A signature of cognitive deficits and brain atrophy that is highly predictive of progression to Alzheimer's dementia

Angela Tam[a,b,c,1], **Christian Dansereau**[a,d], **Yasser Iturria-Medina**[c], **Sebastian Urchs**[c], **Pierre Orban**[a,e,f], **John Breitner**[b,c], **Pierre Bellec**[a,d,1], **and for the Alzheimer's Disease Neuroimaging Initiative**[*]

[a]Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, CA; [b]Douglas Hospital Research Centre, McGill University, Montréal, CA; [c]McGill University, Montréal, CA; [d]Département d'Informatique et de recherche opérationnelle, Université de Montréal, Montréal, CA; [e]Centre de Recherche de l'Institut Universitaire en Santé Mentale de Montréal, Montréal, CA; [f]Département de Psychiatrie, Université de Montréal, Montréal, CA

This manuscript was compiled on July 3, 2018

Patients with mild cognitive impairment (MCI) are at risk of progressing to Alzheimer's dementia, yet only a fraction of them do. We explore here whether a very high-risk MCI subgroup can be identified using additional cognitive assessments and structural neuroimaging. A multimodal signature of Alzheimer's dementia was first extracted using machine learning tools in the ADNI1 sample, and was comprised of cognitive deficits across multiple domains as well as atrophy in temporal, parietal and occipital regions. We then validated the predictive value of this signature on two MCI cohorts. In ADNI1 (N=235), the presence of the signature predicted progression to dementia over three years with 80.4% positive predictive value, adjusted for a "typical" MCI baseline rate of 33% (95.6% specificity, 55.1% sensitivity). These results were replicated in ADNI2 (N=235), with 87.8% adjusted positive predictive value (96.7% specificity, 47.3% sensitivity). Our results demonstrate that, even for widely used markers, marked improvement in positive predictive value over the literature can be achieved by focusing on a subgroup of individuals with similar brain characteristics. The signature can be readily applied for the enrichment of clinical trials.

Alzheimer's disease | mild cognitive impairment | prediction | neuroimaging | cognition

DRAFT

Alzheimer's disease (AD) is an age-related neurodegenerative disorder and a leading cause of dementia worldwide. AD is marked by the abnormal accumulation of amyloid β (Aβ) and hyperphosphorylated tau proteins in the brain, which leads to widespread neuronal damage. These neurobiological events may develop years, or even decades, before the emergence of mild cognitive impairment (MCI), considered a prodromal stage of AD. MCI is a clinical condition not specific of AD; many patients will remain cognitively stable over time or even revert to normal cognition. A diagnosis of MCI is still a risk factor for AD dementia, with up to 36% of patients developing dementia within two years (1). Given the long preclinical and prodromal phases in AD, there is interest in developing treatments that could prevent further decline, which requires biomarkers predictive of progression to dementia from preclinical stages. Identifying MCI patients who will progress to AD dementia with enough specificity has been a challenge for clinical trials, where the inclusion criteria for MCI subjects have had low to moderate positive predictive value (2). This lack of prognostic power could be due to underlying heterogeneity among individuals. Previous work has for example shown four distinct patterns of brain atrophy, two of which were associated with higher-than-expected rates of progression to dementia (3). Yet, one subtype was much more at risk than the other. The implications for prognosis are profound: only a subset of individuals will have clinical trajectories that can be predicted very reliably. In this work, we propose to apply machine learning tools to identify subgroups at very high risk of progression to AD dementia, using commonly available measures, i.e. structural magnetic resonance imaging (MRI) and cognitive tests.

There is a growing field of research dedicated to developing machine learning algorithms to automatically detect progression to AD dementia in patients with MCI. Cognitive tests measuring episodic memory and language have been identified as features with high predictive accuracy (4), with specificity and sensitivity values of at least 70%. It has been suggested that the combination of several cognitive domains, notably memory, language, and executive function tests, yields the best performance and balance between specificity and sensitivity (4). With structural MRI, it is well established that significant tissue loss occurs in the medial temporal lobes in AD patients (5). Medial temporal lobe atrophy may be an early marker of AD as it can be seen also in MCI patients relative to cognitively normal individuals (6). As the disease progresses and clinical symptoms become more severe, atrophy

Significance Statement

It has been challenging to reliably identify individuals who will develop Alzheimer's dementia during the years preceding the diagnosis. We describe here a subgroup of patients that are at high risk of being diagnosed with Alzheimer's dementia in the next three years. These patients share a signature composed of spatial patterns of brain atrophy and mild cognitive deficits, which was identified by a recent machine learning method, validated here for the first time on a large sample of subjects. Despite the widely acknowledged heterogeneity of clinical groups, our work demonstrates the existence of a relatively common and high-risk subtype of patients, which can improve subject selection for clinical trials.

P.B. conceived the study. C.D., Y.I-M, S.U., P.O., and A.T. designed the methods. A.T. analyzed data. J.B. reviewed the manuscript. A.T. and P.B. wrote the paper.

The authors declare no conflicts of interest.

[*]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

[1]To whom correspondence should be addressed. E-mail: angela.tam@mail.mcgill.ca; pierre.bellec@criugm.qc.ca

spreads to the parietal and frontal lobes, while leaving the sensorimotor and visual cortices relatively intact (6). Grey matter volumes (GMV) of regions of interest, typically defined by an atlas, have been used as features for the classification of progressor MCI (pMCI) and stable MCI (sMCI) (7, 8). Cortical thickness of brain regions have similarly been used as features to classify sMCI and pMCI (9–11). Models using structural features alone have achieved variable performance, ranging from 38-84% specificity and 63-95% sensitivity. Models that have combined structural MRI markers with cognitive measurements have outperformed models with either structural MRI or cognition alone at predicting progression to AD dementia (12–14). For models combining structural MRI and cognition, in terms of accuracy, the best has been reported at 79% (76% specificity, 83% sensitivity) (13). Overall, the state-of-the-art, which used Aβ positron emission tomography (PET) scans, has so far achieved 82% accuracy (87% specificity, 71% sensitivity) (15). Note that accuracy is defined as the proportion of subjects that were correctly identified, either as positives or negatives. Positive predictive value, on the other hand, is defined as the proportion of true positives out of all subjects that were classified as positive. A model could have high accuracy yet moderate positive predictive value if the number of true positives is low relative to the number of true negatives. In the case of predicting incipient AD, relatively few MCI subjects progress to dementia (up to 36% within two years (1)). Despite promising accuracy, the positive predictive values of models predicting AD progression remain moderate, ranging from 50 to 75% across the literature (16). This implies that up to half of subjects who were identified as progressors by published algorithms did not in fact progress to dementia. Furthermore, predictive power decreases as the time to conversion increases (9), so there is even more room for improvement for long-term prognosis (beyond 2 years).

