Gene expression correlates of hippocampal atrophy in cognitively normal elderly and MCI

Liana G. Apostolova[1,2], Kristy S. Hwang[1,2], Omid Kohannim[1,2], Giovanni Copolla[1], Dimitar Zlatev[2], Fuying Gao[1], Jeffrey L. Cummings[3], Paul M. Thompson[1,2]

[1] Department of Neurology, University of California Los Angeles, Los Angeles, CA, United States
[2] Laboratory of Neuro Imaging, University of California Los Angeles, Los Angeles, CA, United States
[3] Stanford University, Stanford, CA, USA
[3] Lou Ruvo Center for Brain Health, Cleveland Clinic, Las Vegas, NV, USA

Background: Genome-wide association and gene expression studies have revealed the influences of multiple genes on human health. Imaging genetics has also yielded important insights into genetic influences on the biology of Alzheimer's disease (AD). TOMM40, PICALM, CR1 and CLU were recently identified risk genes for late onset AD. Peripheral blood gene expression changes involved in protein metabolism, inflammation, DNA oxidative injury, transcription and translation, and apoptosis have already been described in AD.

Objective: We investigated the accuracy of a novel multimodal biomarker classifier for differentiating cognitively normal elderly (NC) from subjects with amnestic mild cognitive impairment (aMCI). We hypothesized that by combining imaging and genetic biomarker data, we would achieve higher accuracy in differentiating the diagnostic groups.

Patients: We analysed the imaging, genetic and gene expression data data of 46 cognitively NC and 35 amnestic MCI (aMCI) participants in the ImaGene project. Diagnosis of MCI was established using Petersen criteria. Demographic data are listed in **Table 1**.

Table 1. Demographic and volumetric data [mean (SD)]

Variable	NC (N=46)	aMCI (N=35)	P-value
Age, yr	68.8 (7.3)	69.6 (8.8)	0.7
Education, yr	17.5 (2.1)	15.7 (2.4)	**0.001**
Gender, M/F	25:21	15:20	0.3
MMSE	28.9 (1.2)	27.7 (1.7)	**0.001**

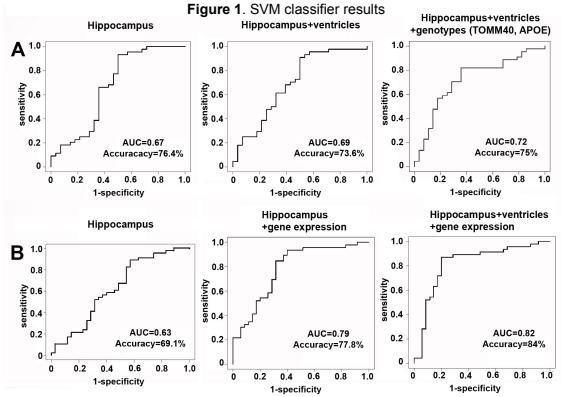
Methods: All magnetic resonance imaging (MRI) scans (TR 28 ms, TE min, 256x192 matrix, FOV 22 cm, 1.5 mm slice thickness, no gaps) were co-registered to the ICBM53 template and intensity normalized. The hippocampi were automatically segmented using AdaBoost as previously described. Briefly, a human expert (LGA) traced the hippocampi of a small representative sample of randomly selected subjects (10 NC and 10 MCI). Using this training set AdaBoost developed an automated segmentation algorithm whereby selecting out of over 14,000 features at each voxel a combination of features that achieved best accuracy classifying voxels as belonging to the hippocampus or not. The algorithm's performance was tested on an independent group of images from the same study. The fully trained AdaBoost model was then used to segment the hippocampi of all subjects in our data set.

The lateral ventricles were automatically segmented using a multi-atlas fluid registration approach. Briefly, a human expert (DZ) manually traced the lateral ventricles of 4 subjects (termed atlases). Using fluid registration techniques each atlas was separately warped to match and thereby extract the shape of the lateral ventricle of each new subject's scan. This step resulted in four lateral ventricle segmentations per subject that were then averaged to create one final ventricular model.

