Journal of Alzheimer's Disease 45 (2015) 1061–1076
DOI 10.3233/JAD-142118
IOS Press

An Integrated Bioinformatics Approach for Identifying Genetic Markers that Predict Cerebrospinal Fluid Biomarker p-tau$_{181}$/A$\beta_{1\text{-}42}$ Ratio in ApoE4-Negative Mild Cognitive Impairment Patients

Ying Sun[a,b,*], Anders Bresell[c], Mattias Rantalainen[d], Kina Höglund[a,e], Thibaud Lebouvier[f], Hugh Salter[a,b] and the Alzheimer Disease Neuroimaging Initiative[1]
[a]*AstraZeneca Translational Science Centre, Personalised Healthcare & Biomarkers, Innovative Medicines, AstraZeneca R&D, Sweden*
[b]*Department Clinical Neuroscience, Science for Life Laboratory, Karolinska Institutet, Solna, Sweden*
[c]*Clinical Informatics, CNS/Pain Clinical Development, AstraZeneca R&D Södertälje, Sweden*
[d]*Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden*
[e]*Institute of Neuroscience & Physiology, Department of Psychiatry & Neurochemistry, The Sahlgrenska Academy at University of Gothenburg, Sweden*
[f]*Department of Immunology, Genetics and Pathology, Uppsala University, Sweden*

Handling Associate Editor: Henrik Zetterberg

Accepted 12 January 2015

Abstract. Alzheimer's disease (AD) is the most common form of dementia, with no disease-modifying treatment yet available. Early detection of patients at risk of developing AD is of central importance. Blood-based genetic signatures can serve as early detection and as population-based screening tools. In this study, we aimed to identify genetic markers and gene signatures associated with cerebrospinal fluid (CSF) biomarkers levels of t-tau, p-tau$_{181}$, and with the two ratios t-tau/A$\beta_{1\text{-}42}$ and p-tau$_{181}$/A$\beta_{1\text{-}42}$ in the context of progression from mild cognitive impairment (MCI) to AD, and to identify a panel of genetic markers that can predict CSF biomarker p-tau$_{181}$/A$\beta_{1\text{-}42}$ ratio with consideration of *APOE* ε4 stratification. We analyzed genome-wide the Alzheimer's Disease Neuroimaging Initiative dataset with up to 48 months follow-up. In the first part of the analysis, the main effect of single nucleotide polymorphisms (SNPs) under an additive genetic model was assessed for each of the four CSF biomarkers. In the second part of the analysis, we performed an integrated analysis of genome-wide association study results with pathway enrichment analysis, predictive modeling and network analysis in the subgroup of ApoE4-negative subjects. We identified a panel of five SNPs, rs6766238, rs1143960, rs1249963, rs11975968, and rs4836493, that are predictive for

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://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/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

*Correspondence to: Ying Sun, AstraZeneca Translational Science Centre, Personalised Healthcare & Biomarkers, Innovative Medicines, AstraZeneca R&D, Science for Life Laboratory, Tomtebodavägen 23A, SE-171 65, Solna, Sweden. Tel.: +46 8 55329330; Fax: +46 8 524 81425; E-mail: Ying.Sun2015@gmail.com.

ISSN 1387-2877/15/$35.00

p-tau$_{181}$/Aβ_{1-42} ratio (high/low) with a sensitivity of 66% and a specificity of 70% (AUC 0.74). These results suggest that a panel of SNPs is a potential prognostic biomarker in ApoE4-negative MCI patients.

Keywords: Alzheimer's disease, cerebrospinal fluid, genome-wide association study, mild cognitive impairment, multivariate analysis, pathway analysis, predictive model

INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia, with no disease-modifying treatment yet available. It has been estimated that 35.6 million people lived with dementia worldwide in 2010 and that AD accounts for about 60–80% of these cases [1]. Patients with amnestic mild cognitive impairment (aMCI) have mild but measurable changes in cognitive abilities, especially executive memory, which is considered a prodromal stage of AD [2] when supported by the presence of abnormal biomarkers. The rate of progression from aMCI to AD is up to 10% per year [3]. In an aging society, early detection as well as early therapy is widely considered to be an important goal for researchers. Therefore, aMCI is an important clinical group in which to study longitudinal changes associated with the development of AD.

The emerging criteria for diagnosis of AD require the presence of an appropriate clinical AD phenotype together with one or more pathophysiological biomarker(s) consistent with the presence of AD pathology [4–6]. Biomarkers that can be utilized as surrogate markers of underlying pathological change have become of central importance for detection of early and preclinical AD. Efforts have been made by researchers worldwide to identify and validate different biomarkers for the early diagnosis and/or prediction of progression from MCI to AD, including positron emission tomography (PET) imaging ligands, archetypically Pittsburgh compound B (PiB), which bind to amyloid-β (Aβ), PET imaging with ^{18}F-FDG to measure local glucose metabolism [7–10], structural magnetic resonance imaging (MRI) and cerebrospinal fluid (CSF) biochemical biomarkers, especially Aβ_{1-42}, total tau (t-tau), and tau phosphorylated at threonine 181 (p-tau$_{181}$) either alone, or in combination with imaging and CSF biomarkers [11–14]. It has been reported that the combination of increased CSF concentrations of t-tau or p-tau$_{181}$ and decreased concentration of Aβ_{1-42} improves sensitivity and specificity in the diagnosis of AD, and that these markers are predictive of future conversion from MCI to AD [15–18]. However, both PET imaging and CSF biomarkers have only had limited use in population-based screening, because they are invasive and relatively expensive. Therefore, blood-based biomarkers are needed to develop more affordable and more widely accessible diagnostic and prognostic tests.

Genetic markers may facilitate improved methods for early detection and for patient segmentation, as well as illuminating potential therapeutic avenues. Mutations have been identified in genes that encode the amyloid precursor protein (*APP*), presenilin-1 (*PS1*), and presenilin-2 (*PS2)* in familial AD [19]. In sporadic late-onset AD (LOAD), presence of alleles encoding the apolipoprotein E4 isoform (ApoE4) is a strong risk factor associated with AD [20, 21]. Recently, genome-wide association studies (GWAS) have identified common variants in genes *PLD3, CD2AP, CD33, MS4A/MS4A6E*, and *EPHA1* as novel candidates associated with LOAD diagnosis [22, 23] and other GWASs have identified variants in *PPP3R1* and *MAPT* as associated with progression of AD [24]. However, genetic variants discovered by single-locus based GWAS typically identified variants with small effects sizes in their association with the phenotype, here AD or AD progression, and therefore the individual SNPs do not by themselves constitute potent biomarkers for disease diagnosis and monitoring of AD progression. Multivariate panels of genetic variants may, however, provide more powerful means for diagnostic and prognostic applications. To the best of our knowledge, no studies to date have reported multivariate panels of SNPs for prediction of progression from MCI to AD. Moreover, there is relatively little knowledge about diagnostic accuracy and marker selection in ApoE4-negative patients specifically. A recent study reported that use of CSF biomarkers as a predictor of conversion from MCI to AD performed better in ApoE4-negative subjects than ApoE4-positive subjects [25]. To address these questions, we conducted an integrated analytic approach to search for predictive genetic markers as surrogate markers of CSF biomarkers aimed at improving the possibility to predict progression from MCI to AD alongside *APOE* ε4 status.

In our present study, our strategy was to identify significant pathway-related SNPs derived from GWAS findings, and to construct a multivariate predictive model with selected pathway-related SNPs. We initially applied a GWAS approach to identify common genetic polymorphisms with the strongest association with each one of four quantitative traits, t-tau, p-tau$_{181}$ levels, and the two ratios t-tau/A$\beta_{1\text{-}42}$ and p-tau$_{181}$/A$\beta_{1\text{-}42}$, representing progression/conversion from MCI to AD. Due to limited statistical power as well as limitation of a single-locus analysis approach which may lead to false negatives in respect to SNPs contributing to joint genetic effects, we performed pathway enrichment analysis to select significant pathway-related genes/SNPs. Finally, we applied Random Forest (RF) [26], a machine learning method, to determine a candidate panel of five SNPs with the ability to predict p-tau$_{181}$/A$\beta_{1\text{-}42}$ ratio level (high/low) in ApoE4-negative subjects.

MATERIAL AND METHODS

ADNI

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). ADNI study was launched in 2004 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 MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitivity and specificity of 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. The initial goal of ADNI (ADNI1) was to recruit 800 subjects from over 50 sites across the U.S. and Canada.

Samples and genotyping

Genotype data of the subjects in ADNI cohort, who meet entry criteria for the clinical diagnosis of normal cognition, amnestic MCI or probable AD were downloaded from the LONI website (http://adni.loni.usc.edu/data-samples/access-data/). Population stratification was observed on multidimensional scaling plots of the genome-wide identity-by-state (IBS) pairwise distance matrix (Fig. 1), therefore, only Caucasian subjects were included in further analysis. A total of 177 MCI subjects with 48 months follow-up and available CSF biomarkers data for A$\beta_{1\text{-}42}$, p-tau$_{181}$ and t-tau at baseline were analyzed. Subjects were defined as MCI to AD converter (MCI-con) if they converted from MCI to AD at any time within 48 months and the remainder defined as MCI stable (MCI-stable).