AD dementia prognosis may be difficult due to the inherent heterogeneity present across the population. While brain amyloidosis is a defining feature of AD, it is clearly insufficient at causing clinical impairment since 10-44% of cognitively normal subjects aged 50 to 90 years are amyloid-positive, based on a meta-analysis of 55 studies with PET and cerebrospinal fluid (17). Furthermore, a meta-analysis of 29 cohorts revealed that, on average, 88% of clinically diagnosed AD patients have amyloid-negative PET scans (18). Similar prevalence rates have been corroborated by post-mortem neuropathological assessments (19, 20), revealing that clinical diagnoses often do not match with pathology. Previous work has also found that guidelines for a neuropathological diagnosis of AD were too restrictive, rendering 33% of cases unclassifiable (19). It is apparent that we need to better characterize heterogeneity within populations if we are to reach the precision levels required for individualized medicine.

To address some of this heterogeneity, several studies have explored subtypes of atrophy in AD, where homogeneous subgroups of AD patients show different patterns of gray matter loss relative to healthy controls. For example, a number of works have reproduced a subtype characterized by primarily medial temporal lobe atrophy, a second subtype characterized by parietal, occipital, and lateral temporal lobe atrophy, and a third subtype with diffuse atrophy (21–24). The subgroup with parietal, occipital, and lateral temporal lobe atrophy tended to be younger and had the worst cognition (21, 23), while the medial temporal lobe atrophy subtype contained a greater proportion of carriers of the APOE4 allele (24), a genetic risk factor for sporadic late-onset AD, and the diffuse atrophy subtype had the least Aβ and tau burden (24). Heterogeneity is also present in the symptomatic expression of MCI patients. Often, MCI patients are categorized as amnestic MCI, subjects who have memory impairment, non-amnestic MCI, those who have impairments in non-memory domains, and multi-domain MCI, subjects with concurrent impairments in multiple kinds of cognition (25). This would result in a patient being labeled as one of four MCI subtypes: 1) amnestic single domain MCI, 2) non-amnestic single domain MCI, 3) amnestic multidomain MCI, 4) non-amnestic multidomain MCI (26). Amnestic MCI subjects are more likely to be diagnosed with dementia than non-amnestic MCI (27, 28), but the group at highest risk of conversion is multi-domain MCI (29). It has been reported 59% of multi-domain MCI subjects progressed after two years of follow-up, while only 18% of amnestic single domain MCI progressed, and 70% of non-amnestic single domain MCI actually improved (30), which illustrates how outcomes can vary substantially between subgroups. Certain subtypes of brain patterns may predispose individuals to worse cognitive trajectories. Subjects with widespread brain atrophy, and especially those who also express greater loss in the temporal lobes, have worse cognitive decline compared to other subjects with more localized atrophy (3).

While brain structure and cognition appear to offer complementary information for prognosis of dementia, the complementarity of those modalities to identify high-risk subtypes is unclear. It is possible that a predictive signature based on atrophy may identify a subgroup of subjects that is separate from those labeled by a signature based on cognition, for example. Then despite the many works that describe subtypes of brain atrophy in AD populations (21–24, 31, 32), little has been done to leverage the heterogeneity to predict incipient dementia from a subgroup of MCI patients with high positive predictive value. We have previously shown that the combined expression of certain subtypes of cortical thinning patterns, grey matter volumes, and resting-state network functional connectivity was highly predictive of future AD progression in a subgroup of MCI subjects with 90% positive predictive value (16). While the results from Dansereau et al. (16) are promising, the sample size was small, as resting-state functional MRI data is not available for the majority of ADNI participants. Large scale validation of a highly predictive signature could still be done using more widely available markers, such as cognitive assessments and structural MRI, but such validation has not been done yet to the best of our knowledge.

The main aim of this work was to develop a multimodal signature that is highly predictive of progression to AD dementia in a subgroup of MCI patients, based on cognition and subtypes of grey matter atrophy. We also aimed to evaluate the complementarity of features derived from cognition and atrophy patterns. Although this has been extensively studied for prognosis of dementia in a general MCI population, the complementarity of these measures is not documented in the specific context of a high risk signature. We applied a cluster analysis on structural magnetic resonance images to identify subtypes of brain atrophy in a sample containing both patients with AD dementia and cognitively normal (CN) individuals. We then used a two-step machine learning algorithm (16) to

DRAFT

train a model to identify three signatures that were highly predictive of AD dementia: 1) an anatomical signature, 2) a cognitive signature, 3) a multimodal anatomical and cognitive signature. After identifying MCI patients carrying these signatures, we examined cognitive decline, Aβ and tau burden, and progression to dementia in these individuals to explore whether a highly predictive signature represented a prodromal stage of AD. We analysed whether these three signatures identified separate subgroups of subjects and how they performed against each other in terms of positive predictive value, specificity, and sensitivity at identifying progressors.

Results

Brain atrophy subtypes. We extracted spatial patterns of brain atrophy and cognitive measures that were used for our predictive models. For the structural MRI measures, we took the baseline T1-weighted MRI scans of patients suffering from AD dementia (N = 205) and cognitively normal (CN) individuals (N = 165) from ADNI1 and applied a voxel-based morphometry (VBM) analysis to generate maps of local grey matter volume. To measure the spatial similarity of brain atrophy between subjects, we generated spatial Pearson's correlation coefficients between all pairs of individual maps of grey matter volume. A clustering analysis on these correlation coefficients revealed the presence of 7 subgroups of subjects with homogeneous grey matter topography (Figure 1). A subtype template was generated for each subgroup by averaging the maps of individuals within that subgroup (Figure 1b). We then generated a measure of similarity between each individual map and each subtype template, using spatial correlation coefficients, referred to as weights (Figure 1a). These weights were used as features for the predictive models involving VBM in the rest of this work. As in our previous works (16, 33), we chose to use weights, which can be interpreted as continuous measures for subtype affinity, over discrete subtype membership because the latter is less informative as most individuals express similarity to multiple subtypes (31). Note that although we chose to present our findings with 7 subtypes, we examined how the number of subtypes may impact our subsequent predictions. There was no significant difference in model performance when we changed the number of subtypes (see Table S1 in supplementary material).

