combinations computing the relative proportion of either AD or cMCI in the respective populations. Lastly, we evaluated the possibility to use all biomarker combinations to predict the time to AD diagnosis as an index of future cognitive decline. For

this we computed regression analyses to predict TTC using z-transformed probabilities provided by the NB classifiers for cMCI data for each classifier and using only biomarker data acquired before conversion to AD. To provide a more quantitative metric of the predictive power of each classifier for TTC, we reported Pearson's correlation coefficients between observed TTC values and those predicted by the regression analyses. A squared Pearson's correlation coefficient (determination coefficient) provides the percentage of variance explained in the target variable by the variables used as predictors.

To enable a clearer interpretation of all above mentioned analyses, we have further ranked all biomarker constellations by sensitivities matched for TTC, specificities and correlations with observed TTC. All biomarker combinations were then sorted by the average rank of these three metrics.

## RESULTS

### *Demographic and neuropsychological results*

The groups did not differ with respect to age and gender (Table 1). There was a significant difference in education. *post-hoc t*-tests revealed differences in education only between AD and HC ($t(254) = 3.5$; $p = 0.001$) but not between cMCI and sMCI ($t(440) = 0.2$; $p = 0.875$). Comparisons of neuropsychological scores revealed significant between-group differences in all six measures (Table 1). Subsequent *post-hoc t*-tests revealed significant differences between AD and HC (all $p < 0.001$) in all measures. When comparing cMCI and sMCI all measures except for GDS ($p = 1.0$) were also significantly different.

### *Classification results*

Classification results for all biomarker combinations are displayed in Table 2 and Fig. 2. All classifiers performed significantly above chance level for differentiating between sMCI and cMCI (all $p < 0.01$). For single biomarkers, highest balanced accuracy (74.5%), specificity (83.9%), and AUC (0.824) were obtained using FDG-PET only (Fig. 3a). In contrast, highest sensitivity of 85.4% but on cost of a very low specificity (52.4%) was obtained using a classifier based on neuropsychological scores. Adjustment for TTC resulted in an even higher sensitivity of 92.9% for this classifier (Table 2). A strong increase in sensitivity when adjusting for TTC was also observed for sMRI. Lowest single modality classifier performance with a balanced accuracy of 59.5% was obtained for APOE followed by AV45-PET with 63.5%. At time of conversion, highest sensitivity of 100% for single biomarkers was obtained for neuropsychological scores followed by FDG-PET with 90% being the most sensitive imaging biomarker.

When evaluating all possible combinations of imaging biomarkers with APOE and neuropsychological information, highest balanced accuracy of 85% and highest sensitivity of 100% were obtained for the combination of AV45-PET and sMRI with neuropsychological scores (Table 2, Fig. 3b). In contrast, when adjusting the testing data for TTC, 86.8% was the highest accuracy observed for the combination of APOE, FDG-PET, and sMRI. This combination also showed the highest specificity of 86.1% and an AUC of 0.84. The constellation of biomarkers providing the lowest balanced accuracy was with 60.6% the combination of APOE with AV45- and FDG-PET. All classifier results were comparable when matching for size of the training cohorts (Table 3).

The use of only amyloid negative controls and amyloid positive AD did not significantly change accuracies [$t(16) = -2.024$; $p = 0.36$], sensitivities [$t(16) = -2.083$; $p = 1.0$], and specificities [$t(16) = -2.083$; $p = 0.3240$] for classifiers that did not include AV45-PET (Table 4). The only strong change for these classifiers was observed for the combination of neuropsychological profiles with FDG-PET and sMRI information for which the sensitivity increased by 10% whilst the specificity decreased by 14%. In contrast, significant changes with average sensitivity increases by 7% [$t(17) = -4.244$; $p = 0.006$] and specificity decreases by 12% [$t(17) = -7.965$; $p < 0.001$] were observed for biomarker combinations that included AV45-PET. Differences in accuracies, though on average lower for amyloid thresholded data, were not significant [$t(17) = -1.940$; $p = 0.42$].

When using z-transformed NB probabilities as predictors for TTC, several biomarker combinations showed a significant association with TTC. The strongest and significant correlations with TTC of $r = 0.65$ ($p < 0.001$) were observed when using classifier predictions based on neuropsychological scores and either FDG-PET or sMRI (Table 2, Fig. 4b, c). For output of single modality classifiers, the strongest and only significant correlation with TTC ($r = 0.41$) was observed for neuropsychological profiles (Fig. 4a).

Table 2
Classification results for differentiation between sMCI and cMCI using all training data