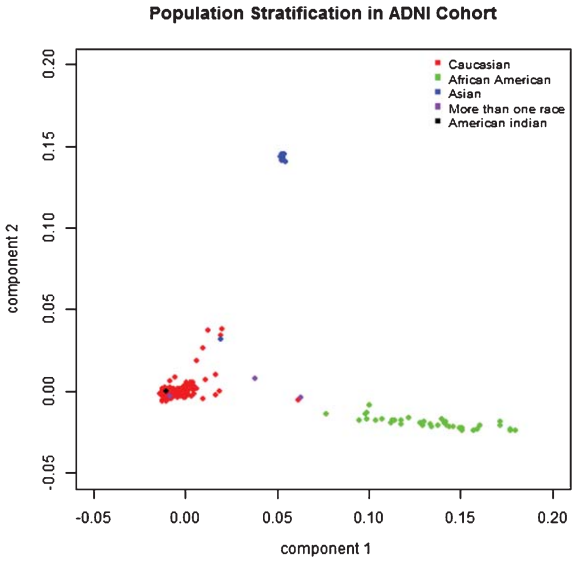

Fig. 1. Multidimensional scaling plot subjects in ADNI. Each dot represents a subject and the distance between dots represents overall genetic similarity calculated using whole genome SNP data. Majority of subjects are self-declared Caucasian (red color), the second large group is African American (green), and third large group is Asian (blue).

Since in this longitudinal study the baseline samples from the same subject were analyzed in different visiting times with other samples from later visits, we estimated and adjusted for any batch effect in the baseline CSF values using a linear model in which batch was included as an adjustment variable. Genotyping data from the Human610-Quad BeadChip (Illumina, Inc., San Diego, CA) included 620,901 SNP and copy number variation (CNV) markers and was completed on all ADNI subjects using the protocol as described previously [27]. All samples also had an *APOE* genotype available in the ADNI database.

Quality control (QC)

The following QC procedures were implemented prior to GWAS analysis. SNPs and individuals were

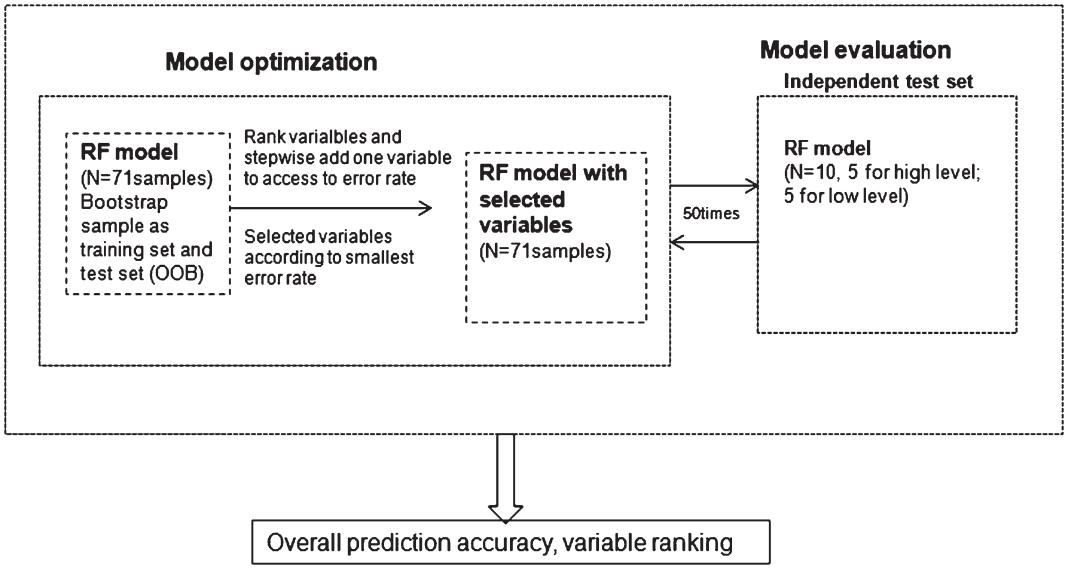

Fig. 2. Nested cross-validation machine learning schema to evaluate model and estimate predication accuracy.

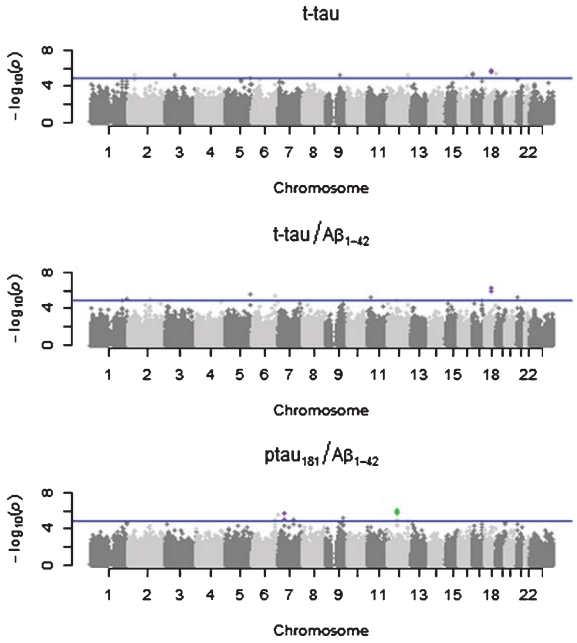

Fig. 3. Manhattan plots of GWAS of tau, tau/A$\beta_{1\text{-}42}$, and ptau$_{181}$/A$\beta_{1\text{-}42}$ as quantitative traits, respectively. The purple dots represent the candidate SNPs which are reached the threshold $p < 5\times10^{-5}$ and at least two SNPs are in high LD (R^2 >0.8) as well. In "t-tau" subplot and "t-tau/A$\beta_{1\text{-}42}$" subplot, rs1445093 and rs12327358 are represented by purple dots. In "ptau$_{181}$/A$\beta_{1\text{-}42}$" subplot, rs11975968 and rs17161127 are represented by purple dots; the SNP rs1249963 from GWAS is represented by green dot.

filtered out if 1) minor allele frequency <5%; 2) genotype rate per SNP <95%; 3) genotype rate per individual <90%; and 4) Hardy-Weinberg Equilibrium $p < 10^{-6}$. In total, 514,932 SNPs were included in subsequent analysis. Genotype calls and QC was performed using PLINK (version 1.07) [28].

A total of 514,932 SNPs and 177 subjects that passed QC were included in GWAS analysis.

Genome-wide association analysis

We selected CSF biomarkers levels of t-tau, p-tau$_{181}$, and the two ratios t-tau/A$\beta_{1\text{-}42}$ and p-tau$_{181}$/A$\beta_{1\text{-}42}$ as quantitative traits (endophenotypes) for GWAS. CSF t-tau, p-tau$_{181}$ concentrations, and the two ratios t-tau/A$\beta_{1\text{-}42}$ and p-tau$_{181}$/A$\beta_{1\text{-}42}$ were all observed to be approximately normally distributed after $\log_2$ transformation.

The main effect of SNPs was assessed on $\log_2$-transformed t-tau, p-tau$_{181}$, and two ratios of $\log_2$-transformed t-tau/A$\beta_{1\text{-}42}$ and p-tau$_{181}$/A$\beta_{1\text{-}42}$ as quantitative traits, separately. Linear models were fitted to identify associations dependent additively upon the minor allele, with adjustment for age, gender, and *APOE* ε4 status. Minor allele homozygotes were coded as 2, heterozygotes coded as 1, and major allele homozygotes were coded as 0. The four models based on four quantitative traits were designated as p-tau$_{181}$ model, t-tau model, p-tau$_{181}$/A$\beta_{1\text{-}42}$ model, and t-tau/A$\beta_{1\text{-}42}$ model. For each trait, the linear regression model used to test for main effect of SNP was:

$$Y = \mu + \beta1^{*}SNP + \beta2^{*}AGE + \beta3^{*}GENDER + \beta4^{*}APOE\,\varepsilon4 + \varepsilon$$