Prediction of AD dementia vs cognitively normal individuals. We used a linear support vector machine (SVM) to develop three different models to classify AD vs CN in ADNI1 based on the following features: 1) VBM-only subtypes, 2) cognitive features from neuropsychological assessments across multiple domains (COG only), 3) a combination of of VBM subtypes and cognitive features (VCOG). In all three models, age, sex, mean grey matter volume (GMV) and total intracranial volume (TIV) were also included as features. The model trained using the VCOG features achieved 94.5% positive predictive value (95.6% specificity, 93.9% sensitivity) when classifying AD vs CN in ADNI1. Such high performance was expected given the marked cognitive deficits associated with clinical dementia. COG features only actually reached excellent performance as well (97.6% positive predictive value, 98.0% specificity, 96.4% sensitivity), while using VBM features only yielded markedly lower performances (86.4% positive predictive value, 89.3% specificity, 79.6% sensitivity) (see Figure 2a). Note that the

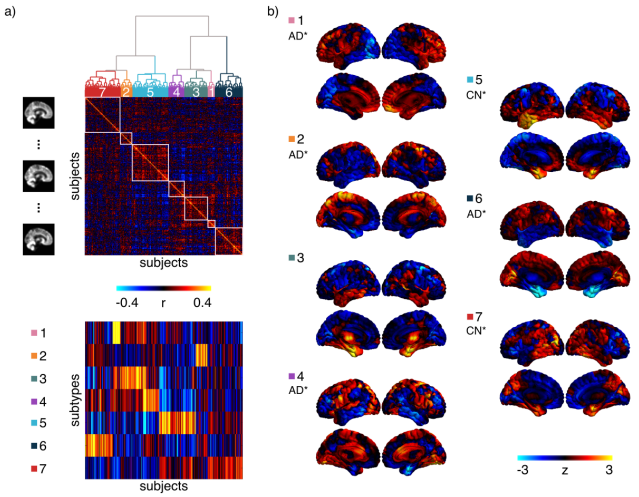

Fig. 1. Subtyping procedure and resulting subtypes. a) A hierarchical clustering procedure identified 7 subtypes, or subgroups, of individuals with similar patterns of grey matter topography within the ADNI1 cohort of CN and AD subjects (top). A measure of spatial similarity, called subtype weight, between a single individual's grey matter volume map and the average of a given subtype was calculated for all individuals and all subtypes (bottom). b) Maps of the 7 subtypes showing the distribution of grey matter across all voxels relative to the average. CN* and AD* denote significant associations between the subtype weights and diagnoses of cognitively normal (CN) or Alzheimer's dementia (AD) respectively.

performance metrics in ADNI1 were estimated through cross-validation, and represent an average performance for several models trained on different subsets of ADNI1. We then trained a model on all of ADNI1, and estimated its performance on an independent dataset, ADNI2. Using VCOG predictors, the ADNI1 model reached 92.0% positive predictive value (96.3% specificity, 92.0% sensitivity), when applied on ADNI2 AD vs CN data. Again the performance was comparable with COG predictors only (92.2% positive predictive value, 96.3% specificity, 94.3% sensitivity), and lower performance with VBM features only (57.3% positive predictive value, 79.8% specificity, 56.7% sensitivity) (see Figure 2a). Note that positive predictive value is dependent on the proportion of patients and controls for a given sensitivity and specificity. Since the ADNI2 sample had a substantially smaller proportion of AD subjects compared to ADNI1, the resulting positive predictive value was reduced. When we adjusted the baseline rate of AD subjects in ADNI2 to the same rate in ADNI1, the positive predictive values were 95.2%, 95.3%, and 70.2% for the VCOG, COG, and VBM models respectively.

Identification of easy AD cases for prediction. We tested the stability of the SVM prediction with VCOG features, using random subsamples of ADNI1. We calculated hit probabilities, representing the frequency at which a given subject would be correctly identified as AD, across all subsamples that did not include that subject as part of the training of the model. When implementing this procedure on all of ADNI1, we found that 88% of subjects had a 100% hit probability, which we defined as easy cases for prediction. A logistic regression was then trained to predict the easy AD cases. The logistic regression was regularized, which means here that it selected a subset of features to perform the prediction. We called the combination of these features a highly predictive signature

NOT PAGE

(HPS) of AD. The VCOG HPS model achieved 99.2% positive predictive value (99.5% specificity, 77.6% sensitivity) in classifying easy AD subjects in ADNI1. These performance scores were estimated by cross-validation of the entire two-stage process (training of SVM, estimation of hit probability, identification of HPS). However, the hyperparameters of the two-stage model were optimized on classifying pMCI vs sMCI in ADNI1, as described below. We next trained a single model on all of ADNI1, which we applied on an independent sample (ADNI2). The ADNI1 AD VCOG HPS model reached 98.6% positive predictive value (99.5% specificity, 79.5% sensitivity) on ADNI2. As was previously observed with the conventional SVM analysis, the VCOG HPS model had similar performance to the COG HPS model (ADNI1: 100% positive predictive value, 100% specificity, 87.3% sensitivity; ADNI2: 98.7% positive predictive value, 99.5% specificity, 88.6% sensitivity), and outperformed the VBM HPS model (ADNI1: 92.3% positive predictive value, 96.1% specificity, 54.6% sensitivity; ADNI2: 65.2% positive predictive value, 91.5% specificity, 33.3% sensitivity), see Figure 2a. When adjusted to the same baseline rate of AD subjects as ADNI1, the positive predictive values in ADNI2 were 99.2%, 99.3%, and 76.7% for the VCOG, COG, and VBM HPS models respectively.

The HPS models consistently outperformed the base SVM classifiers with respect to specificity ($p < 0.001$) in the classifications of AD vs CN and pMCI vs sMCI in both ADNI1 and ADNI2, regardless of the features that the models contained. The HPS also had greater positive predictive value ($p < 0.05$) adjusted for a typical prevalence of 33.6% pMCI in a given sample of MCI subjects (34). However, these increases in specificity and positive predictive value for the HPS model came at a significant cost of reduced sensitivity compared to the base classifier, across all models in both ADNI1 and ADNI2 ($p < 0.05$) (Figure 2). Note that this shift towards lower sensitivity and higher specificity and positive predictive value could be achieved by adjusting the threshold of the SVM analysis (see ROC analysis, Figure S1 in supplementary material), and is not unique to the two-stage procedure we implemented.