| | BA | BA adjusted for TTC | Sensitivity before conversion | Sensitivity adjusted for TTC | Sensitivity at conversion | Specificity | AUC | N Training | N cMCI before conversion | N cMCI at conversion | N sMCI | Correlation with TTC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| APOE | 0.597 | 0.607 | 0.650 | 0.670 | 0.684 | 0.543 | 0.606 | 256 | 177 | 76 | 265 | 0.103 |
| NP | 0.689 | 0.726 | 0.854 | 0.929 | 1.000 | 0.524 | 0.718 | 750 | 301 | 76 | 657 | 0.406** |
| AV45 | 0.635 | 0.635 | 0.667 | 0.667 | 0.600 | 0.604 | 0.664 | 148 | 21 | 10 | 164 | 0.158 |
| FDG | 0.745 | 0.724 | 0.651 | 0.609 | 0.900 | 0.839 | 0.824 | 245 | 43 | 10 | 223 | 0.164 |
| sMRI | 0.648 | 0.743 | 0.618 | 0.808 | 0.778 | 0.678 | 0.625 | 422 | 55 | 18 | 354 | 0.263 |
| APOE + NP | 0.694 | 0.726 | 0.870 | 0.934 | 1.000 | 0.518 | 0.722 | 750 | 301 | 76 | 657 | 0.409** |
| AV45 + FDG | 0.612 | 0.612 | 0.571 | 0.571 | 0.900 | 0.652 | 0.724 | 143 | 21 | 10 | 164 | 0.049 |
| AV45 + sMRI | 0.736 | 0.729 | 0.800 | 0.786 | 0.333 | 0.673 | 0.748 | 105 | 15 | 9 | 107 | 0.057 |
| FDG + sMRI | 0.811 | 0.830 | 0.774 | 0.813 | 0.778 | 0.848 | 0.834 | 177 | 31 | 9 | 151 | 0.170 |
| AV45 + FDG + sMRI | 0.750 | 0.743 | 0.800 | 0.786 | 1.000 | 0.701 | 0.794 | 102 | 15 | 9 | 107 | –0.156 |
| **NP+** | | | | | | | | | | | | |
| AV45 | 0.743 | 0.743 | 0.810 | 0.810 | 0.900 | 0.677 | 0.797 | 148 | 21 | 10 | 164 | 0.501* |
| FDG | 0.740 | 0.803 | 0.744 | 0.870 | 1.000 | 0.735 | 0.771 | 245 | 43 | 10 | 223 | 0.652** |
| sMRI | 0.693 | 0.783 | 0.782 | 0.962 | 1.000 | 0.605 | 0.721 | 422 | 55 | 18 | 354 | 0.644** |
| AV45 + FDG | 0.732 | 0.732 | 0.762 | 0.762 | 0.900 | 0.701 | 0.811 | 143 | 21 | 10 | 164 | 0.322 |
| AV45 + sMRI | 0.850 | 0.850 | 1.000 | 1.000 | 0.889 | 0.701 | 0.829 | 105 | 15 | 9 | 107 | 0.557* |
| FDG + sMRI | 0.748 | 0.784 | 0.742 | 0.813 | 1.000 | 0.755 | 0.833 | 177 | 31 | 9 | 151 | 0.588** |
| AV45 + FDG + sMRI | 0.807 | 0.802 | 0.867 | 0.857 | 0.889 | 0.748 | 0.830 | 102 | 15 | 9 | 107 | 0.252 |
| **APOE+** | | | | | | | | | | | | |
| AV45 | 0.647 | 0.647 | 0.714 | 0.714 | 0.600 | 0.579 | 0.665 | 148 | 21 | 10 | 164 | 0.178 |
| FDG | 0.759 | 0.748 | 0.674 | 0.652 | 0.900 | 0.843 | 0.824 | 245 | 43 | 10 | 223 | 0.223 |
| sMRI | 0.652 | 0.718 | 0.636 | 0.769 | 0.833 | 0.667 | 0.638 | 422 | 55 | 18 | 354 | 0.284 |
| AV45 + FDG | 0.606 | 0.606 | 0.571 | 0.571 | 0.900 | 0.640 | 0.725 | 143 | 21 | 10 | 164 | 0.069 |
| AV45 + sMRI | 0.736 | 0.729 | 0.800 | 0.786 | 0.333 | 0.673 | 0.753 | 105 | 15 | 9 | 107 | 0.083 |
| FDG+sMRI | 0.834 | 0.868 | 0.806 | 0.875 | 0.778 | 0.861 | 0.840 | 177 | 31 | 9 | 151 | 0.193 |
| AV45 + FDG + sMRI | 0.750 | 0.743 | 0.800 | 0.786 | 0.778 | 0.701 | 0.796 | 102 | 15 | 9 | 107 | –0.126 |
| **APOE + NP+** | | | | | | | | | | | | |
| AV45 | 0.740 | 0.740 | 0.810 | 0.810 | 0.900 | 0.671 | 0.794 | 148 | 21 | 10 | 164 | 0.506* |
| FDG | 0.738 | 0.800 | 0.744 | 0.870 | 1.000 | 0.731 | 0.770 | 245 | 43 | 10 | 223 | 0.653** |
| sMRI | 0.684 | 0.783 | 0.764 | 0.962 | 1.000 | 0.605 | 0.724 | 422 | 55 | 18 | 354 | 0.646** |
| AV45 + FDG | 0.755 | 0.755 | 0.810 | 0.810 | 1.000 | 0.701 | 0.810 | 143 | 21 | 10 | 164 | 0.340 |
| AV45 + sMRI | 0.850 | 0.850 | 1.000 | 1.000 | 0.889 | 0.701 | 0.830 | 105 | 15 | 9 | 107 | 0.553* |
| FDG + sMRI | 0.768 | 0.818 | 0.774 | 0.875 | 1.000 | 0.762 | 0.833 | 177 | 31 | 9 | 151 | 0.591** |
| AV45 + FDG + sMRI | 0.807 | 0.802 | 0.867 | 0.857 | 0.889 | 0.748 | 0.833 | 102 | 15 | 9 | 107 | 0.256 |

APOE, apolipoprotein E; AUC, area under the curve; AV45, florbetapir positron emission tomography; BA, balanced accuracy; cMCI, mild cognitive impairment patients converting to Alzheimer's disease; FDG, fluorodeoxyglucose positron emission tomography; N, number of observations; NP, neuropsychological profiles; sMCI, stable mild cognitive impairment; sMRI, structural magnetic resonance imaging; TTC, time to conversion to Alzheimer's disease. $^{*}p < 0.05$, $^{**}p < 0.001$.

## DISCUSSION

In this study we demonstrated that a fully independent classifier built only on AD and HC data, which includes imaging, genetic and neuropsychological biomarkers, can reliably discriminate between sMCI and cMCI outperforming previously reported accuracies. We show that combinations of biomarkers reflecting several pathophysiological mechanisms, genetics and cognition provide greatest sensitivities in the MCI population. Further, we identify biomarker combinations providing very accurate estimations of TTC as an indicator of future disease progression. By controlling all of the evaluated combinations for potential differences in TTC and size of the training data we additionally account for some known aspects which might have biased the observed accuracies. In the single biomarker setting, highest sensitivity and strongest association with disease staging is found for neuropsychological information. In contrast, highest specificity and the overall accuracy are achieved by FDG-PET.

By controlling for TTC and combining APOE with structural and glucose imaging, we obtain an accuracy of about 87% for differentiation between cMCI and sMCI, outperforming all other combinations evaluated in our study. The observed accuracy also substantially outperforms most previously reported accuracies for this discrimination problem [10, 19, 20, 27, 51–57]. The improved discrimination when adding APOE to both imaging modalities can be explained by its known

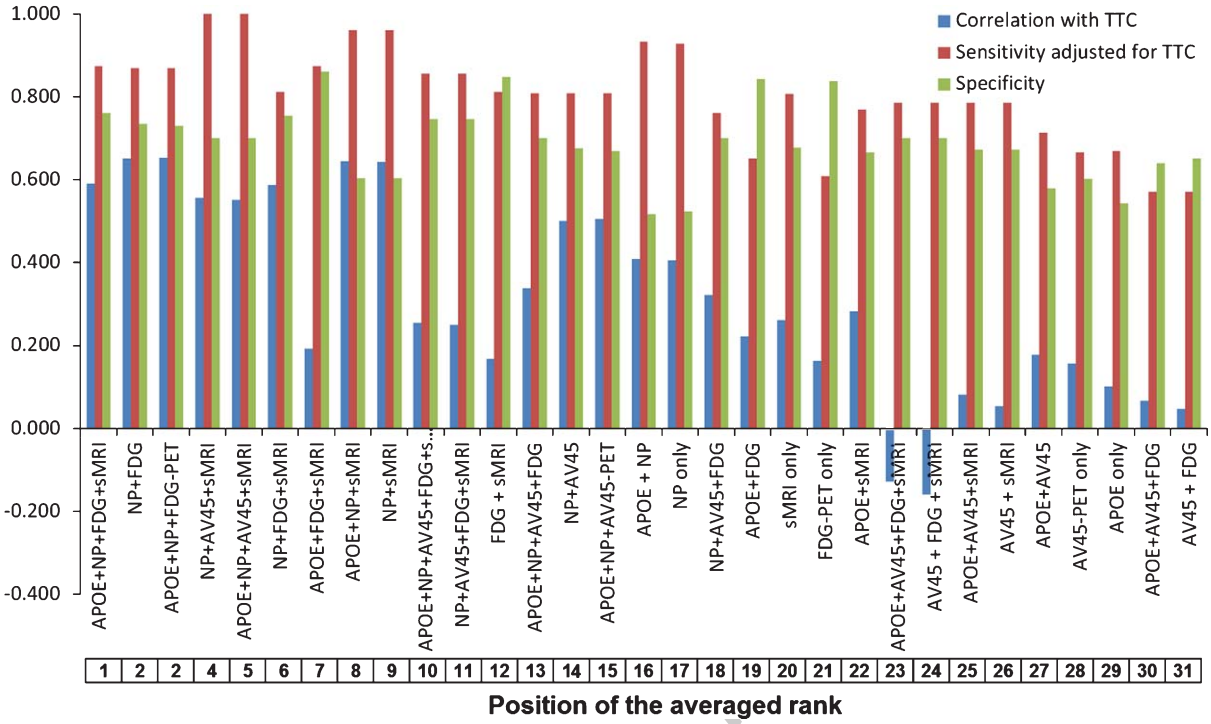


Fig. 2. Classification results for differentiation between stable mild cognitive impairment patients and those converting to Alzheimer's disease during the follow-up. Correlations with time to conversion (TTC), sensitivities matched for TTC, and specificities are displayed for all biomarker constellations sorted by the average rank across these three metrics.