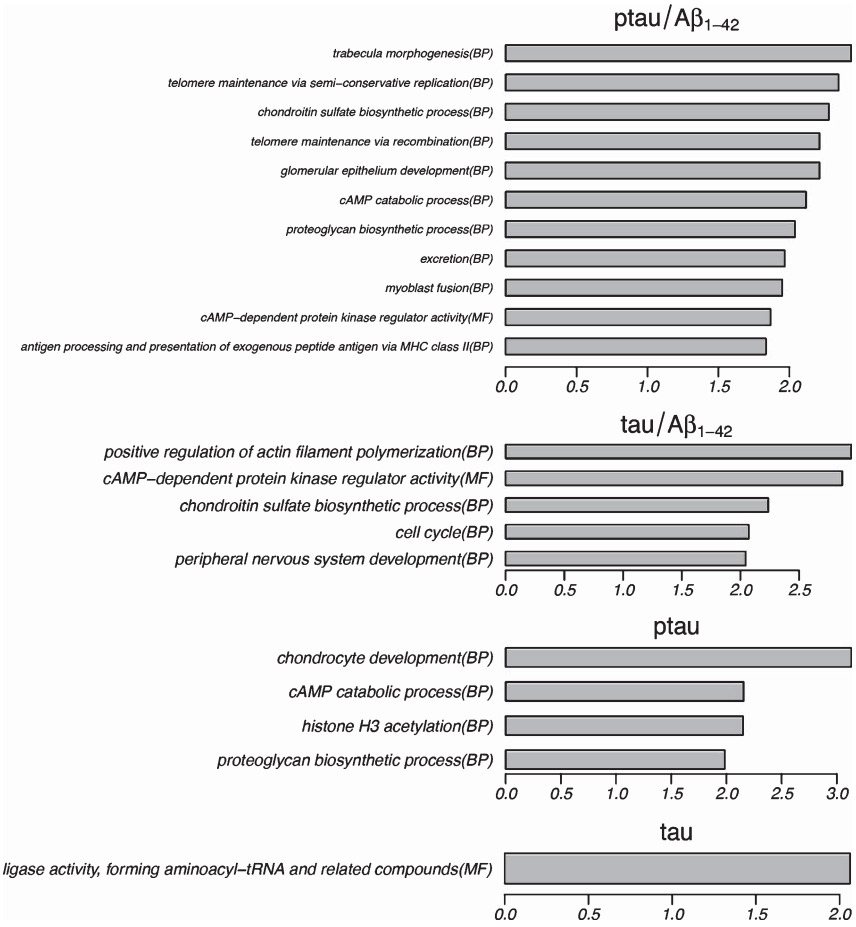

Fig. 4A. Enriched Gene Ontology terms for p-tau$_{181}$. x-axis represents –log$_{10}$(*p*-value).

The false discovery rate (FDR) for reported candidate SNPs were calculated using Benjamini and Hochberg's method [29].

Pathway and network analysis

SNPs from the four GWAS analyses p-tau$_{181}$, p-tau$_{181}$/A$\beta_{1\text{-}42}$ ratio, t-tau, and t-tau/A$\beta_{1\text{-}42}$ were assigned to genes from the NCBI build 36, using ProxyGeneLD which facilitates the conversion of genome-wide genetic marker lists to representative gene lists [30]. The two major features of this software are assigning gene-based significance by accounting for high linkage disequilibrium (LD) structure and correcting the *p*-value according to marker density due to the gene size. By consideration of high LD structure (user-specified r^2) the software iteratively groups SNPs into clusters. Cluster estimation uses LD information to allow for associated markers beyond gene windows and reduces false positive hits by accounting for interdependence of SNPs. Thus, the most significant marker assigned to a specific gene can in turn reside outside of the specified gene boundaries. Two user-specified parameters in software included gene-boundary windows and the LD threshold, which parameters we defined as 1 kb upstream of transcription initiation sites to the end of the 3 UTR of the longest known splice form and $r^2 > 0.8$ for LD threshold, respectively.

In order to characterize the functional role of top genes in the list, and to identify significant pathway-related SNPs for further predictive models, the top three percent genes in each of the four gene lists converted from the four GWAS SNPs lists were used in pathway enrichment analysis with the public database Gene Ontology (GO) and the Ingenuity® pathway analysis. The hypergeometric test and right-tailed Fisher's Exact test to find over-represented pathways were applied to the output from GO and IPA, respectively. Due to the acyclic GO structure, hypergeometric tests

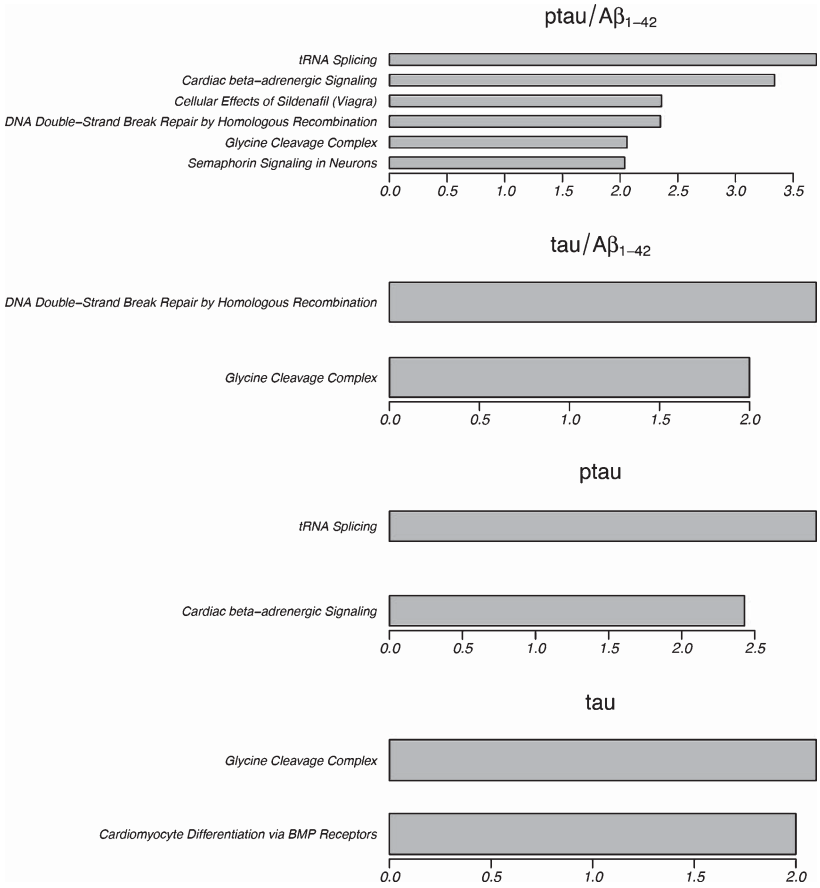

Fig. 4B. Significant canonical pathways from Ingenuity pathway analysis for p-tau$_{181}$ (A), p-tau$_{181}$/A$\beta_{1\text{-}42}$ ratio (B), t-tau (C), and t-tau/A$\beta_{1\text{-}42}$ (D), respectively. x-axis represents –log$_{10}$(p-value).

were performed with consideration for the GO structure, such that tests were first performed for those terms with no child terms. Before testing the terms whose children had already been tested, all genes annotated as significant children from the parent's gene list were removed. This was continued until all terms had been tested. The analysis was restricted to gene sets containing 10–200 genes since small pathways can exhibit spurious phenotype associations due to large single locus effects, and large pathways are more likely to exhibit association by chance alone. Significant pathways were selected with p value $\leq$0.01. We note that ProxyGeneLD as used for conversion of SNPs to genes treated the high LD block as a single signal and chooses the p value of the best single marker (lowest p value in LD block) in that LD block for pre-adjustment significance level, which means that the same marker may be assigned to a number of genes in a high LD block and thus have the potential to inflate pathway enrichment analysis. To avoid this, we manually checked the significant pathway results from the enrichment analysis. If a significant pathway included several genes in a LD block, only one gene was selected at random and the enrichment analysis run again to test if the pathway was still significant. A protein–protein interaction network was constructed based on Ingenuity Knowledge Base using IPAs defined algorithm [31].

Receiver operating characteristic curve analysis

In order to define cut-off points for subsequent binary classification from identified pathways, receiver-operator characteristics (ROC) curves of the MCI-con versus the MCI-stable group were plotted by varying a cutoff from the baseline of t-tau and p-tau$_{181}$ levels and the two ratios p-tau$_{181}$/A$\beta_{1\text{-}42}$ and t-tau/A$\beta_{1\text{-}42}$. AUCs under the ROC curves, sensitivities, specificities, and optimal thresholds were

calculated using the pROC package (version 1.5.4) in the R software environment (version 3.0.0). The optimal threshold point was calculated using the Youden index method [32].

Random Forest multivariate modeling

Random Forest (RF) modeling was applied to develop a predictive model for discrimination between low and high levels of p-tau$_{181}$/Aβ_{1-42} ratio, and to identify an associated biomarker panel consisting of a small number of predictors (SNPs). The SNPs were encoded as 0, 1, and 2, by minor allele count, when used as predictors in the RF model. To jointly evaluate the prediction performance and optimize RF parameters, a nested (double) cross-validation strategy was applied (Fig. 2); the samples were randomly divided into one training set (71 samples) used in the innermost cycle for parameter optimization and one test set (10 samples) used in the outermost cycle for prediction performance assessment. This process was repeated for 50 random partitions of the samples into training and test sets. Within the inner cycle, a RF model was trained, consisting of 1000 trees. A variable importance (VI) score), Mean Decrease Accuracy, was calculated for all variables, and variables were ranked in descending order by VI score. Then, new models were constructed sequentially by stepwise addition of variables, one-at-a-time, and prediction error rates were evaluated. An RF model was optimized, with minimum number of variables (SNPs) and smallest prediction error rate, using the cross-validation training set in each inner cross-validation round. For each outer cross-validation round the samples in the independent test set were classified into low and high level of p-tau$_{181}$/Aβ_{1-42} ratio to assess prediction performance. The final prediction accuracy was calculated as the percentage of correctly classified samples across the 50 outer cross-validation rounds. The average of VI scores was calculated for each variable across the 50 iterations using training sets.