High confidence prediction of progression to AD dementia. We next assessed whether the HPS AD model trained on ADNI1 could be used to predict pMCI from sMCI (within three years of follow-up). We applied the AD HPS model trained on ADNI1 on 235 subjects within the ADNI1 MCI sample. We stratified the MCI patients into three categories: negative individuals were classified as controls by the baseline SVM model, HPS+ individuals were identified as patients by the AD HPS model, and non-HPS+ individuals were identified as patients by the baseline SVM model, but not by the AD HPS model. When using the full VCOG features, 87 MCI patients were selected as HPS+ in ADNI1, out of which 81 (93.1% positive predictive value) were pMCI within 36 months of follow-up. This represented a large, significant increase over the baseline rate of progressors in the entire ADNI1 MCI sample (37.4%) ($p < 0.001$). This was also a significant increase over the SVM's predictions, where 83.9% of subjects that it had labeled as hits were true progressors ($p < 0.001$). When adjusted to a 33.6% baseline rate of progressors, more typical of MCI populations, the positive predictive value of HPS+ for prognosis of dementia was 80.4% (93.2% specificity, 55.1% sensitivity).

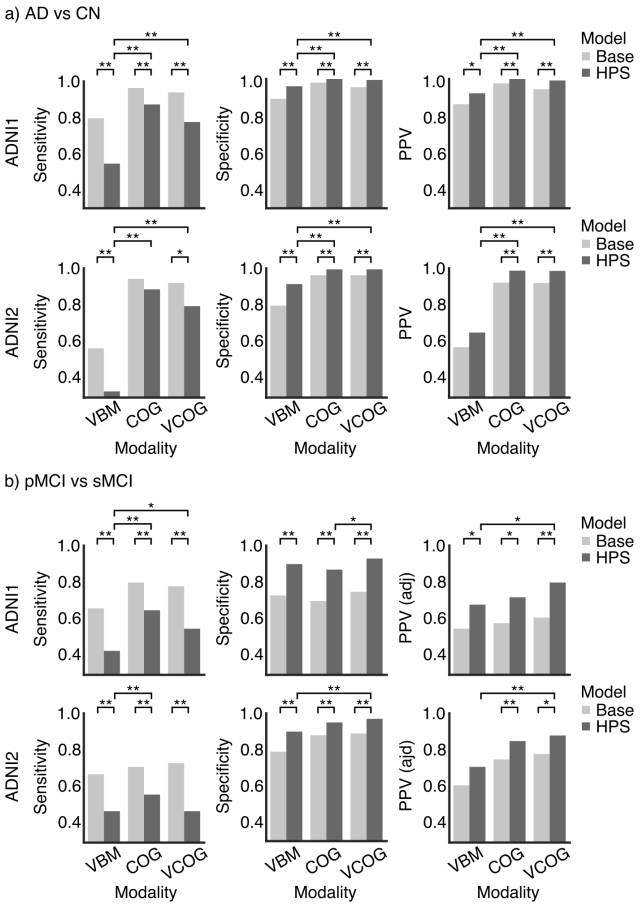

Fig. 2. Specificity, sensitivity, and positive predictive value (PPV) for the base SVM and highly predictive signature (HPS) classifiers in the classifications of a) patients with AD dementia (AD) and cognitively normal individuals (CN) and b) patients with mild cognitive impairment who progress to AD (pMCI) and stable MCI (sMCI) in ADNI1 and ADNI2. VBM represents the model trained with VBM subtypes, COG represents the model trained with baseline cognitive scores, and VCOG represents the model trained with both VBM subtypes and cognition. Significant differences are denoted by $*$ for $p < 0.05$ and $**$ for $p < 0.001$). Positive predictive value was adjusted (PPV (adj)) for a prevalence of 33.6% pMCI in a sample of MCI subjects for both ADNI1 and ADNI2 MCI cohorts.

We replicated these analyses in the MCI sample from ADNI2 (N=235). Using VCOG features, 32 subjects were identified as HPS+, 26 of which progressed to AD dementia within 36 months follow-up (81.2% positive predictive value, specificity of 96.7%, sensitivity of 47.3%, 87.8% positive predictive value adjusted to a 33.6% baseline rate). This represented a significantly higher prevalence than the 30.6% baseline rate in the entire ADNI2 MCI cohort ($p < 0.001$). This was also a significant increase over the SVM's predictions, where 67.8% of subjects it had labeled as hits were true progressors ($p < 0.001$). As expected, the HPS+ model had lower sensitivity and higher specificity/positive predictive value than the baseline model, both in ADNI1 and ADNI2, and for all three feature sets (VBM, COG and VCOG), see Figure 2b. The VCOG features also lead to higher positive predictive value than VBM and COG features taken independently, both in ADNI1 and ADNI2. That increase was large and significant between VCOG and VBM (up to 17%) and marginal and non-significant between VCOG and COG (up to 8%), see Figure 2b.

Characteristics of MCI subjects with a highly predictive VCOG signature of AD. HPS+ MCI subjects with the VCOG signature were more likely to be progressors (Figure 3a) compared to non-HPS+ subjects and negative subjects (ADNI1: $p < 0.001$; ADNI2: $p < 0.001$). HPS+ MCI subjects were also more likely to be APOE4 carriers (Figure 3b) (ADNI1: $p < 0.005$; ADNI2: $p < 0.05$). There was no difference in sex across the HPS groupings in the MCI subjects of either the ADNI1 or ADNI2 cohorts (Figure 3c). This was consistent with the whole sample, where there were equal proportions of progressors across both sexes in each dataset (ADNI1: $\chi^2 = 0.015, p = 0.90$; ADNI2: $\chi^2 = 0.0002, p = 0.99$). The HPS+ class was also significantly enriched for Aβ-positive subjects in ADNI1 ($p < 0.05$). However, this result was not replicated in the ADNI2 MCI subjects (Figure 3d). Similarly with tau, we found a significant increase in tau-positive subjects in the HPS+ group of ADNI1 ($p < 0.05$), but not in ADNI2 (Figure 3e). We found a significant age difference across the HPS classes in ADNI2 ($F = 5.68, p < 0.005$), where the HPS+ subjects were older than the Negative subjects by a mean of 4.4 years. However, age did not differ across the HPS classes in ADNI1 (Figure 3f). Finally, HPS+ subjects had significantly steeper cognitive declines compared to the Non-HPS+ and Negative groups (Figure 3g): there were significant interactions between the HPS groupings and time in ADNI1: (HPS+ $\beta = -0.147, t = -7.56, p < 0.001$; Non-HPS+ $\beta = -0.055, t = -2.46, p < 0.05$) and ADNI2 (HPS+ $\beta = -0.194, t = -8.69, p < 0.001$; Non-HPS+ $\beta = -0.072, t = -3.31, p = 0.001$). The HPS+ in ADNI1 and ADNI2 respectively gained 1.8 and 2.3 more points each year on the ADAS13 compared to the Non-HPS+ and Negative groups. Note that higher scores on the Alzheimer's Disease Assessment Scale - Cognitive subscale (13 items) (ADAS13) represent worse cognitive function.