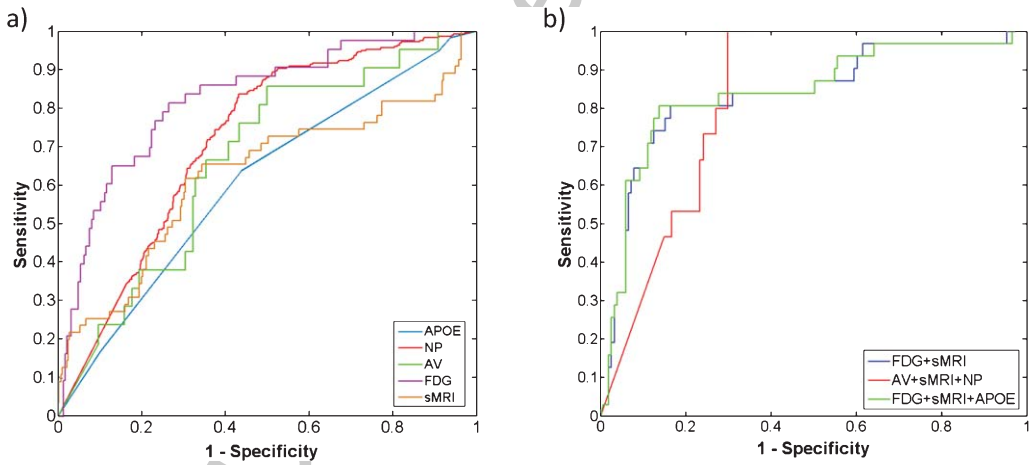


Fig. 3. Classification results for differentiation between stable mild cognitive impairment patients and those converting to Alzheimer's disease during the follow-up. Receiver operating characteristics (ROC) curves are displayed for differentiation based on single biomarkers (a) and for three classifier combinations showing the highest balanced accuracy rates (b).

strong positive and negative predictive value for particular allele combinations as illustrated in Fig. 5 [2, 58–60]. Furthermore, these results demonstrate that a known genetic risk factor combined with neuropsychological information and two *in vivo* measures of neuropathological mechanisms like brain atrophy and neurodegeneration better predict the final phenotype of conversion. Importantly, as compared to most previous studies the high accuracy was achieved using fully independent training and testing data therewith reducing the risk of overfitting. Though differential sensitivities of different combinations of imaging,

Table 3
Classification results for differentiation between sMCI and cMCI using equally sized training data

| | BA | Sensitivity | Specificity | AUC |
|---|---|---|---|---|
| APOE | 0.597 | 0.650 | 0.543 | 0.615 |
| NP | 0.685 | 0.857 | 0.513 | 0.716 |
| AV | 0.629 | 0.667 | 0.591 | 0.667 |
| FDG | 0.727 | 0.628 | 0.825 | 0.818 |
| MRI | 0.629 | 0.509 | 0.749 | 0.623 |
| APOE + NP | 0.679 | 0.850 | 0.508 | 0.710 |
| AV45 + FDG | 0.686 | 0.714 | 0.659 | 0.726 |
| AV45 + sMRI | 0.732 | 0.800 | 0.664 | 0.746 |
| FDG + sMRI | 0.756 | 0.677 | 0.834 | 0.803 |
| AV45 + FDG + sMRI | 0.750 | 0.800 | 0.701 | 0.794 |
| **NP+** | | | | |
| AV45 | 0.749 | 0.810 | 0.689 | 0.778 |
| FDG | 0.727 | 0.674 | 0.780 | 0.748 |
| sMRI | 0.692 | 0.782 | 0.602 | 0.728 |
| AV45 + FDG | 0.735 | 0.762 | 0.707 | 0.795 |
| AV45 + sMRI | 0.850 | 1.000 | 0.701 | 0.825 |
| FDG + sMRI | 0.768 | 0.742 | 0.795 | 0.818 |
| AV45 + FDG + sMRI | 0.807 | 0.867 | 0.748 | 0.830 |
| **APOE+** | | | | |
| AV45 | 0.647 | 0.714 | 0.579 | 0.665 |
| FDG | 0.761 | 0.651 | 0.870 | 0.831 |
| sMRI | 0.652 | 0.618 | 0.686 | 0.634 |
| AV45 + FDG | 0.665 | 0.714 | 0.616 | 0.715 |
| AV45 + sMRI | 0.736 | 0.800 | 0.673 | 0.759 |
| FDG + sMRI | 0.779 | 0.710 | 0.848 | 0.820 |
| AV45 + FDG + sMRI | 0.750 | 0.800 | 0.701 | 0.796 |
| **APOE + NP+** | | | | |
| AV45 | 0.720 | 0.714 | 0.726 | 0.797 |
| FDG | 0.752 | 0.698 | 0.807 | 0.787 |
| sMRI | 0.638 | 0.727 | 0.548 | 0.690 |
| PPL | 0.615 | 0.929 | 0.300 | 0.640 |
| AV45 + FDG | 0.735 | 0.762 | 0.707 | 0.800 |
| AV45 + sMRI | 0.817 | 0.933 | 0.701 | 0.827 |
| FDG + sMRI | 0.744 | 0.806 | 0.682 | 0.814 |
| AV45 + FDG + sMRI | 0.807 | 0.867 | 0.748 | 0.833 |

APOE, apolipoprotein E; AUC, area under the curve; AV45, florbetapir positron emission tomography; BA, balanced accuracy; cMCI, mild cognitive impairment patients converting to Alzheimer's disease; FDG, fluorodeoxyglucose positron emission tomography; NP, neuropsychological profiles; sMCI, stable mild cognitive impairment; sMRI, structural magnetic resonance imaging.



genetic, and neuropsychological biomarkers for discrimination between cMCI and sMCI were repeatedly reported in previous studies, these estimates were mostly computed in different MCI subpopulations, e.g., not each patient had each imaging biomarker. This aspect limits the comparability of accuracies of different imaging biomarker combinations due to potential differences in diagnostics, training sets, or other between-group differences in clinical or demographic characteristics across the MCI populations included for different modalities. By evaluating all imaging biomarkers in the same MCI subjects we account for these potential biases. We identify FDG-PET as the most accurate single modality biomarker differentiating between cMCI and sMCI. This finding is consistent with conclusions of a recent comprehensive meta-analysis reporting higher accuracies for FDG-PET as compared to other imaging and clinical biomarkers to detect AD related pathology [61].