For selection of the most important variables, an overall VI score was calculated using the full set of samples. The correlation between overall VI score and average VI score from 50 iterations was calculated using Spearman's rank correlation test. Analysis was performed with the Random Forest package (version 4.6–7) in the R software environment. In order to validate the statistical significance of predictive performance of AUC from ROC analysis based on the small set of selected predictors, RF models were fitted to the same number of randomly selected SNPs/predictors from GWAS as in the obtained model using a bootstrap approach. This procedure was repeated 100 times in order to obtain AUC values. The empirical p value of the AUC was then estimated as the proportion of sampled AUC values where AUC is greater or equal to the observed AUC value.

High performance computing

GWAS, cross-validation, and bootstrap validation for RF models used the High Performance Computing (HPC) cluster resource at AstraZeneca, Mölndal, Sweden. The HPC consists of Dell M1000E blade centers, 176 nodes with two six-core Xeon processors, with 2–8GB of memory per core. The cluster contains 2,112 cores in total, all connected with Infiniband QDR for node inter-communication and storage access.

RESULTS

Demographic characteristics and CSF biomarkers of patients with MCI in ADNI

From the ADNI1-cohort, 177 MCI subjects with 48 months follow-up time and with a useable CSF sample at baseline were included in the analysis, 81 of these converted to AD and 96 were stable after to 48 months follow-up (Table 1). *APOE* ε4 allele status was significantly associated with MCI to AD conversion according to the one-sided Fisher's exact test ($p = 0.02313$).

Median and mean $\pm$ SD values ($\log_2$ transformed) for p-tau$_{181,}$ t-tau, and the two ratios of t-tau/Aβ_{1-42} and p-tau$_{181}$/Aβ_{1-42} at baseline from the selected MCI-stable and MCI-con group in ADNI are summarized in Table 2. The CSF levels at baseline of p-tau$_{181}$ and the p-tau$_{181}$/Aβ_{1-42} ratio were significantly increased in the MCI-con group as compared with MCI-stable group, respectively (for p-tau$_{181}$: $p = 0.01$; for p-tau$_{181}$/Aβ_{1-42}: $p = 0.006$), but not for t-tau or t-tau/Aβ_{1-42} ($p = 0.21$; $p = 0.097$).

Table 1
Demographic characteristics of MCI subjects who provided a cerebrospinal fluid sample at baseline visit and with 48 months follow-up time in ADNI cohort

	male/female ‘	median AGE at baseline (range)	*APOE* ε4 carrier
MCI-stable ($n = 96$)	66/30	75(55–89)	45 (47%)
MCI-converter ($n = 81$)	52/29	75(55–89)	51 (63%)

Table 2
Cerebrospinal fluid biomarker concentration and ratios for selected MCI subjects at baseline

	ptau$_{181}$	ptau$_{181}$/A$\beta_{1\text{-}42}$	tau	tau/A$\beta_{1\text{-}42}$
MCI-stable				
median (pg/ml)	23.5	0.15	74.60	0.50
mean ± SD	30.3 ± 18.3	0.21 ± 0.17	94.79 ± 63.03	0.65 ± 0.59
MCI-con				
median (pg/ml)	36.25	0.26	94.16	0.68
mean ± SD	37.02 ± 15.82	0.27 ± 0.15	106.70 ± 46.87	0.79 ± 0.48

GWAS of CSF biomarkers t-tau, p-tau$_{181}$ and the two ratios of p-tau$_{181}$/A$\beta_{1\text{-}42}$ and t-tau/A$\beta_{1\text{-}42}$

In order to perform pathway-based analysis and select significant pathway-related SNPs for downstream analysis, GWAS analysis was performed as described above. After quality control, 514,932 SNPs were individually fitted in linear regression models with covariate adjustment for age, gender, and *APOE* ε4 allele status to evaluate the association of SNPs with p-tau$_{181}$ concentration, t-tau concentration, p-tau$_{181}$/A$\beta_{1\text{-}42}$ ratio, and t-tau/A$\beta_{1\text{-}42}$ ratio, respectively. After calculation of genomic inflation factors [33] (p-tau$_{181}$ $\lambda = 1.007$, p-tau$_{181}$/A$\beta_{1\text{-}42}$ $\lambda = 1.012$, t-tau $\lambda = 1.005$, t-tau/A$\beta_{1\text{-}42}$ $\lambda = 1.012$), no inflation was observed, which indicated that population stratification alone was unlikely to account for the GWAS results.

Top-ranked (p value $< 10^{-5}$) candidate SNPs associated with CSF biomarkers in the context of conversion from MCI to AD for the models t-tau, t-tau/A$\beta_{1\text{-}42}$, p-tau$_{181}$/A$\beta_{1\text{-}42}$, and p-tau$_{181}$ are listed in Table 3. SNPs rs1445093 and rs12327358, which are in high LD ($r^2 = 0.97$), were associated with t-tau/A$\beta_{1\text{-}42}$ ratio, with p values of 2.80×10^{-7} and 5.76×10^{-7}, respectively (Fig. 3, purple dots in "t-tau/A$\beta_{1\text{-}42}$" subplot). These two SNPs were also associated with t-tau with p value 3.73×10^{-7} and 6.34×10^{-7} (Fig. 3, purple dots in "t-tau" subplot). SNPs rs1445093 and rs12327358 are located in an intergenic region on chromosome 18 and are located circa 95kb and 90kb upstream, respectively, of the gene encoding the Netrin receptor DCC (*DCC*). From the p-tau$_{181}$/A$\beta_{1\text{-}42}$ model, we observed that rs1249963 was associated with p-tau$_{181}$/A$\beta_{1\text{-}42}$ with p value 8.85×10^{-7} (Fig. 3, green dot in the "ptau$_{181}$/A$\beta_{1\text{-}42}$" subplot). rs1249963 is located in an intergenic region on chromosome 12 and 7kb upstream of *PPP1R1A* (protein phosphatase 1, regulatory (inhibitor) subunit 1A). SNPs rs11975968 (p-value: 1.53×10^{-6}) and rs17161127 (p-value: 7.71×10^{-6}), are located in the first intron of phosphodiesterase 1C, calmodulin-dependent (*PDE1C*) on chromosome 7, and are in high LD ($r^2 > 0.8$) (Fig. 3, purple dots in "ptau$_{181}$/A$\beta_{1\text{-}42}$" subplot).

Pathway analysis

In order to characterize the functional role of SNPs and the corresponding genes and gene products associated with the four traits (p-tau$_{181}$, t-tau, and two ratios p-tau$_{181}$/A$\beta_{1\text{-}42}$ and t-tau/A$\beta_{1\text{-}42}$), respectively, we performed a marker-to-gene conversion and analyzed the top 3% of genes for pathway enrichment analysis for each trait. This procedure yielded four lists of genes associated with each trait ranked by significance. From GO analysis, we identified several statistically significantly enriched (p value <0.01) GO terms in the "biological processes" and "molecular function" that were associated with the top 3% of genes from the p-tau$_{181}$, p-tau$_{181}$/A$\beta_{1\text{-}42}$, t-tau, and t-tau/A$\beta_{1\text{-}42}$ models, respectively (Fig. 4A). Shared enriched GO terms between the p-tau$_{181}$ and p-tau$_{181}$/A$\beta_{1\text{-}42}$ traits were "cAMP catabolism process" and "proteoglycan biosynthesis". Shared enriched GO terms between p-tau$_{181}$/A$\beta_{1\text{-}42}$ and t-tau/A$\beta_{1\text{-}42}$ traits were "cAMP-dependent protein kinase regulation" and "chondroitin sulphate biosynthesis". We also observed that pathways for "positive regulation of actin filament polymerization" (*CDC42EP2, FMN1, RHOA, RAC1, CCL21*, and *CCL24*) and "peripheral nervous system development" (FOXD2, PMP22, ISL2, TBCE, and RUNX1) significantly associated with t-tau/A$\beta_{1\text{-}42}$.

Significantly associated canonical pathways (p value <0.01) were similarly identified using Ingenuity Pathway Analysis (IPA) (Fig. 4B). The "Glycine cleavage complex" pathway was associated to the p-tau$_{181}$/A$\beta_{1\text{-}42}$, t-tau and t-tau/A$\beta_{1\text{-}42}$ models. GCSH (glycine cleavage system H protein) and AMT (aminomethyltransferase) were associated with the "Glycine cleavage complex" pathway (energy production, lipid metabolism). The "cardiac beta-adrenergic signalling" pathway was shared by p-tau$_{181}$ and p-tau$_{181}$/A$\beta_{1\text{-}42}$ traits. A number of genes, *PDE1B, PDE1C, PDE3B, PDE6D*, and *PDE8A*, encoding cyclic nucleotide phosphodiesterases (PDEs) were found to be associated with this pathway.