COG, VBM and VCOG signatures. The COG signature was mainly driven by scores from the ADAS13, which measures overall cognition, ADNI-MEM, a composite score that measures memory (36), and ADNI-EF, a composite score that measures executive function (37) (coefficients were 5.49, -4.80 and -2.50 respectively). In this model, sex, age, mean GMV, and

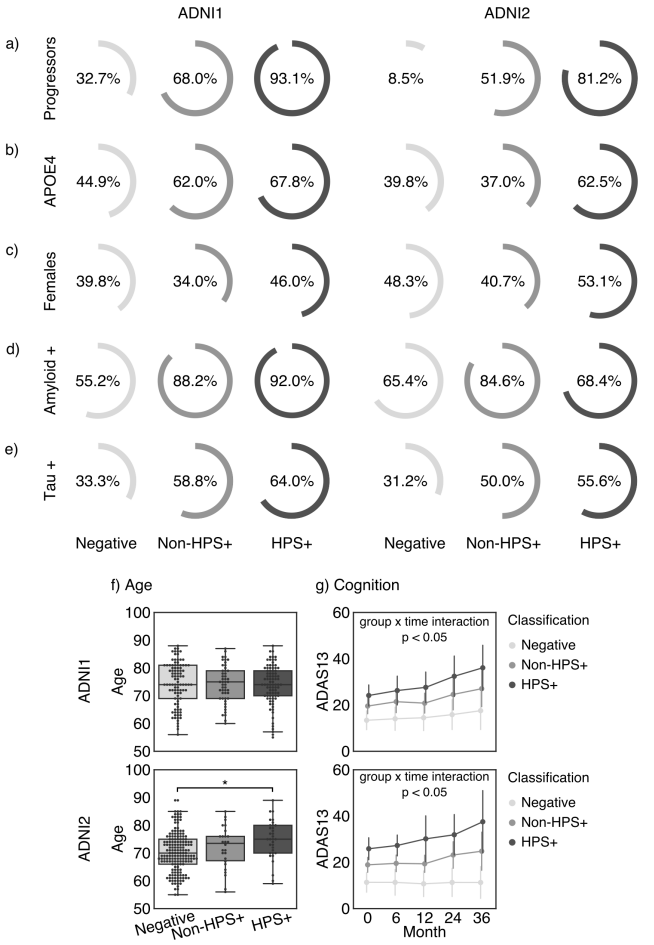

Fig. 3. Characteristics of MCI subjects with the VCOG signature in ADNI1 and ADNI2. We show the percentage of MCI subjects who a) progressed to dementia, were b) APOE4 carriers, c) female, d) positive for Aβ measured by a cut-off of 192 pg/mL in the CSF (35), and e) positive for tau measured by a cut-off of 93 pg/mL in the CSF (35) in each classification (HPS+, Non-HPS+, and Negative). f) Age and g) cognitive trajectories, measured by the Alzheimer's Disease Assessment Scale - Cognitive subscale with 13 items (ADAS13), across the three classes. Significant differences are denoted by $*$ for family-wise error rate-corrected $p < 0.05$.

TIV contributed very little, relative to the cognitive features (Figure 4b). Note that these coefficients should be interpreted as pseudo z-scores as the features had been normalized to zero mean and unit variance.

Almost all grey matter subtypes contributed to the VBM signature. Mean GMV, subtype 1 and subtype 6 had the highest weights in the model (coefficients were -5.07, 4.87, and 3.98 respectively) (Figure 4c). Subtype 1 was characterized by reduced relative GMV in the occipital, parietal and posterior temporal lobes. Subtype 6 was characterized by reduced relative GMV in the temporal lobes, notably the medial temporal regions. We had anticipated the larger contribution of these two subtypes as they have been described in previous AD subtyping work (21–24). Diagnosis (CN, sMCI, pMCI, AD) accounted for a substantial amount of variance in these subtype weights (subtype 1: $F = 8.51, p < 0.001$; subtype 2: $F = 34.27, p < 0.001$). Post-hoc t-tests showed AD subjects had significantly higher weights compared to CN (Figure 1), making these subtypes associated with a diagnosis of AD (subtype 1: $t = 2.88, p < 0.05$; subtype 6: $t = 7.68, p < 0.001$).

The ADAS13, memory (ADNI-MEM) and executive function (ADNI-EF) scores contributed the most to the VCOG HPS signature (coefficients were 6.27, -7.43 and -3.95 respectively, Figure 4a). Of the VBM features, subtypes 2, 3 and 7 contributed the most to the signature (coefficients were 1.36, -2.12 and -2.83 respectively). Subtypes 1 and 6, which had the highest positive weights in the VBM model, were given marginal weights in the VCOG model, which is potentially indicative of redundancy with COG features. Subtype 2, which was associated with AD, was characterized by greater relative GMV in medial parts of the parietal and occipital lobes and the cingulate cortex, but less GMV everywhere else. Subtype 3 was characterized by greater relative GMV in the temporal lobes, insula and striatum. Subtype 7, which was associated with healthy controls, was characterized by greater relative GMV in the parietal, occipital and temporal lobes. Note that the weights for subtypes 3 and 7 were negative in the model, which means that predicted AD and pMCI cases had brain atrophy patterns that were spatially dissimilar to those subtypes.

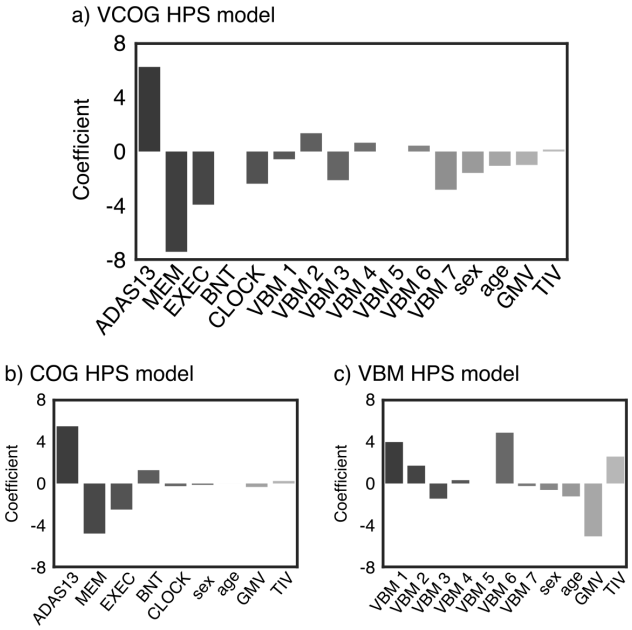

Fig. 4. Coefficients of the high confidence prediction a) VCOG model, b) COG model, and c) VBM model. ADAS13 = Alzheimer's Disease Assessment Scale - Cognitive, MEM = ADNI-MEM score; EXEC = ADNI-EF score, BNT = Boston Naming Test, CLOCK = clock drawing test, VBM 1-7 = VBM subtype weights, GMV = mean grey matter volume, TIV = total intracranial volume.