Also consistently with previous studies, we find that a combination of FDG-PET and sMRI results in a substantially improved accuracy for early AD detection [47, 62–64]. Adding APOE genotype to this combination further increases the observed accuracy. This combination also results in the highest specificity of 86%. We observed a similarly high accuracy for the combination of neuropsychological profiles with AV45-PET and sMRI. However, this good performance is mostly driven by a very high sensitivity whilst the specificity is comparably low. Correspondingly, these two combinations of biomarkers might provide alternative enrichment strategies for clinical

Table 4
Classification results for differentiation between sMCI and cMCI using amyloid positive AD and amyloid negative controls

| No AV45 classifiers | BA | Sensitivity | Specificity | Δ BA | Δ sensitivity | Δ specificity |
|---|---|---|---|---|---|---|
| APOE | 0.597 | 0.650 | 0.543 | 0.000 | 0.000 | 0.000 |
| NP | 0.698 | 0.834 | 0.562 | 0.009 | –0.020 | 0.038 |
| FDG | 0.738 | 0.651 | 0.825 | –0.007 | 0.000 | –0.013 |
| sMRI | 0.650 | 0.600 | 0.701 | 0.002 | –0.018 | 0.023 |
| APOE + NP | 0.705 | 0.841 | 0.569 | 0.011 | –0.030 | 0.052 |
| FDG + sMRI | 0.791 | 0.742 | 0.841 | –0.019 | –0.032 | –0.007 |
| NP + FDG | 0.720 | 0.767 | 0.673 | –0.020 | 0.023 | –0.063 |
| NP + sMRI | 0.693 | 0.782 | 0.605 | 0.000 | 0.000 | 0.000 |
| NP + FDG + sMRI | 0.727 | 0.839 | 0.616 | –0.021 | 0.097 | –0.139 |
| APOE + FDG | 0.754 | 0.674 | 0.834 | –0.004 | 0.000 | –0.009 |
| APOE + sMRI | 0.659 | 0.636 | 0.681 | 0.007 | 0.000 | 0.014 |
| APOE + FDG + sMRI | 0.818 | 0.774 | 0.861 | –0.016 | –0.032 | 0.000 |
| APOE + FDG + NP | 0.716 | 0.767 | 0.664 | –0.022 | 0.023 | –0.067 |
| APOE + sMRI + NP | 0.700 | 0.782 | 0.619 | 0.016 | 0.018 | 0.014 |
| APOE + FDG + sMRI + NP | 0.731 | 0.839 | 0.623 | –0.037 | 0.065 | –0.139 |
| Average change | – | – | – | –0.007 | 0.006 | –0.020 |
| **AV45 classifiers** | **BA** | **Sensitivity** | **Specificity** | **Δ BA** | **Δ sensitivity** | **Δ specificity** |
| AV45 | 0.652 | 0.762 | 0.543 | 0.017 | 0.095 | –0.061 |
| AV45 + FDG | 0.655 | 0.762 | 0.549 | 0.043 | 0.190 | –0.104 |
| AV45 + sMRI | 0.732 | 0.867 | 0.598 | –0.004 | 0.067 | –0.075 |
| AV45 + FDG + sMRI | 0.737 | 0.867 | 0.607 | –0.013 | 0.067 | –0.093 |
| NP + AV45 | 0.724 | 0.905 | 0.543 | –0.019 | 0.095 | –0.134 |
| NP + AV45 + FDG | 0.706 | 0.857 | 0.555 | –0.026 | 0.095 | –0.146 |
| NP + AV45 + sMRI | 0.714 | 0.933 | 0.495 | –0.136 | –0.067 | –0.206 |
| NP + AV45 + FDG+sMRI | 0.766 | 1.000 | 0.533 | –0.041 | 0.133 | –0.215 |
| APOE + AV45 | 0.652 | 0.762 | 0.543 | 0.006 | 0.048 | –0.037 |
| APOE + FDG | 0.754 | 0.674 | 0.834 | –0.004 | 0.000 | –0.009 |
| APOE + AV45 + FDG | 0.655 | 0.762 | 0.549 | 0.050 | 0.190 | –0.091 |
| APOE + AV45 + sMRI | 0.732 | 0.867 | 0.598 | –0.004 | 0.067 | –0.075 |
| APOE + AV45 + FDG + sMRI | 0.732 | 0.867 | 0.598 | –0.018 | 0.067 | –0.103 |
| APOE + AV45 + NP | 0.721 | 0.905 | 0.537 | –0.019 | 0.095 | –0.134 |
| APOE + AV45 + FDG + NP | 0.706 | 0.857 | 0.555 | –0.049 | 0.048 | –0.146 |
| APOE + AV45 + sMRI + NP | 0.728 | 0.933 | 0.523 | –0.122 | –0.067 | –0.178 |
| APOE + AV45 + FDG + sMRI + NP | 0.762 | 1.000 | 0.523 | –0.045 | 0.133 | –0.224 |
| Average change | – | – | – | –0.023 | 0.074* | –0.119* |

APOE, apolipoprotein E; AUC, area under the curve; AV45, florbetapir positron emission tomography; BA, balanced accuracy; cMCI, mild cognitive impairment patients converting to Alzheimer's disease; FDG, fluorodeoxyglucose positron emission tomography; NP, neuropsychological profiles; sMCI, stable mild cognitive impairment; sMRI, structural magnetic resonance imaging. Δ BA, Δ sensitivity, and Δ specificity refer to differences from results obtained using all training data (as displayed in Table 2). $^{*}p < 0.05$.

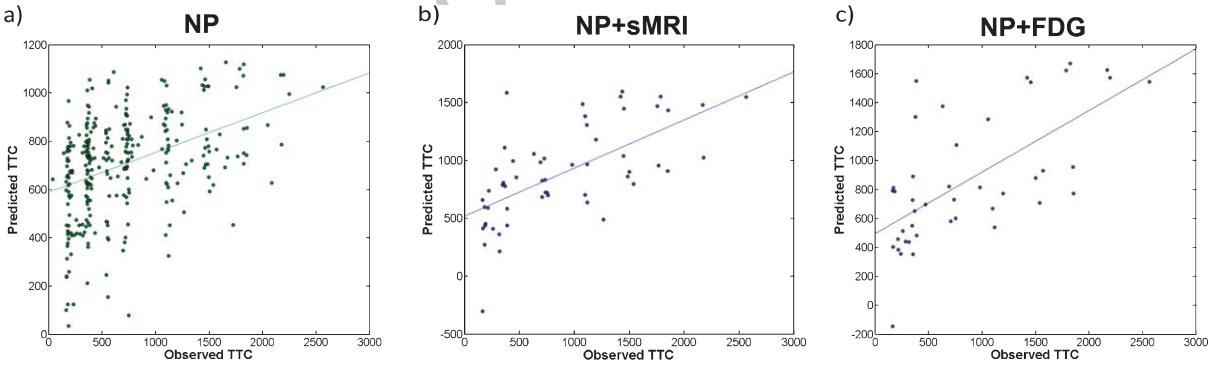


Fig. 4. Results of regression analyses between Naïve Bayes z-transformed probabilities and time to conversion (TTC). Results for classifiers showing the strongest correlation with TTC in a single biomarker setting (a) and when combining different modalities (b and c) are displayed. Predicted time to conversion values are displayed on the y axis. The lines indicate the regression slopes. FDG, fluorodeoxyglucose positron emission tomography; NP, neuropsychological scores; sMRI, structural magnetic resonance imaging.