Predictive models for pathway-derived SNPs/genes associated with p-tau$_{181}$ /A$\beta_{1\text{-}42}$ in ApoE4-negative patients

In order to identify genetic markers predictive of CSF biomarkers that are known in turn to predict

Table 3
SNPs (p value $<1 \times 10^{-5}$) selected from GWAS for each trait. In each model, *represents SNPs in high LD (LD >0.8)

SNPs from GWAS of t-tau				
SNPs name	p_{value}	FDR	chromosome	Gene
rs1445093*	3.73E-07	0.16	18	–
rs12327358*	6.34E-07	0.16	18	–
rs4239351	1.29E-06	0.22	18	–
rs11078506	4.95E-06	0.58	17	–
rs11124499	6.13E-06	0.58	2	–
rs3885648	7.47E-06	0.58	3	TMEM132C
rs1466134	8.01E-06	0.58	16	GPR56
SNPs from GWAS of t-tau/ A$\beta_{1\text{-}42}$				
rs1445093*	2.80E-07	0.14	18	
rs12327358*	5.76E-07	0.15	18	
rs7131051	3.08E-06	0.40	11	NAV2
rs2824765	3.29E-06	0.40	21	TMPRSS15
rs4869001	3.90E-06	0.40	5	–
rs12130076	6.91E-06	0.55	1	–
rs1249963	7.81E-06	0.55	12	–
rs4577811	9.16E-06	0.55	6	PLEKHG1
SNPs from GWAS of p-tau$_{181}$/A$\beta_{1\text{-}42}$				
rs1249963	8,85E-07	0.37	12	–
rs11975968*	1,53E-06	0.37	7	PDE1C
rs10945919	2,13E-06	0.37	6	–
rs2157673	4,32E-06	0.51	9	
rs4895598	7,31E-06	0.51	6	–
rs17161127*	7,71E-06	0.51	7	PDE1C
rs1716355	7,90E-06	0.51	12	GLYCAM1
rs1143960	8,47E-06	0.51	12	*PPP1R1A*
rs2107284	8,91E-06	0.51	7	–
SNPs from GWAS of p-tau$_{181}$.				
rs11975968	2,99E-06	0.55	7	PDE1C
rs12809589	4,45E-06	0.55	12	TMEM132C
rs11795346	4,75E-06	0.55	9	ZNF169
rs10945919	6,92E-06	0.55	6	–
rs11059821	7,87E-06	0.55	12	TMEM132C
rs11795331	8,13E-06	0.55	9	ZNF169

MCI to AD conversion [15, 16], we performed ROC analysis for CSF biomarkers with stratification by *APOE* ε4 allele status. Results showed that all four CSF biomarker classifiers performed best in the ApoE4-negative subgroup (Supplementary Figure 1 and Supplementary Table 1). The optimized cut-offs for p-tau$_{181}$, ptau$_{181}$/A$\beta_{1\text{-}42}$, t-tau, and t-tau/A$\beta_{1\text{-}42}$ to predict MCI to AD conversion for ApoE4-negative subject were 20 pg/ml, 70 pg/ml, 0.13 and 0.25, respectively. The p-tau$_{181}$/A$\beta_{1\text{-}42}$ ratio was selected as response variable (binned response variable for two categories of high/low) in further predictive models considering that of both sensitivity and specificity were greater than 60%, thus higher in comparison with other CSF biomarkers (Supplementary Table 2).

From GO and IPA enrichment analysis, 51 non-redundant genes were identified as related to 49 SNPs (Supplementary Table 3) and associated with GO terms or IPA canonical pathways in turn significantly associated to the ptau/A$\beta_{1\text{-}42}$ ratio trait, as described above. Having identified pathway-associated genes/SNPs in the GWAS analysis, and having identified optimal cutoff points for binary classifiers (low/high level of ptau$_{181}$/A$\beta_{1\text{-}42}$ ratio), we performed predictive modeling in the ApoE4-negative group patients using these 49 SNPs as predictors and low/high level of ptau$_{181}$/A$\beta_{1\text{-}42}$ ratio with respect to the optimized cut-off as the response variable.

An RF model was constructed for prediction of high or low ratio of p-tau$_{181}$/A$\beta_{1\text{-}42}$ in the ApoE4-negative group. Cross-validation was applied to evaluate prediction performance of the RF model, and the average sensitivity was estimated to 66% and specificity to 70%. The AUC value was 0.74 from ROC analysis (p value = 0.01) (Fig. 5). The correlation of variable rankings according to VI score using all samples and average VI score from 50 iteration of cross-validation was highly significant (Spearman's rank correlation test, p-value $<2.2e^{-16}$). Five SNPs rs6766238, rs1143960, rs1249963, rs11975968, and rs4836493 were selected as important variables/SNPs, these SNPs were ranked as top five in both in the

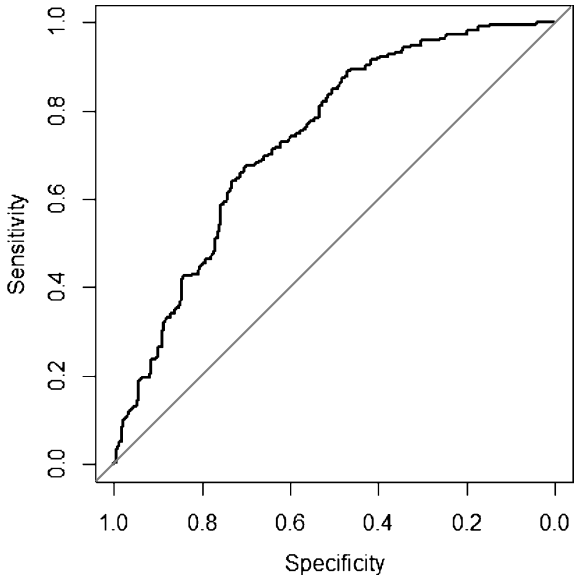

Fig. 5. Receiver operating characteristic analysis for prediction the baseline level of p-tau$_{181}$/A$\beta_{1\text{-}42}$ ratio (high/low) using SNP data based on RF model in ApoE4-negative subjects. The straight diagonal line illustrates prediction by pure chance (AUC 0.5), and the curved line describes the predictions using the RF models when varying the cut-off values (AUC 0.74, p value = 0.01).

ranking list based on overall VI score using all samples and in the list based on the average VI score from cross-validation. The corresponding candidate genes of these SNPs are listed in Table 4 together with the observed protein expression in human brain from the Human Protein Atlas (HPA) [34]. We note that SNPs assigned to a specific gene can reside outside of specified gene boundaries due to consideration of LD structure as discussed above, and that LD blocks may contain more than one gene. Distribution of p-tau$_{181}$/A$\beta_{1\text{-}42}$ ratio by genotype for the five SNPs is shown in Fig. 6.

In order to characterize the functional role of the identified genes, potential interconnections among the seven candidate protein products was explored and a functional network was built upon the Ingenuity pathway knowledge base (Fig. 7). From the protein-protein interaction network, we could observe that several candidate gene products indirectly connected with Ca^{2+}, which might indicate that the progression of AD pathology is correlated with changes of intracellular Ca^{2+} concentration. Most of these genes encode proteins that demonstrate moderate to strong protein expression in cerebral cortex, lateral ventricle, hippocampus, or cerebellum, suggesting that these genes might indeed play an important role in CNS function and in AD pathology.

Table 4
RNA/protein expression level in CNS from Human Protein Atlas database (HPA) for identified candidate genes

Representative SNP	Gene	CNS(brain) from HPA
rs1143960	PDE1B	Strong in cerebral cortex, lateral ventricle, and cerebellum
rs11975968	PDE1C	Strong in cerebral cortex
rs1249963	*PPP1R1A*	Strong in cerebral cortex, hippocampus, lateral ventricle, and cerebellum
rs6766238	AMT	Strong in cerebral cortex, cerebellum and moderate in hippocampus
rs6766238	RHOA	Moderate in cerebral cortex, hippocampus and cerebellum
rs6766238	DAG1	Moderate expression in cerebral cortex
rs4836493	CHSY3	Moderate in cerebral cortex

DISCUSSION

In this study, we developed and applied an integrated analysis approach that combined GWAS, pathway enrichment analysis, and predictive modeling to identify genetic factors predictive of baseline CSF biomarkers, which may thus ultimately be predictive of AD progression from MCI to AD. For single SNP analysis, we applied a conventional GWAS approach to identify top-ranked SNPs associated with CSF biomarkers t-tau, p-tau$_{181}$, t-tau/A$\beta_{1\text{-}42}$, and p-tau$_{181}$/A$\beta_{1\text{-}42}$. In addition, we also performed pathway enrichment analysis in order to get biological insight and take into account potential joint genetic effects. Through these analyses, we could prioritize a limited number of candidate SNPs for building the predictive models. We have been able to identify a panel of five SNPs, rs6766238, rs11975968, rs1143960, rs1249963, and rs4836493, that are predictively informative for baseline p-tau$_{181}$/A$\beta_{1\text{-}42}$ ratio (high/low, cutoff 0.13) with 66% sensitivity and 70% specificity in ApoE4-negative subjects. It is notable that the optimal cutoff used for the predictive model (0.13) is close to the median ratio for p-tau$_{181}$/A$\beta_{1\text{-}42}$ from the MCI ApoE4-negative carrier group (0.12), suggesting that the genetic test could have a significant clinical impact after further validation. rs6766238 is representative for an LD block containing protein coding genes for *RHOA*, *AMT*, and *DAG1*. It has been shown that *RHOA* protein abundance is decreased in the AD brain hippocampus, and *RHOA* colocalized with hyperphosphorylated tau in Pick's disease, a neurodegenerative disorder characterized by hyperphosphorylated tau accumulation [35]. It has been