Discussion

We developed a highly precise and specific MRI and cognitive-based model to predict AD dementia. Our two-stage predictive model reached 93.2% specificity and 93.1% positive predictive value (80.4% when adjusted for 33.6% prevalence of progressors) in ADNI1 when classifying progressor vs stable MCI patients (within 3 years follow-up). We replicated these results in ADNI2 where the model reached 96.7% specificity and 81.2% positive predictive value (87.8% adjusted positive predictive value). With respect to specificity and positive predictive value, these results are a substantial improvement over previous works combining structural MRI and cognition on the same prediction task, that have reported up to 76% specificity and 65% positive predictive value (adjusted for 33.6% prevalence of progressors) (13). We also report the highest positive predictive value compared to the current state-of-the-art predictive model using Aβ PET scans, which reported 74% (adjusted) positive predictive value (15). Finally, our results also reproduced our past work which developed a model that optimizes specificity and positive predictive value (16). Our performance is close to the 90% positive predictive value reported by (16), which used a combination of structural and functional MRI measures and a two-stage predictive model, with the limitation of a smaller sample size (N=56 MCI patients) due to limited availability of functional MRI data in ADNI.

An ideal model to predict conversion to AD dementia would have both high sensitivity and high specificity, meaning it would be able to detect all progressor MCI and reject all stable MCI individuals. However, there is a balance between sensitivity and specificity that should be considered depending on the context. In the case of progression to AD dementia, the pathophysiological heterogeneity of clinical diagnosis will prevent highly accurate prediction linking brain features to clinical trajectories. We argue that, faced with heterogeneity, it is necessary to sacrifice sensitivity to focus on a subgroup of individuals with similar brain abnormalities. In a predictive model, one needs to choose a threshold to control the proportions of false positives and false negatives. A receiver operating characteristic (ROC) curve illustrates the trade off between the true positive and false positive rates as the threshold for positivity changes and is often used to determine an optimal threshold. Generally, increasing the threshold leads to higher specificity and lower sensitivity. The high specificity of our two-stage model indeed came at a cost of reduced sensitivity (55.1% in ADNI1 and 47.3% in ADNI2 for classifying pMCI vs sMCI), which is much lower than sensitivity values of 64%-95% reported by other groups (16). The two stage procedure, based on prediction stability, did not offer gains compared to a simpler SVM model, if the threshold of the SVM model could be selected *a priori* to match the specificity of the two stage procedure (see ROC curves in Figure S1 in supplementary ma-

DRAFT

terial). The two-stage prediction model offered the advantage of a principled approach to train the prediction model in a high-specificity regime, based on stability. The choice of a L1 regularized logistic regression also lead to a compact and interpretable subset of features for the HPS.

Favoring specificity over sensitivity could be useful in settings where false positives need to be minimized and positive predictive value needs to be high, such as expensive clinical trials. Here, with our HPS VCOG model, we report the high positive predictive values for progression to AD from the MCI stage (up to 87.8%, adjusted for 33.6% prevalence of progressors). Importantly, the proposed HPS+ model used tools that are already widely used by clinicians. The present work could be used as a screening tool for recruitment in clinical trials that target MCI subjects who are likely to progress to dementia within three years. The implementation of an automated selection algorithm could also result in groups of MCI subjects with more homogeneous brain pathology. However, we note that HPS+ subjects did not all present with significant amyloid burden (92.0% and 68.4% of HPS+ subjects in ADNI1 and ADNI2 respectively, Figure 3), which means that not all HPS+ individuals are likely to have prodromal AD, even when progressing to dementia.

When we trained our model with cognitive features only, tests for general cognition, memory, and executive function were chosen as the strongest predictors of AD dementia. Our COG HPS model thus supports previous research that reported general cognition, memory, and executive function as important neuropsychological predictors of dementia (4, 29, 38, 39). Compared to the state-of-the-art multi-domain cognition-based predictive model, which reported 87.1% specificity and 81.8% positive predictive value (77.5% when adjusted to 33.6% pMCI prevalence) (40), our COG HPS model achieved similar performance reaching between 87.5%-95% specificity and 72.3%-85.1% (adjusted) positive predictive value. As general cognition was the strongest feature in our model to predict progression, this supports previous findings that MCI patients with deficits across multiple domains are at the highest risk for dementia (29, 30).

For our VBM model, we extracted a number of gray matter atrophy subtypes that recapitulated previously reported subtypes, namely the medial temporal lobe and parietal dominant subtypes (21–24), which were associated strongly with a diagnosis of AD dementia. Weights for the parietal dominant and medial temporal lobe subtypes (Subtypes 1 and 6 from Figure1b, respectively) contributed substantially to the highly predictive signature in the VBM model. The atrophy pattern of subtype 6 is spatially similar to the spread of neurofibrillary tangles in Braak stages III and IV (41), which may support previous findings that tau aggregation mediates neurodegeneration (42). The contributions of the parietal dominant and medial temporal lobe subtypes in the VBM model are also in line with previous works, which have reported that cortical thickness and volumes of the medial temporal lobes, inferior parietal cortex, and precuneus are strong predictors of progression to dementia (9, 13).

When combined with cognitive tests in the VCOG model, the structural subtypes were given marginal weights. This suggests some redundancy between atrophy and cognition, and that cognitive features have higher predictive power than structural features in the ADNI MCI sample. This conclusion is consistent with the observation that the COG model significantly outperformed the VBM model, similar to previous work (13). Although cognitive markers were stronger features, the VCOG model assigned large negative weights for the structural subtypes 3, which showed greater relative GMV in the temporal lobes, and 7, which showed greater relative GMV in the parietal, occipital, and temporal lobes. This means that these features were predictive of stable MCI in the VCOG model, in line with previous work showing that atrophy in these regions is predictive of progression to dementia (9, 13). Furthermore, we demonstrated that combining MRI data with cognitive markers significantly improves upon a model based on MRI features alone. This result is again in line with the literature (14, 39), yet was shown for the first time for a model specifically trained for high positive predictive value. Note that in the current study, the predictive model was trained exclusively on images acquired on 1.5T scanners from ADNI1. Good generalization to ADNI2 with 3T scanners demonstrates robustness of imaging structural subtypes across scanner makes.