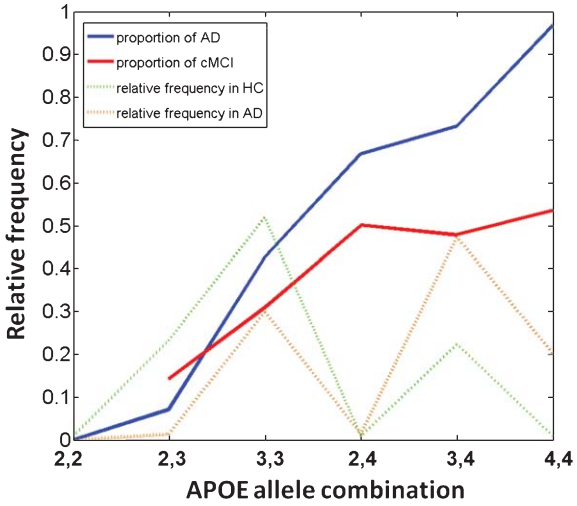


Fig. 5. Visualization of apolipoprotein E (APOE) genotype profiles observed in the training and testing dataset. Blue solid line indicates the relative frequency of Alzheimer's disease (AD) patients out of all subjects having the particular APOE allele combination in the training cohort. Red solid line indicates the relative frequency of mild cognitive impairment patients converting to AD (cMCI) out to all MCI having the particular APOE allele combination in the testing cohort. Green dotted line indicates the relative frequency of the particular APOE allele combination in healthy controls in the training dataset. Orange dotted line indicates the relative frequency of the particular APOE allele combination in AD in the training dataset.

trials where high sensitivity or specificity might be prioritized.

Most importantly, for the first time we identified biomarker combinations which not only allow a very accurate discrimination of cMCI and sMCI but are also strongly predictive at an individual subject's level to the future cognitive decline as indicated by TTC. In a single biomarker setting only neuropsychological scores are a significant predictor of future disease as indicated by a 0.4 correlation with TTC. However, combining these with either FDG-PET or sMRI increases the explained variance in TTC to above 40% (squared Pearson correlation coefficients). These findings suggest that both biomarker combinations are highly sensitive to future disease progression. This aspect might be particularly important in clinical trials aiming to identify MCI patients and earlier and/or more homogeneous disease stages. Though many previous studies focused on identification of biomarker combinations that increase the risk of conversion to AD, only few of the studies so far also evaluated the link between identified biomarkers and TTC [22, 25, 28, 53]. By focusing on hazard ratios these studies identified risk factors associated with TTC at group level. These factors do not yet necessarily allow an accurate prediction of progression for individual patients. Furthermore, none of the above mentioned studies performed an exhaustive comparison of different imaging, genetic and neuropsychological biomarker to identify constellations that are most sensitive to TTC. The strongly significant association identified in our study for the combinations of neuropsychological scores with either FDG-PET or sMRI information indicates a high potential of these modality combinations to provide prognostic information for individual MCI patients. Beside this already highly important information for the patients, the established relationships can be also applied in clinical trials to identify MCI patients at particular disease stages. Considering that several promising phase III studies targeting mechanisms in AD have recently failed [65–67] with *post-hoc* analyses of these failures suggesting that the inclusion of AD patients at quite advanced disease stages might partially explain the lack of observed treatment effects [68]. The identified biomarker combinations might provide a sensitive stratification mechanism to identify patients at earlier disease stages in future clinical studies. More recent AD trials therefore aim to focus on more prodromal AD stages as the primary intervention window [69]. Crucial for their success might be therefore the capability to accurately diagnose AD at its early disease stages. Contrary to our prior expectations of getting more accurate classifiers when using only amyloid positive AD and amyloid negative HC to train the classifiers, no benefit in terms of accuracies is observed for discrimination of cMCI and sMCI. Interestingly, we see a strong dissociation between classifiers with and without AV45-PET in terms of obtained sensitivities and specificities. Whilst no major changes are observed for classifiers not including AV45-PET, we see a strong shift towards an increased sensitivity and decreased specificity in classifier combinations including this biomarker. Considering that previous epidemiological studies have shown that amnestic sMCI as included in the ADNI are at high risk of conversion to AD when followed for a period of up to 10 years [9, 70], the assignment of a higher percentage of sMCI patients as AD might in fact more closely reflect the true differentiation between AD and non-AD MCI than the criteria of a stable follow-up of two years we apply for sMCI. However, these considerations remain speculative until sMCI populations with longer follow-up than the one included in our study become available.

Lastly, we observed even reduced accuracy as compared to single biomarkers when AV45- and FDG-PET

information is combined. This finding suggests a low consistency between these two imaging modalities in the evaluated MCI population. A potential reason for this might be that amyloid depositions are rather dissociated from disease progression as reflected by functional imaging markers. The lack of clinical benefits in pharmacological trials aiming at amyloid clearance despite successful reductions of those depositions supports this assumption [71–73].

Even though in the present study we aimed to account for most potential limitations and biases common to these types of studies, several limitations still need to be considered prior to interpretation of the reported findings. First of all, in our effort to identify a homogeneous subpopulation of the ADNI cohort having the constellation of all biomarkers included in our study, we had to discard a large amount of data available in the ADNI dataset. In particular for the constellations of biomarkers including AV45-PET and at the time point of conversion applying these filtering criteria resulted in a very low and varying number of MCI testing cases depending on the biomarker constellation. The data loss is mostly due to the fact that AV45-PET was only included in the ADNI-GO and 2 and to sparse acquisition of some of the imaging measures. For this reason, we limited our discussion of accuracies obtained for data at the time point of conversion as they need to be validated in samples that are significantly larger than evaluated in the current study. Correspondingly, the low numbers of testing data need to be considered when interpreting sensitivities obtained using the affected combinations. For the reason of varying and small numbers of testing cases for each biomarker constellation we also did not directly compare the classifier performances to each other but only to chance level performance. This formal testing needs to be performed when a sufficiently large amount of data becomes available, covering all of the studied modalities in exactly the same MCI population. A second limitation of our study is related to the pre-processing pipelines applied for imaging data. Numerous studies including our own previous work have provided evidence that particular pre-processing steps omitted in our study, e.g., partial volume effect correction or adjustment for age-related effects, can further improve the sensitivity of the single imaging modalities for discrimination between AD and HC or cMCI and sMCI [40, 42, 43, 74]. Due to the high sparsity of the available imaging data, applying these pre-processing steps would have resulted in further exclusion of a substantial amount of imaging data eventually leading to a very limited sample size of the training and testing datasets. As demonstrated by earlier studies cited above, having more optimal pre-processing pipelines should further increase the accuracies observed here for the single biomarker classifiers. Therefore, if anything, our results are likely to underestimate the achievable accuracies.