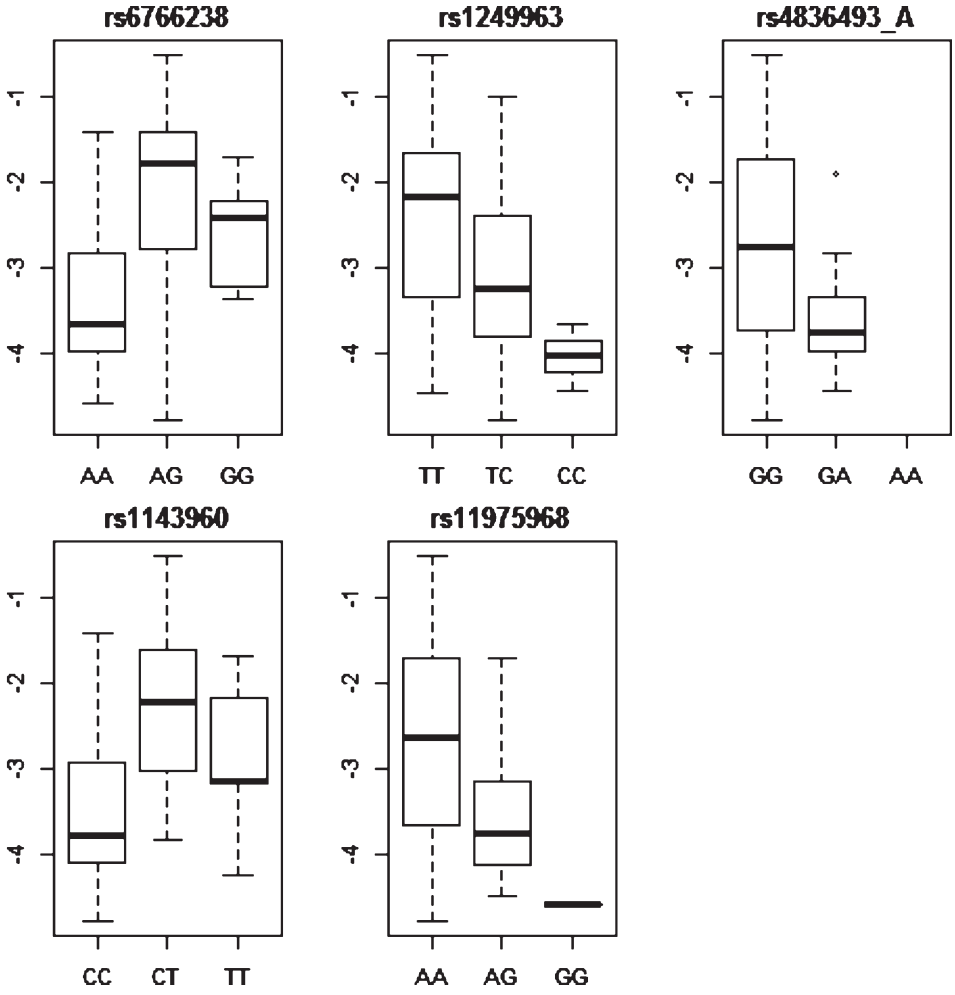

Fig. 6. Boxplots of ratio of ptau/A$\beta_{1\text{-}42}$ (log2 transformed) grouped by genotype for five candidate SNPs in APOE4-negative subjects. Each boxplot is in the order homozygous major allele, heterozygous, homozygous minor allele at each SNP.

shown that *DAG1* encoding protein alpha-dystroglycan exhibits higher levels in AD CSF as compared with normal subjects [36]. rs1143960 is representative for *PDE1B* (phosphodiesterase 1B, Calcium/calmodulin-dependent). rs1249963 is located in intergenic region of genome, but it is representative for *PPP1R1A* due to LD structure. rs11975968 is located in the intronic portion of *PDE1C* (phosphodiesterase 1C, calmodulin-dependent). rs4836493 is located in the intron region of gene *CHSY3* (chondroitin sulfate synthase 3). *PDE1B* and *PDE1C* are members of the PDE1 family. *PDE1* family is activated by the binding of calmodulin in the presence of Ca^{2+} and is capable of acting on both cAMP and cGMP. *PDE1* has several isoforms; *PDE1B* is expressed in brain and mainly in dopaminergic regions [37]. *PDE1C* is highly expressed in the heart, but also found in the CNS [38, 39]. *PPP1R1A* is involved in long-term potentiation of synapses (LTP). It has been shown that deficiency of this gene in mouse causes different degrees of impairment of LTP, indicating that these genes play a role in synaptic plasticity [40]. Bossers et al. showed that using Braak staging for neurofibrillary changes as an objective indicator of progression of AD, *PPP1R1A* expression was decreased in early Braak stages, followed by an increase in expression in later stages [41]. In summary, the genes identified in the model have face validity as potential modulators of calcium regulation and hence control of phosphorylation in neuronal signal transduction, and thus may contribute to the modulation of AD progression by interaction with pathways downstream from Aβ. This study also suggests that pharmacological intervention at these targets may have future utility.

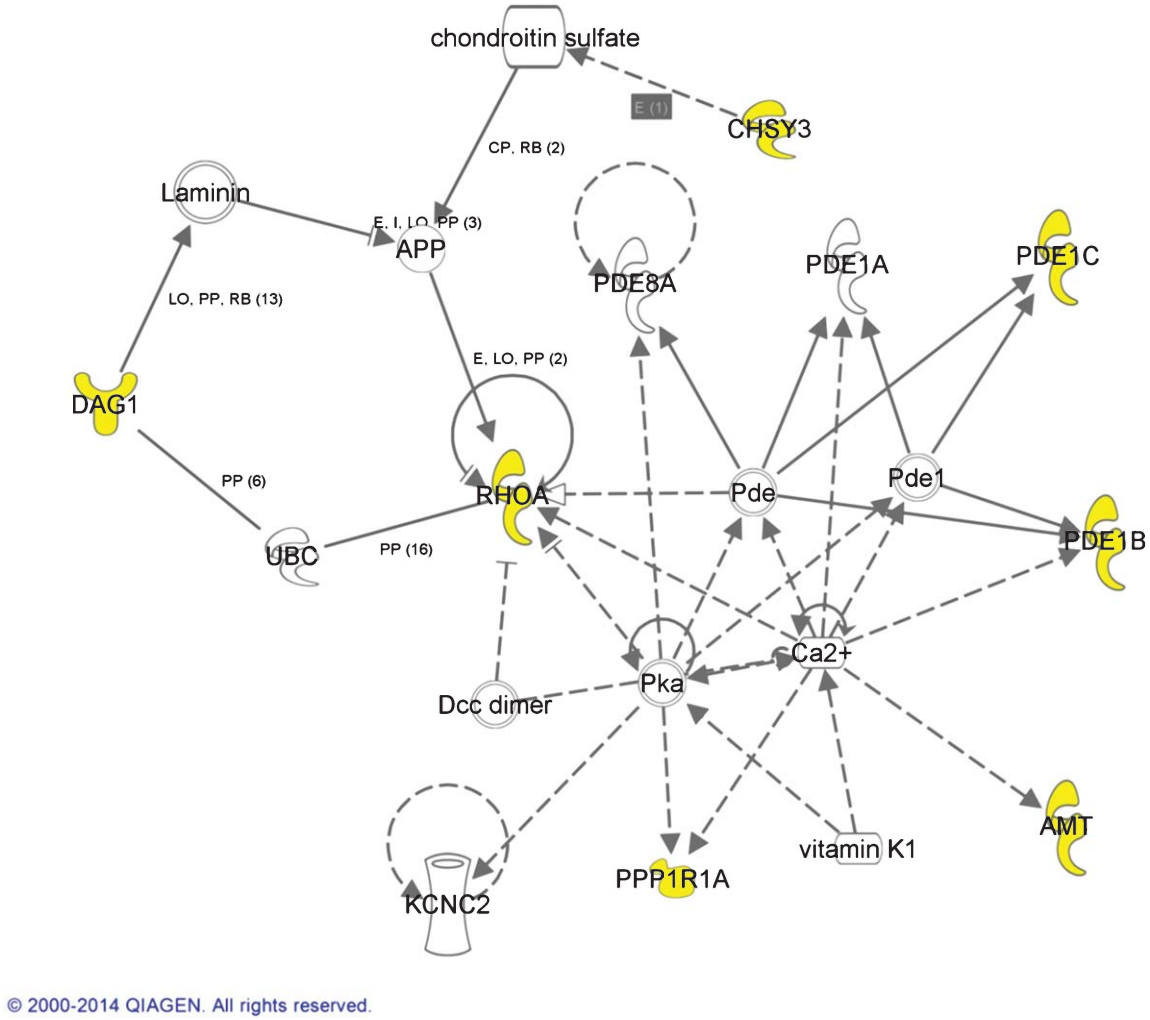

Fig. 7. Protein–protein interaction network from the Ingenuity pathway knowledge base. Protein products of selected candidate genes are highlighted in yellow. A solid line indicates the known experiment-confirmed direct interaction and a dashed line indicates indirect interaction.

In pathway enrichment analysis, proteoglycan biosynthesis was significantly associated with the p-tau$_{181}$ and p-tau$_{181}$/Aβ_{1-42} traits. It is known that proteoglycans are associated with amyloid deposition and it has been reported that heparin sulphate proteoglycans are involved in the pathogenesis of AD [42]. Furthermore, it has been shown that chondroitin sulphate-containing proteoglycan is found in senile plaques of human AD tissue [43]. *NCAN* (neurocan), *CHSY3* (chondroitin sulfate synthase 3), and *CHST7* (carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 7) were associated with the "proteoglycan biosynthetic process" pathway. Interestingly, the pathway of telomere maintenance via semi-conservative replication was significantly associated with p-tau$_{181}$/Aβ_{1-42}. There is some evidence linking shortened telomeres to AD [44, 45].