While the VCOG model identified the majority of MCI subjects who were highly likely to progress to dementia, there were sizable groups of subjects who were uniquely identified as progressors by the unimodal signatures. For example, out of all the MCI subjects that were picked up by the three HPS models, 12.9% and 35.3% of subjects from ADNI1 and ADNI2 respectively were identified as progressors exclusively by the VBM HPS model (Supplementary Figure S2). Although these subjects presented an AD-like brain atrophy pattern, they may exert compensatory mechanisms or have a high cognitive reserve that may allow them to maintain a higher level of cognitive function than subjects who were also selected by the COG signature. Out of all the HPS+ MCI subjects across the three models, the COG HPS model also exclusively labeled 20.4% and 10.8% of subjects from ADNI1 and ADNI2 respectively as progressors (Supplementary Figure S2). As these subjects were not labeled by the VBM model, they may represent subjects who have patterns of neurodegeneration that are atypical of AD or subjects who have a non-AD pathology. Structural MRI and cognition therefore provide complementary information about AD risk, and it appears that brain atrophy can add to the predictive value of cognitive markers. Our results converge with clinical practices that utilize both neuropsychological testing and brain imaging to inform diagnosis and prognosis, although our approach has the added benefits of being data-driven, automated and optimized for high positive predictive value.

The VCOG highly predictive signature might reflect a late disease stage. We looked at the ratio of early to late MCI subjects in the ADNI2 sample (note that ADNI1 did not have early MCI subjects). Of the MCI subjects identified as HPS+ by the VCOG model, 84.4% were late MCI subjects, compared to a rate of 34.9% of late MCI subjects in the entire ADNI2 MCI sample (Supplementary Figure S3). In both ADNI1 and ADNI2, the VCOG HPS+ subjects had on average the worst baseline cognition and the worst cognitive decline over time (see Figure 3g and Supplementary Figure S4). The HPS+ subjects in ADNI1 also had significantly greater amounts of Aβ and tau pathology compared to the whole sample, although these findings were not replicated in ADNI2 (Figure 3d,e). Given that the majority of HPS+ subjects

DRAFT

were late MCI subjects and had worse cognition, the two-stage predictive model appears to be effective at identifying individuals at later disease stages who are at the cusp of progressing to dementia. This result is to be expected since the model was built on cognitive variables that are designed to separate cognitively normal and individuals with dementia at certain cut-offs. This approach may not be optimal for early detection of future cognitive decline. Training a model to classify MCI progressors and non-progressors to dementia could be done in order to capture future progressors in earlier preclinical stages (e.g. early MCI). Finally, we focused on structural MRI and neuropsychological batteries as features in our models due to their wide availability and established status as clinical tools. However, we believe adding other modalities such as PET imaging, CSF markers, functional MRI, genetic factors, or lifestyle factors could result in higher predictive power, especially at earlier preclinical stages of AD.

Conclusion

In summary, we found a subgroup of patients with MCI who share a signature of cognitive deficits and brain atrophy, that put them at very high risk to progress from MCI to AD dementia within a time span of three years. We validated the signature in two separate cohorts that contained both stable MCI patients and MCI patients who progressed to dementia. The model was able to predict progression to dementia in MCI patients with up to 93.1% positive predictive value and up to 96.7% specificity. The signature was present in about half of all progressors, demonstrating that gains in positive predictive value can be made by focusing on a homogeneous, yet relatively common subgroup. Our model could potentially improve subject selection in clinical trials and identify individuals at a higher risk of AD dementia for early intervention in clinical settings.

Materials and Methods

Data. 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 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. 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 Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org. We took baseline T1-weighted MRI scans from the ADNI1 (228 CN, 397 MCI, 192 AD) and ADNI2 (218 CN, 354 MCI, 103 AD) studies. For a detailed description of MRI acquisition details, see http://adni.loni.usc.edu/methods/documents/mri-protocols/. All subjects gave informed consent to participate in these studies, which were approved by the research ethics committees of the institutions involved in data acquisition. Consent was obtained for data sharing and secondary analysis, the latter being approved by the ethics committee at the CRIUGM. For the MCI groups, each individual must have had at least 36 months of follow-up for inclusion in our analysis. We also further stratified the MCI groups into stable (sMCI), who never received any change in their diagnosis, and progressors (pMCI), who received a diagnosis of AD dementia within 36 months of follow-up. pMCI who progressed to AD dementia after 36 months were excluded. After applying these inclusion/exclusion criteria, we were left with 280 and 268 eligible MCI subjects in ADNI1 and ADNI2 respectively.

Structural features from voxel-based morphometry. Images were processed and analyzed with the NeuroImaging Analysis Kit (NIAK) version 0.18.1 (https://hub.docker.com/r/simexp/niak-boss/), the MINC toolkit (http://bic-mni.github.io/) version 0.3.18, and SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/) under CentOS with Octave (http://gnu.octave.org) version 4.0.2. Preprocessing of MRI data was executed in parallel on the Cedar supercomputer (https://docs.computecanada.ca/wiki/Cedar), using the pipeline system for Octave and Matlab - PSOM (43). Each T1 image was linearly co-registered to the Montreal Neurological Institute (MNI) ICBM152 stereotaxic symmetric template (44), using the CIVET pipeline (45), and then re-oriented to the AC-PC line. Each image was segmented into grey matter, white matter, and CSF probabilistic maps. The DARTEL toolbox (46) was used to normalize the grey matter segmentations to a predefined grey matter template in MNI152 space. Each map was modulated to preserve the total amount of signal and smoothed with with a 8 mm isotropic Gaussian blurring kernel. After quality control of the normalized grey matter segmentations, we were left with 621 subjects in ADNI1 (out of 700, 88.7% success rate) and 515 subjects in ADNI2 (out of 589, 87.4% success rate).

We extracted subtypes to characterize variability of grey matter distribution with the CN and AD samples from ADNI1. In order to reduce the impact of factors of no interest that may have influenced the clustering procedure, we regressed out age, sex, mean grey matter volume (GMV), and total intracranial volume (TIV), using a mass univariate linear regression model at each voxel. We then derived a spatial Pearson's correlation coefficient between all pairs of individual maps after confound regression. This defined a subject x subject (377 x 377) similarity matrix which was entered into a Ward hierarchical clustering procedure (Figure 1a). Based on visual inspection of the similarity matrix, we identified 7 subgroups (Figure 1b). Each subtype was defined as the average map of each subgroup. For each subject, we computed spatial correlations between their map and each subtype, which we call weights (Figure 1a). The weights formed a n subject x n subtypes matrix, which was included in the feature space for all predictive models including VBM throughout this work.