A further limitation of our study is the pre-selection of neuropsychological and clinical tests used in our study to differentiate between stable and converter MCI. Our major motivation to do a pre-selection of tests from the extensive test battery included in the ADNI was to cover major domains affected in AD with a reasonable number of tests that could be integrated in a standard clinical setting. However, inclusion of other neuropsychological and functional measures might further increase accuracies achievable with these types of biomarkers.

Lastly, though we validated the obtained classifier using fully independent testing data, the reported classifier performances remain limited to the ADNI dataset with its restrictive inclusion and exclusion criteria. They are therefore likely to overestimate accuracies achievable in a standard clinical setting in the presence of other possible dementia syndromes [75].

To summarize, in our study we provide strong evidence that fully automated classifiers based on combination of imaging, genetic, and/or neuropsychological biomarkers can reliably and very accurately discriminate between stable and converter MCI. Further, we demonstrate the high sensitivity of the some of the identified biomarker combinations to future disease progression as indicating by the time to conversion to AD. The result of our study further confirms the high degree of pathological and clinical heterogeneity of AD [76], thus suggesting that the combined use of genetic and imaging and neuropsychological biomarkers in the framework of endophenotypes for this disorder could increase the power of identifying individuals at risk for conversion. Notably, these biomarker combinations could be used for enrichment of clinical trials to identify MCI patients at earlier AD stages.

## ACKNOWLEDGMENTS

Juergen Dukart, Fabio Sambataro, and Alessandro Bertolino are full-time employees of F.Hoffmann-La Roche, Basel, Switzerland. The authors received no specific funding for this work. F.Hoffmann-La Roche provided support in the form of salary for authors Juergen Dukar, Fabio Sambataro and Alessandro

Bertolino, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' disclosures available online (http://j-alz.com/manuscript-disclosures/15-0570r1).

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: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec, Inc.; Bristol-Myers Squibb Company; 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.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc, Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Rev December 5, 2013 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 Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

## SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-150570.

## REFERENCES

[1] Braak H, Braak E (1997) Diagnostic criteria for neuropathologic assessment of Alzheimer's disease. *Neurobiol Aging* **18**, S85-S88.

[2] Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921-923.

[3] Heindel WC, Salmon DP, Shults CW, Walicke PA, Butters N (1989) Neuropsychological evidence for multiple implicit memory systems: A comparison of Alzheimer's, Huntington's, and Parkinson's disease patients. *J Neurosci* **9**, 582-587.

[4] Hodges JR, Salmon DP, Butters N (1992) Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge? *Neuropsychologia* **30**, 301-314.

[5] Mielke R, Pietrzyk U, Jacobs A, Fink GR, Ichimiya A, Kessler J, Herholz K, Heiss WD (1994) HMPAO SPET and FDG PET in Alzheimer's disease and vascular dementia: Comparison of perfusion and metabolic pattern. *Eur J Nucl Med* **21**, 1052-1060.

[6] Rabinovici GD, Seeley WW, Kim EJ, Gorno-Tempini ML, Rascovsky K, Pagliaro TA, Allison SC, Halabi C, Kramer JH, Johnson JK, Weiner MW, Forman MS, Trojanowski JQ, Dearmond SJ, Miller BL, Rosen HJ (2007) Distinct MRI atrophy patterns in autopsy-proven Alzheimer's disease and frontotemporal lobar degeneration. *Am J Alzheimers Demen* **22**, 474-488.

[7] Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P (2007) Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-746.

[8] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 263-269.

[9] Visser PJ, Kester A, Jolles J, Verhey F (2006) Ten-year risk of dementia in subjects with mild cognitive impairment. *Neurology* **67**, 1201-1207.

[10] Adaszewski S, Dukart J, Kherif F, Frackowiak R, Draganski B, Initiative ADN (2013) How early can we predict Alzheimer's disease using computational anatomy? *Neurobiol Aging* **34**, 2815-2826.

[11] Desikan RS, Cabral HJ, Hess CP, Dillon WP, Glastonbury CM, Weiner MW, Schmansky NJ, Greve DN, Salat DH, Buckner RL (2009) Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain* **132**, 2048-2057.

[12] Forsberg A, Engler H, Almkvist O, Blomquist G, Hagman G, Wall A, Ringheim A, Långström B, Nordberg A (2008) PET imaging of amyloid deposition in patients with mild cognitive impairment. *Neurobiol Aging* **29**, 1456-1465.

[13] Gerardin E, Chételat G, Chupin M, Cuingnet R, Desgranges B, Kim HS, Niethammer M, Dubois B, Lehéricy S, Garnero L, Eustache F, Colliot O, Alzheimer's Disease Neuroimaging, Initiative (2009) Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *Neuroimage* **47**, 1476-1486.

[14] Johnson KA, Sperling RA, Gidicsin C, Carmasin J, Maye JE, Coleman RE, Reiman EM, Sabbagh MN, Sadowsky CH, Fleisher AS, Murali Doraiswamy P, Carpenter AP, Clark CM, Joshi AD, Lu M, Grundman M, Mintun MA, Pontecorvo MJ, Skovronsky DM, AV45-A11 study group (2013) Florbetapir (F18-AV-45) PET to assess amyloid burden in Alzheimer's disease dementia, mild cognitive impairment, and normal aging. *Alzheimers Dement J Alzheimers Assoc* **9**, S72-S83.

[15] McEvoy LK, Fennema-Notestine C, Roddey JC, Hagler DJ Jr, Holland D, Karow DS, Pung CJ, Brewer JB, Dale AM, Alzheimer's Disease Neuroimaging, Initiative (2009) Alzheimer disease: Quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. *Radiology* **251**, 195-205.

[16] Misra C, Fan Y, Davatzikos C (2009) Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI. *Neuroimage* **44**, 1415-1422.

[17] Mosconi L, Tsui WH, Herholz K, Pupi A, Drzezga A, Lucignani G, Reiman EM, Holthoff V, Kalbe E, Sorbi S, Diehl-Schmid J, Perneczky R, Clerici F, Caselli R, Beuthien-Baumann B, Kurz A, Minoshima S, de Leon MJ (2008) Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. *J Nucl Med* **49**, 390-398.

[18] Da X, Toledo JB, Zee J, Wolk DA, Xie SX, Ou Y, Shacklett A, Parmpi P, Shaw L, Trojanowski JQ, Davatzikos C, Alzheimer's Neuroimaging Initiative (2014) Integration and relative value of biomarkers for prediction of MCI to AD progression: Spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers. *Neuroimage Clin* **4**, 164-173.

[19] Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack CR Jr, Feldman HH, Bokde AL, Alexander GE, Scheltens P, Vellas B, Dubois B, Weiner M, Hampel H, North American Alzheimer's Disease Neuroimaging Initiative (ADNI) (2010) Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiol Aging* **33**, 1203-1214.

[20] Hinrichs C, Singh V, Xu G, Johnson SC, Alzheimer's Disease Neuroimaging, Initiative (2011) Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *Neuroimage* **55**, 574-589.

[21] Jack CR Jr, Lowe VJ, Senjem ML, Weigand SD, Kemp BJ, Shiung MM, Knopman DS, Boeve BF, Klunk WE, Mathis CA, Petersen RC (2008) 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. *Brain* **131**, 665-680.