Peripheral blood gene expression (GE) and single nucleotide polymorphism (SNP) data for *APOE, TOMM40, PICALM, CLU, CR1, MAPT* and *PCDH11X* were obtained using standard methods. We first ran correlation analyses to determine the strength of associations between our log2-transformed absolute gene expression levels and hippocampal volume while adjusting for age and gender. We limited our 3D analyses to the top ten positive and top ten negative associations between gene expression and hippocampal volume, but also included eleven additional genes with a known CNS function who ranked in the top 100.

We used a novel automated machine-learning classifier based on a support vector machine (SVM) algorithm to test the ability of hippocampal and ventricular volume, AD risk genes and GE alone and in combination, to differentiate NC from aMCI subjects. We used the e1071 package for classification and regression in R (http://www.r-project.org/). All classifiers included age and gender. We performed cross-validation using the leave-one-out approach and assessed the accuracy and the trade-off between sensitivity and specificity with receiver operating characteristic (ROC) statistics.

Figure 1. SVM classifier results



Results: As hypothesized, combining imaging and genetic biomarkers led to substantially improved NC vs. aMCI classifier performance.

Imaging/SNP classifier: At the time of the analyses we had 72 subjects with SNP data available (44 NC and 28 aMCI). The N=72 Imaging/SNP classifier results can be seen in **Figure 1A**. The classifier that used hippocampal volume only achieved 76.4% diagnostic accuracy (area under the curve, AUC= 0.67) compared to the classifier based on ventricular volume only – accuracy 69.4% (AUC= 0.56) and the combined hippocampal-ventricular classifier – accuracy 74% (AUC= 0.7). The addition of SNP variables led to hippocampal-SNP classifier accuracy of 76% (AUC 0.74) and hippocampal-ventricular-SNP classifier accuracy of 72% (AUC 0.69). Of the 7 SNPs entered, *TOMM40* was selected in both classifiers, *PICALM* was selected in the hippocampal-SNP and *APOE* in the hippocampal-ventricular-SNP classifier only. The remaining SNPs were not included in the optimal classification algorithm.

Imaging/GE classifier: All 81 subjects had GE data available. The N=81 Imaging/GE classifier can be seen in **Figure 1B.** 12 expressed genes and 8 expressed genes were selected as being useful for improving classification, by the final hippocampal-GE and hippocampal-ventricular-GE combined classifiers, respectively. These genes were previously reported to take part in misfolded protein degradation and clearance, protein autophosphorylation, oxidative stress, inflammation, neurogenesis, axonal growth and connectivity, synapse assembly and synaptic transmission and lipid metabolism.

Combined Imaging/SNP/GE classifier: The most comprehensive classifier model (ROC not shown) achieved accuracy of 78.6% and AUC of 0.78. Of the 7 SNP variables only *CLU* and *APOE* were selected.

Conclusion: Combining imaging, genetic and gene expression biomarkers resulted in significantly improved diagnostic accuracy. Automated classifiers show great promise for diagnostic analyses and potentially for predicting future conversion to AD.

References:
1. **Kohannim O, Hua X, Hibar DP, et al.** Boosting power for clinical trials using classifiers based on multiple biomarkers. *Neurobiol Aging* 2010;31(8):1429-42.
2. **Chou Y, Lepore N, Zubicaray GI, et al.** Automated 3D mapping and shape analysis of the lateral ventricles via fluid registration of multiple surface-based atlases. *Proc 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* 2007:1288-91.
3. **Morra JH, Tu Z, Apostolova LG, et al.** Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. *Neuroimage* 2008;43(1):59-68.

This study was generously supported by the Easton Consortium for Alzheimer's Drug Discovery and Biomarker Development, NIA P50 AG16570, NIBIB EB00856, NIBIB EB008281, NIBIB EB007813, NICHD HD 050735, NCRR P41 RR013642.