In this study, we investigated CSF biomarkers as predictors of conversion from MCI to AD, stratified by the presence of the *APOE* ε4 allele. We observed that CSF biomarkers in ApoE4–negative subjects outperformed ApoE4-positive subjects in discriminating MCI-con and MCI-stable. Our observation is supported by data recently published by Apostolova et al. who showed that using CSF biomarker as a classifier to predict conversion from MCI to AD performed better in ApoE4-negative subjects than ApoE4-positive subjects [25]. One explanation based on our analysis is that there are significant differences in baseline levels of CSF biomarkers between MCI-

con and MCI-stable in the ApoE4-negative group, while there is no significant difference between MCI-con and MCI-stable in ApoE4-positive group. We also observed that baseline levels of CSF t-tau, p-tau181, t-tau/Aβ_{1-42}, and p-tau$_{181}$/Aβ_{1-42} were greater in the ApoE4-positive group as compared with the ApoE4-negative group. High CSF t-tau or p-tau$_{181}$ and low CSF Aβ_{1-42} are linked to the tau and Aβ pathologies of AD and *APOE* ε4 is a strong risk factor for AD development.

One of the main aims of this study was to identify candidate genetic markers to predict CSF biomarkers in the context of progression from MCI to AD using an integrated analytics approach rather than single-locus based GWAS approach. Our results provided AUC estimated to 0.74 (p value = 0.01), suggesting that the RF model based on the candidate biomarker panel provides a means to discriminate between high and low ratio of p-tau$_{181}$/Aβ_{1-42} in ApoE4-negative subjects. genes implicated by the identified predictive markers show moderate to strong protein expression in cerebral cortex, lateral ventricle, hippocampus or cerebellum, indicating that these genes might be correlated to CNS function. It is clear that additional work is required to replicate and validate the findings. Replication in additional cohorts and additional analysis to determine likely underlying functional variants to investigate the clinical, histological, and functional cell biological consequences of those changes are all warranted as objectives of further studies.

A limitation of the current study is the small sample size of MCI subjects for GWAS analysis and present lack of an independent cohort for validation. We were limited to 177 subjects in ADNI1 that had CSF biomarkers data and so could be included into this model approach at the time of our study. Taking into account the known heterogeneity of the aMCI clinical diagnosis and the small effect size the comment genetic variants, we will have low power to detect genome-wide significant associations ($p < 5^{*}10^{-7}$) of individual variants in the GWAS analysis using conventional Bonferroni multiple test correction. However, Bonferroni's method is overly conservative because the independence assumption does not hold due to the LD structure among SNPs. Therefore, the top-ranked GWAS SNPs were primarily used to identify significant pathways through pathway analysis. The reported top-ranked SNPs in this study should be considered as potential candidates for replication and validation in future studies. Future studies with potential for replication include ADNI-GO and ADNI-2, where early MCI subjects are being recruited and all patients in ADNI-2 undergo lumbar punctures for CSF data collection, which will increase sample size and statistical power. Moreover, the World Wide ADNI (WW-ADNI) consortium is actively developing broader collaboration efforts to contribute in this community. Large datasets from WW-ADNI are likely to be become available in the future and could provide more replication samples.

Diagnostic criteria in the field of AD remain in development, and even advanced clinical AD remains a probable diagnosis that is not confirmable antemortem. Emerging guidelines suggest that the addition of one or two biomarkers to the clinical status would add value [4], however, no genetic markers are yet validated in this context. Moreover, current biomarker methodologies are invasive and expensive (primarily lumbar puncture and PET imaging). Much attention is thus focused on the identification of less invasive alternatives. The identification of genetic signatures that can complement or even ultimately replace biochemical biomarkers in a diagnostic or prognostic scenario is thus a potential future avenue. Despite the limitation of sample size for this study as we discussed, the final SNP model is designed to demonstrate that a genetic signature panel can predict level of baseline biochemical biomarkers, which in turn are predictive of future conversion of MCI to AD. Hence, the marker panel identified in this study may have utility in screening and/or stratification for future treatments.

ACKNOWLEDGMENTS

The authors wish to acknowledge valuable discussions with Mun-Gwan Hong at SciLifeLab, School of Biotechnology, KTH-Royal Institute of Technology, Stockholm and Andreas Loong in High Performance Computing, AstraZeneca, Mölndal, Sweden.

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; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics,

N.V.; 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; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. 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 Rev December 5, 2013 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.

Authors' disclosures available online (http://j-alz.com/manuscript-disclosures/14-2118r2).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-142118.

REFERENCES

[1] Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP (2013) The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimers Dement* **9**, 63-75.e2.

[2] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999) Mild cognitive impairment: Clinical characterization and outcome. *Arch Neurol* **56**, 303-308.

[3] Manly JJ, Tang MX, Schupf N, Stern Y, Vonsattel JP, Mayeux R (2008) Frequency and course of mild cognitive impairment in a multiethnic community. *Ann Neurol* **63**, 494-506.

[4] Cummings JL, Dubois B, Molinuevo JL, Scheltens P (2013) International Work Group criteria for the diagnosis of Alzheimer disease. *Med Clin North Am* **97**, 363-368.

[5] Dubois B, Epelbaum S, Santos A, Di Stefano F, Julian A, Michon A, Sarazin M, Hampel H (2013) Alzheimer disease: From biomarkers to diagnosis. *Rev Neurol (Paris)* **169**, 744-751.

[6] Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, DeKosky ST, Gauthier S, Selkoe D, Bateman R, Cappa S, Crutch S, Engelborghs S, Frisoni GB, Fox NC, Galasko D, Habert MO, Jicha GA, Nordberg A, Pasquier F, Rabinovici G, Robert P, Rowe C, Salloway S, Sarazin M, Epelbaum S, de Souza LC, Vellas B, Visser PJ, Schneider L, Stern Y, Scheltens P, Cummings JL (2014) Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *Lancet Neurol* **13**, 614-629.

[7] Cohen AD, Klunk WE (2014) Early detection of Alzheimer's disease using PiB and FDG PET. *Neurobiol Dis* **72PA** 117-122.

[8] Herholz K, Westwood S, Haense C, Dunn G (2011) Evaluation of a calibrated (18)F-FDG PET score as a biomarker for progression in Alzheimer disease and mild cognitive impairment. *J Nucl Med* **52**, 1218-1226.

[9] Salas-Gonzalez D, Gorriz JM, Ramirez J, Illan IA, Lopez M, Segovia F, Chaves R, Padilla P, Puntonet CG (2010) Feature selection using factor analysis for Alzheimer's diagnosis using 18F-FDG PET images. *Med Phys* **37**, 6084-6095.

[10] Brown RK, Bohnen NI, Wong KK, Minoshima S, Frey KA (2014) Brain PET in suspected dementia: Patterns of altered FDG metabolism. *Radiographics* **34**, 684-701.

[11] Cui Y, Liu B, Luo S, Zhen X, Fan M, Liu T, Zhu W, Park M, Jiang T, Jin JS, Alzheimer's Disease Neuroimaging Initiative (2011) Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. *PLoS One* **6**, e21896.

[12] De Riva V, Galloni E, Marcon M, Di Dionisio L, Deluca C, Meligrana L, Bolner A, Perini F (2014) Analysis of combined CSF biomarkers in AD diagnosis. *Clin Lab* **60**, 629-634.

[13] 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: Predicting future clinical change. *Neurology* **73**, 294-301.

[14] Samtani MN, Raghavan N, Shi Y, Novak G, Farnum M, Lobanov V, Schultz T, Yang E, DiBernardo A, Narayan VA, Alzheimer's Disease Neuroimaging Initiative (2013) Disease progression model in subjects with mild cognitive impairment from the Alzheimer's disease neuroimaging initiative: CSF biomarkers predict population subtypes. *Br J Clin Pharmacol* **75**, 146-161.

[15] Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, Dean R, Siemers E, Potter W, Lee VM, Trojanowski JQ, Alzheimer's Disease Neuroimaging Initiative (2009) Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol* **65**, 403-413.

[16] Hansson O, Zetterberg H, Buchhave P, Londos E, Blennow K, Minthon L (2006) Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: A follow-up study. *Lancet Neurol* **5**, 228-234.

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

[18] De Meyer G, Shapiro F, Vanderstichele H, Vanmechelen E, Engelborghs S, De Deyn PP, Coart E, Hansson O, Minthon L, Zetterberg H, Blennow K, Shaw L, Trojanowski JQ, Alzheimer's Disease Neuroimaging Initiative (2010) Diagnosis-independent Alzheimer disease biomarker signature in cognitively normal elderly people. *Arch Neurol* **67**, 949-956.

[19] Bertram L, Tanzi RE (2012) The genetics of Alzheimer's disease. *Prog Mol Biol Transl Sci* **107**, 79-100.

[20] Sorbi S, Nacmias B, Forleo P, Latorraca S, Gobbini I, Bracco L, Piacentini S, Amaducci L (1994) ApoE allele frequencies in Italian sporadic and familial Alzheimer's disease. *Neurosci Lett* **177**, 100-102.