[22] Jack CR Jr, Wiste HJ, Vemuri P, Weigand SD, Senjem ML, Zeng G, Bernstein MA, Gunter JL, Pankratz VS, Aisen PS, Weiner MW, Petersen RC, Shaw LM, Trojanowski JQ, Knopman DS, Alzheimer's Disease Neuroimaging, Initiative (2010) Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. *Brain* **133**, 3336-3348.

[23] Karow DS, McEvoy LK, Fennema-Notestine C, Hagler DJ, Jennings RG, Brewer JB, Hoh CK, Dale AM, Alzheimer's Disease Neuroimaging, Initiative (2010) Relative capability of MR imaging and FDG PET to depict changes associated with prodromal and early Alzheimer disease. *Radiology* **256**, 932-942.

[24] Perrin RJ, Fagan AM, Holtzman DM (2009) Multimodal techniques for diagnosis and prognosis of Alzheimer's disease. *Nature* **461**, 916-922.

[25] Vemuri P, Wiste HJ, Weigand SD, Shaw LM, Trojanowski JQ, Weiner MW, Knopman DS, Petersen RC, Jack CR Jr, Alzheimer's Disease Neuroimaging, Initiative (2009) MRI and CSF biomarkers in normal, MCI, and AD subjects: Diagnostic discrimination and cognitive correlations. *Neurology* **73**, 287-293.

[26] Westman E, Muehlboeck J, Simmons A (2012) Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *Neuroimage* **62**, 229-238.

[27] Young J, Modat M, Cardoso MJ, Mendelson A, Cash D, Ourselin S (2013) Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neuroimage Clin* **2**, 735-745.

[28] Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack CR Jr, Feldman HH, Bokde AL, Alexander GE, Scheltens P, Vellas B, Dubois B, Weiner M, Hampel H, North American Alzheimer's Disease Neuroimaging Initiative (ADNI) (2012) Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiol Aging* **33**, 1203-1214.

[29] Beach TG, Monsell SE, Phillips LE, Kukull W (2012) Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *J Neuropathol Exp Neurol* **71**, 266-273.

[30] Toledo JB, Cairns NJ, Da X, Chen K, Carter D, Fleisher A, Householder E, Ayutyanont N, Roontiva A, Bauer RJ, Eisen P, Shaw LM, Davatzikos C, Weiner MW, Reiman EM, Morris JC, Trojanowski JQ, Alzheimer's Disease Neuroimaging, Initiative (2013) Clinical and multimodal biomarker correlates of ADNI neuropathological findings. *Acta Neuropathol Commu* **1**, 65.

[31] Clark CM, Schneider JA, Bedell BJ, Beach TG, Bilker WB, Mintun MA, Pontecorvo MJ, Hefti F, Carpenter AP, Flitter ML, Krautkramer MJ, Kung HF, Coleman RE, Doraiswamy PM, Fleisher AS, Sabbagh MN, Sadowsky CH, Reiman EP, Zehntner SP, Skovronsky DM, AV45-A07 Study, Group (2011) Use of florbetapir-PET for imaging β-amyloid pathology. *JAMA* **305**, 275-283.

[32] Landau SM, Breault C, Joshi AD, Pontecorvo M, Mathis CA, Jagust WJ, Mintun MA, Alzheimer's Disease Neuroimaging, Initiative (2013) Amyloid-β imaging with Pittsburgh compound B and florbetapir: Comparing radiotracers and quantification methods. *J Nucl Med* **54**, 70-77.

[33] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939-944.

[34] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198.

[35] Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO (1983) Development and validation of a geriatric depression screening scale: A preliminary report. *J Psychiatr Res* **17**, 37-49.

[36] Mohs RC, Cohen L (1988) Alzheimer's Disease Assessment Scale (ADAS). *Psychopharmacol Bull* **24**, 627-628.

[37] McMinn MR, Wiens AN, Crossen JR (1987) Rey Auditory-Verbal Learning Test: Development of norms for healthy young adults. *Clin Neuropsychol* **2**, 67-87.

[38] Schmidt M. (1996) *Rey auditory verbal learning test: A handbook,* Western Psychological Services, Los Angeles.

[39] Pfeffer RI, Kurosaki TT, Harrah CH, Chance JM, Filos S (1982) Measurement of functional activities in older adults in the community. *J Gerontol* **37**, 323-329.

[40] Borghammer P, Aanerud J, Gjedde A (2009) Data-driven intensity normalization of PET group comparison studies is superior to global mean normalization. *Neuroimage* **46**, 981-988.

[41] Dukart J, Mueller K, Horstmann A, Vogt B, Frisch S, Barthel H, Becker G, Moller HE, Villringer A, Sabri O, Schroeter ML (2010) Differential effects of global and cerebellar normalization on detection and differentiation of dementia in FDG-PET studies. *Neuroimage* **49**, 1490-1495.

[42] Yakushev I, Landvogt C, Buchholz HG, Fellgiebel A, Hammers A, Scheurich A, Schmidtmann I, Gerhard A, Schreckenberger M, Bartenstein P (2008) Choice of reference area in studies of Alzheimer's disease using positron emission tomography with fluorodeoxyglucose-F18. *Psychiatry Res* **164**, 143-153.

[43] Barthel H, Rullmann M, Dukart J, Herbert S, Luthardt J, Gertz H-J, Reininger C, Sabri O. (2012) Improved diagnostic performance of florbetaben beta-amyloid PET via partial volume effect correction. *J Cereb Blood Flow Metab* **32**, S148-S148.

[44] Samuraki M, Matsunari I, Chen WP, Yajima K, Yanase D, Fujikawa A, Takeda N, Nishimura S, Matsuda H, Yamada M (2007) Partial volume effect-corrected FDG PET and grey matter volume loss in patients with mild Alzheimer's disease. *Eur J Nucl Med Mol Imaging* **34**, 1658-1669.

[45] Thomas BA, Erlandsson K, Modat M, Thurfjell L, Vandenberghe R, Ourselin S, Hutton BF (2011) The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease. *Eur J Nucl Med Mol Imaging* **38**, 1104-1119.

[46] Ashburner J (2007) A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95-113.

[47] Dukart J, Mueller K, Horstmann A, Barthel H, Möller HE, Villringer A, Sabri O, Schroeter ML (2011) Combined evaluation of FDG-PET and MRI improves detection and differentiation of dementia. *PLoS One* **6**, e18111.

[48] Aliferis CF, Tsamardinos I, Statnikov A (2003) HITON: A novel Markov Blanket algorithm for optimal variable selection. In *AMIA Annual Symposium Proceedings,* American Medical Informatics Association, pp. 21.

[49] Statnikov A, Tsamardinos I, Brown LE, Aliferis CF (2010) Causal explorer: A matlab library of algorithms for causal discovery and variable selection for classification. *Causation Predict Chall Chall Mach Learn* **2**, 267.

[50] Joshi AD, Pontecorvo MJ, Clark CM, Carpenter AP, Jennings DL, Sadowsky CH, Adler LP, Kovnat KD, Seibyl JP, Arora A, Saha K, Burns JD, Lowrey MJ, Mintun MA, Skovronsky DM, Florbetapir F, 18 Study, Investigators (2012) Performance characteristics of amyloid PET with florbetapir F 18 in patients with alzheimer's disease and cognitively normal subjects. *J Nucl Med* **53**, 378-384.

[51] Coupé P, Eskildsen SF, Manjón JV, Fonov VS, Pruessner JC, Allard M, Collins DL, Alzheimer's Disease Neuroimaging, Initiative (2012) Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. *Neuroimage Clin* **1**, 141-152.

[52] Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ (2322) (2011) Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol Aging* **32** e19-e27.

[53] Devanand DP, Pradhaban G, Liu X, Khandji A, De Santi S, Segal S, Rusinek H, Pelton GH, Honig LS, Mayeux R, Stern Y, Tabert MH, de Leon MJ (2007) Hippocampal and entorhinal atrophy in mild cognitive impairment: Prediction of Alzheimer disease. *Neurology* **68**, 828-836.

[54] Eskildsen SF, Coupé P, García-Lorenzo D, Fonov V, Pruessner JC, Collins DL, Alzheimer's Disease Neuroimaging, Initiative (2013) Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. *Neuroimage* **65**, 511-521.

[55] Nho K, Shen L, Kim S, Risacher SL, West JD, Foroud T, Jack CR Jr, Weiner MW, Saykin AJ (2010) Automatic Prediction of Conversion from Mild Cognitive Impairment to Probable Alzheimer's Disease using Structural Magnetic Resonance Imaging. *In AMIA Annual Symposium Proceedings*, American Medical Informatics Association, pp. 542.

[56] Vos SJ, van Rossum IA, Verhey F, Knol DL, Soininen H, Wahlund L-O, Hampel H, Tsolaki M, Minthon L, Frisoni GB, Froelich L, Nobili F, van der Flier W, Blennow K, Wolz R, Scheltens P, Visser PJ (2013) Prediction of Alzheimer disease in subjects with amnestic and nonamnestic MCI. *Neurology* **80**, 1124-1132.

[57] Wee C-Y, Yap P-T, Shen D, Alzheimer's Disease Neuroimaging, Initiative (2013) Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. *Hum Brain Mapp* **34**, 3411-3425.

[58] Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, Rimmler JB, Locke PA, Conneally PM, Schmader KE, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet* **7**, 180-184.

[59] Frisoni GB, Govoni S, Geroldi C, Bianchetti A, Calabresi L, Franceschini G, Trabucchi M (1995) Gene dose of the $\varepsilon$4 allele of apolipoprotein E and disease progression in sporadic late-onset Alzheimer's disease. *Ann Neurol* **37**, 596-604.

[60] Roses MD (1996) Apolipoprotein E alleles as risk factors in Alzheimer's disease. *Annu Rev Med* **47**, 387-400.

[61] Bloudek LM, Spackman DE, Blankenburg M, Sullivan SD (2011) Review and meta-analysis of biomarkers and diagnostic imaging in Alzheimer's disease. *J Alzheimers Dis* **26**, 627-645.

[62] Hinrichs C, Singh V, Mukherjee L, Xu G, Chung MK, Johnson SC, Alzheimer's Disease Neuroimaging, Initiative (2009) Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *Neuroimage* **48**, 138-149.

[63] Kawachi T, Ishii K, Sakamoto S, Sasaki M, Mori T, Yamashita F, Matsuda H, Mori E (2006) Comparison of the diagnostic performance of FDG-PET and VBM-MRI in very mild Alzheimer's disease. *Eur J Nucl Med Mol Imaging* **33**, 801-809.

[64] Zhang D, Wang Y, Zhou L, Yuan H, Shen D, Alzheimer's Disease Neuroimaging, Initiative (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* **55**, 856-867.

[65] Doody RS, Raman R, Farlow M, Iwatsubo T, Vellas B, Joffe S, Kieburtz K, He F, Sun X, Thomas RG, Aisen PS, Alzheimer's Disease Cooperative Study Steering Committee, Siemers E, Sethuraman G, Mohs R, Semagacestat Study, Group (2013) A phase 3 trial of semagacestat for treatment of Alzheimer's disease. *N Engl J Med* **369**, 341-350.

[66] Doody RS, Thomas RG, Farlow M, Iwatsubo T, Vellas B, Joffe S, Kieburtz K, Raman R, Sun X, Aisen PS, Siemers E, Liu-Seifert H, Mohs R, Alzheimer's Disease Cooperative Study Steering Committee, Solanezumab Study, Group (2014) Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N Engl J Med* **370**, 311-321.

[67] Salloway S, Sperling R, Fox NC, Blennow K, Klunk W, Raskind M, Sabbagh M, Honig LS, Porsteinsson AP, Ferris S, Reichert M, Ketter N, Nejadnik B, Guenzler V, Miloslavsky M, Wang D, Lu Y, Lull J, Tudor IC, Liu E, Grundman M, Yuen E, Black R, Brashear HR, Bapineuzumab 301, 302 Clinical Trial, Investigators (2014) Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. *N Engl J Med* **370**, 322-333.

[68] Vellas B, Carrillo MC, Sampaio C, Brashear HR, Siemers E, Hampel H, Schneider LS, Weiner M, Doody R, Khachaturian Z, Cedarbaum J, Grundman M, Broich K, Giacobini E, Dubois B, Sperling R, Wilcock GK, Fox N, Scheltens P, Touchon J, Hendrix S, Andrieu S, Aisen P, Task Force EU/US/CTAD, Members (2013) Designing drug trials for Alzheimer's disease: What we have learned from the release of the phase III antibody trials: A report from the EU/US/CTAD Task Force. *Alzheimers Dement* **9**, 438-444.

[69] Novakovic D, Feligioni M, Scaccianoce S, Caruso A, Piccinin S, Schepisi C, Errico F, Mercuri NB, Nicoletti F, Nisticò R (2013) Profile of gantenerumab and its potential in the treatment of Alzheimer's disease. *Drug Des Devel Ther* **7**, 1359-1364.

[70] Fischer P, Jungwirth S, Zehetmayer S, Weissgram S, Hoenigschnabl S, Gelpi E, Krampla W, Tragl KH (2007) Conversion from subtypes of mild cognitive impairment to Alzheimer dementia. *Neurology* **68**, 288-291.

[71] Karran E, Hardy J (2014) A critique of the drug discovery and phase 3 clinical programmes targeting the amyloid hypothesis for Alzheimer's disease. *Ann Neurol* **76**, 185-205.

[72] Kurz A, Perneczky R (2011) Amyloid clearance as a treatment target against Alzheimer's disease. *J Alzheimers Dis 24* **S2**, 61-73.

[73] Reitz C (2012) Alzheimer's disease and the amyloid cascade hypothesis: A critical review. *Int J Alzheimers Dis* **2012**, e369808.

[74] Dukart J, Schroeter ML, Mueller K, Initiative TADN (2011) Age correction in dementia–matching to a healthy brain. *PLoS One* **6**, e22193.

[75] Klöppel S, Peter J, Ludl A, Pilatus A, Maier S, Mader I, Heimbach B, Frings L, Egger K, Dukart J, Schroeter ML, Perneczky R, Häussermann P, Vach W, Urbach H, Teipel S, Hüll M, Abdulkadir A (2015) Applying automated MR-based diagnostic methods to the memory clinic: A prospective study. *J Alzheimers Dis* **47**, 939-954.

[76] Lam B, Masellis M, Freedman M, Stuss DT, Black SE (2013) Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. *Alzheimers Res Ther* **5**, 1.