[21] Strittmatter WJ, Weisgraber KH, Huang DY, Dong LM, Salvesen GS, Pericak-Vance M, Schmechel D, Saunders AM, Goldgaber D, Roses AD (1993) Binding of human

apolipoprotein E to synthetic amyloid beta peptide: Isoform-specific effects and implications for late-onset Alzheimer disease. *Proc Natl Acad Sci U S A* **90**, 8098-8102.

[22] Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R, Harari O, Norton J, Budde J, Bertelsen S, Jeng AT, Cooper B, Skorupa T, Carrell D, Levitch D, Hsu S, Choi J, Ryten M, UK Brain Expression Consortium, Hardy J, Ryten M, Trabzuni D, Weale ME, Ramasamy A, Smith C, Sassi C, Bras J, Gibbs JR, Hernandez DG, Lupton MK, Powell J, Forabosco P, Ridge PG, Corcoran CD, Tschanz JT, Norton MC, Munger RG, Schmutz C, Leary M, Demirci FY, Bamne MN, Wang X, Lopez OL, Ganguli M, Medway C, Turton J, Lord J, Braae A, Barber I, Brown K, UK Alzheimer's Research, Consortium, Passmore P, Craig D, Johnston J, McGuinness B, Todd S, Heun R, Kolsch H, Kehoe PG, Hooper NM, Vardy ER, Mann DM, Pickering-Brown S, Brown K, Kalsheker N, Lowe J, Morgan K, David Smith A, Wilcock G, Warden D, Holmes C, Pastor P, Lorenzo-Betancor O, Brkanac Z, Scott E, Topol E, Morgan K, Rogaeva E, Singleton AB, Hardy J, Kamboh MI, St George-Hyslop P, Cairns N, Morris JC, Kauwe JS, Goate AM (2014) Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* **505**, 550-554.

[23] Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK, Larson EB, Bird TD, Boeve BF, Graff-Radford NR, De Jager PL, Evans D, Schneider JA, Carrasquillo MM, Ertekin-Taner N, Younkin SG, Cruchaga C, Kauwe JS, Nowotny P, Kramer P, Hardy J, Huentelman MJ, Myers AJ, Barmada MM, Demirci FY, Baldwin CT, Green RC, Rogaeva E, St George-Hyslop P, Arnold SE, Barber R, Beach T, Bigio EH, Bowen JD, Boxer A, Burke JR, Cairns NJ, Carlson CS, Carney RM, Carroll SL, Chui HC, Clark DG, Corneveaux J, Cotman CW, Cummings JL, DeCarli C, DeKosky ST, Diaz-Arrastia R, Dick M, Dickson DW, Ellis WG, Faber KM, Fallon KB, Farlow MR, Ferris S, Frosch MP, Galasko DR, Ganguli M, Gearing M, Geschwind DH, Ghetti B, Gilbert JR, Gilman S, Giordani B, Glass JD, Growdon JH, Hamilton RL, Harrell LE, Head E, Honig LS, Hulette CM, Hyman BT, Jicha GA, Jin LW, Johnson N, Karlawish J, Karydas A, Kaye JA, Kim R, Koo EH, Kowall NW, Lah JJ, Levey AI, Lieberman AP, Lopez OL, Mack WJ, Marson DC, Martiniuk F, Mash DC, Masliah E, McCormick WC, McCurry SM, McDavid AN, McKee AC, Mesulam M, Miller BL, Miller CA, Miller JW, Parisi JE, Perl DP, Peskind E, Petersen RC, Poon WW, Quinn JF, Rajbhandary RA, Raskind M, Reisberg B, Ringman JM, Roberson ED, Rosenberg RN, Sano M, Schneider LS, Seeley W, Shelanski ML, Slifer MA, Smith CD, Sonnen JA, Spina S, Stern RA, Tanzi RE, Trojanowski JQ, Troncoso JC, Van Deerlin VM, Vinters HV, Vonsattel JP, Weintraub S, Welsh-Bohmer KA, Williamson J, Woltjer RL, Cantwell LB, Dombroski BA, Beekly D, Lunetta KL, Martin ER, Kamboh MI, Saykin AJ, Reiman EM, Bennett DA, Morris JC, Montine TJ, Goate AM, Blacker D, Tsuang DW, Hakonarson H, Kukull WA, Foroud TM, Haines JL, Mayeux R, Pericak-Vance MA, Farrer LA, Schellenberg GD (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* **43**, 436-441.

[24] Peterson D, Munger C, Crowley J, Corcoran C, Cruchaga C, Goate AM, Norton MC, Green RC, Munger RG, Breitner JC, Welsh-Bohmer KA, Lyketsos C, Tschanz J, Kauwe JS, Alzheimer's Disease Neuroimaging Initiative (2014) Variants in PPP3R1 and MAPT are associated with more rapid functional decline in Alzheimer's disease: The Cache County Dementia Progression Study. *Alzheimers Dement* **10**, 366-371.

[25] Apostolova LG, Hwang KS, Kohannim O, Avila D, Elashoff D, Jack CR Jr, Shaw L, Trojanowski JQ, Weiner MW, Thompson PM, Alzheimer's Disease Neuroimaging Initiative (2014) ApoE4 effects on automated diagnostic classifiers for mild cognitive impairment and Alzheimer's disease. *Neuroimage Clin* **4**, 461-472.

[26] Breiman L (2001) Random Forests. *Machine Learn* **45**, 5-32.

[27] Saykin AJ, Shen L, Foroud TM, Potkin SG, Swaminathan S, Kim S, Risacher SL, Nho K, Huentelman MJ, Craig DW, Thompson PM, Stein JL, Moore JH, Farrer LA, Green RC, Bertram L, Jack CR Jr, Weiner MW, Alzheimer's Disease Neuroimaging Initiative (2010) Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimers Dement* **6**, 265-273.

[28] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575.

[29] Hochberg YBY (1990) More powerful procedures for multiple significance testing. *Stat Med* **9**, 811-818.

[30] Hong MG, Pawitan Y, Magnusson PK, Prince JA (2009) Strategies and issues in the detection of pathway enrichment in genome-wide association studies. *Hum Genet* **126**, 289-301.

[31] Ingenuity pathway analysis. Network Generation Algorithm. (October 29, 2005).

[32] Youden WJ (1950) Index for rating diagnostic tests. *Cancer* **3**, 32-35.

[33] Bacanu SA, Devlin B, Roeder K (2000) The power of genomic control. *Am J Hum Genet* **66**, 1933-1944.

[34] The Human Protein Atlas, http://www.proteinatlas.org/

[35] Huesa G, Baltrons MA, Gomez-Ramos P, Moran A, Garcia A, Hidalgo J, Frances S, Santpere G, Ferrer I, Galea E (2010) Altered distribution of RhoA in Alzheimer's disease and AbetaPP overexpressing mice. *J Alzheimers Dis* **19**, 37-56.

[36] Yin GN, Lee HW, Cho JY, Suk K (2009) Neuronal pentraxin receptor in cerebrospinal fluid as a potential biomarker for neurodegenerative diseases. *Brain Res* **1265**, 158-170.

[37] Medina AE (2011) Therapeutic utility of phosphodiesterase type I inhibitors in neurological conditions. *Front Neurosci* **5**, 21.

[38] Lakics V, Karran EH, Boess FG (2010) Quantitative comparison of phosphodiesterase mRNA distribution in human brain and peripheral tissues. *Neuropharmacology* **59**, 367-374.

[39] Menniti FS, Faraci WS, Schmidt CJ (2006) Phosphodiesterases in the CNS: Targets for drug development. *Nat Rev Drug Discov* **5**, 660-670.

[40] Allen PB, Hvalby O, Jensen V, Errington ML, Ramsay M, Chaudhry FA, Bliss TV, Storm-Mathisen J, Morris RG, Andersen P, Greengard P (2000) Protein phosphatase-1 regulation in the induction of long-term potentiation: Heterogeneous molecular mechanisms. *J Neurosci* **20**, 3537-3543.

[41] Bossers K, Wirz KT, Meerhoff GF, Essing AH, van Dongen JW, Houba P, Kruse CG, Verhaagen J, Swaab DF (2010) Concerted changes in transcripts in the prefrontal cortex precede neuropathology in Alzheimer's disease. *Brain* **133**, 3699-3723.

[42] van Horssen J, Wesseling P, van den Heuvel LP, de Waal RM, Verbeek MM (2003) Heparan sulphate proteoglycans

in Alzheimer's disease and amyloid-related disorders. *Lancet Neurol* **2**, 482-492.

[43] Canning DR, McKeon RJ, DeWitt DA, Perry G, Wujek JR, Frederickson RC, Silver J (1993) beta-Amyloid of Alzheimer's disease induces reactive gliosis that inhibits axonal outgrowth. *Exp Neurol* **124**, 289-298.

[44] Hochstrasser T, Marksteiner J, Humpel C (2012) Telomere length is age-dependent and reduced in monocytes of Alzheimer patients. *Exp Gerontol* **47**, 160-163.

[45] Cai Z, Yan LJ, Ratka A (2013) Telomere shortening and Alzheimer's disease. *Neuromolecular Med* **15**, 25-48.