Cognitive features. We took baseline neuropsychological scores for each subject from several cognitive domains: memory from the composite score ADNI-MEM (36), executive function from the composite score ADNI-EF (37), language from the Boston Naming Test (BNT), visuospatial from the clock drawing test, and global cognition from the Alzheimer's Disease Assessment Scale-Cognitive (ADAS13). We chose measures that span multiple cognitive domains as it has been suggested that the use of a combination of neuropsychological measures is likely the best approach to predicting incipient dementia (4). These scores were included as features for the predictive models involving cognition. Thirteen subjects across both ADNI1 and ADNI2 (8 AD, 5 MCI) had to be excluded due to missing values in their cognitive assessments. See Table 1 for demographic information of subjects who were included in analyses.

Prediction of easy AD dementia cases in ADNI1. We trained a SVM model with a linear kernel, as implemented by Scikit-learn (47) version 0.18 to classify AD vs CN from ADNI1 to get a baseline prediction accuracy. A tenfold cross-validation loop was used to estimate the performance of the trained model. Classes were balanced inversely proportional to class frequencies in the input data for the training. A nested cross-validation loop (stratified shuffle split with 50 splits and 20% test size) was used for the grid search of the hyperparameter C (grid was 10^{-2} to 10^{1} with 15 equal steps). We randomly selected subsamples of the dataset (retaining 50% of participants in each subsample) to replicate the SVM training 500 times. For each 50% subsample, a separate SVM model was trained to predict AD or CN in ADNI1. Predictions were made on the remaining 50% of the sample that was not used for training. For each subject, we calculated a hit probability defined as the frequency of correct classification across all SVM replications in which the test set contained that subject. Easy AD cases were defined as individuals with 100% hit probabilities with the AD label. Next, we trained a logistic regression classifier (48), with L1 regularization on the coefficients, to predict the easy AD cases. A stratified shuffle split (500 splits, 50% test size) was used to estimate the performance of the model for the grid search of the hyperparameter C (grid was

DRAFT

Table 1. Demographic information for post-QC subjects in ADNI1 and ADNI2.

ADNI1	CN	sMCI	pMCI	AD
N	205	88	147	165
Age ± SD	76.1 ± 5.0	74.0 ± 7.6	74.3 ± 7.1	75.4 ± 7.5
Female %	51.7	40.9	40.8	51.5
APOE4+ %	27.8	37.5	68.7	65.4
ADAS13 ± SD	9.5 ± 4.3	14.3 ± 5.5	21.3 ± 5.3	28.6 ± 7.1
MMSE ± SD	29.1 ± 1.0	27.7 ± 1.7	26.7 ± 1.7	23.4 ± 2.0
ADNI2	**CN**	**sMCI**	**pMCI**	**AD**
N	188	180	55	89
Age ± SD	72.8 ± 6.1	70.8 ± 7.3	72.1 ± 7.1	74.4 ± 7.8
Female %	54.0	47.8	49.1	46.1
APOE4+ %	29.4	35.6	65.4	71.3
ADAS13 ± SD	9.1 ± 4.2	11.8 ± 5.3	21.4 ± 6.5	31.6 ± 8.7
MMSE ± SD	29.1 ± 1.1	28.4 ± 1.6	27.3 ± 1.9	23.1 ± 2.3

ADAS13 = Alzheimer's Disease Assessment Scale - Cognitive subscale (13 items); MMSE = Mini Mental State Examination

10^{-2} to 10^{1} with 15 equal steps). See (16) for more information about this two-step prediction. We used the whole CN and AD sample from ADNI1 to obtain three highly predictive signatures (HPS), one using VBM subtypes (VBM only), one using cognitive features (COG only), and one using the combination of VBM and cognitive features (VCOG). In all three signatures, age, sex, mean GMV, and TIV were also included as features.

Prediction of progression to AD dementia from the MCI stage in ADNI1. The logistic regression trained on AD vs CN was used to identify MCI patients who have a HPS of AD dementia in ADNI1. We re-trained our models on AD vs CN after optimizing our hyperparameters (resampling size and resampling ratio) in order to maximize specificity and positive predictive value while keeping a minimum of 30% sensitivity for our classification of sMCI (n=89) vs pMCI (n=155) in ADNI1. This was done for all three signatures. In brief, we used the AD and CN sample from ADNI1 as a training set, the MCI subjects from ADNI1 as a validation set, and ADNI2 served as our test set.

Statistical test of differences in model performance. We used Monte-Carlo simulations to generate confidence intervals on the performance (i.e. positive predictive value, specificity and sensitivity) of both base SVM and HPS models for their predictions of AD vs CN and pMCI vs sMCI. Taking the observed sensitivity and specificity, and using similar sample sizes to our experiment, we replicated the number of true and false positive detection 100000 times using independent Bernoulli variables, and derived replications of positive predictive value, specificity and sensitivity. By comparing these replications to the sensitivity, specificity and positive predictive value observed in both models, we estimated a p-value for differences in model performance (49). A p-value smaller than 0.05 was interpreted as evidence of a significant difference in performance, and 0.001 as strong evidence. We also used this approach to compare the performance of the combined features (VCOG) to the models containing VBM features or cognitive features only. Note that, based on our hypotheses regarding the behaviour of the HPS model, the tests were one-sided for increased specificity and positive predictive value, and one-sided for decreased sensitivity.

Statistical tests of association of progression, AD biomarkers, and risk factors in HPS+ MCI subjects. Based on the classifications resulting from the base SVM and HPS models, we separated the MCI subjects into three different groups: 1) HPS+, subjects who were selected by the HPS model as hits, 2) Non-HPS+, subjects who were selected by the base SVM model as hits but were not selected by the HPS model, and 3) Negative, subjects who were not selected as hits by either algorithm.

We tested statistically if the HPS+ subgroup was enriched for progression to dementia, APOE4 carriers, females, and subjects who were positive for Aβ and tau pathology. Positivity of AD pathology was determined by CSF measurements of Aβ 1-42 peptide and total tau with cut-off values of less than 192 pg/mL and greater than 93 pg/mL respectively (35). We implemented Monte-Carlo simulations, where we selected 100000 random subgroups out of the original MCI sample. By comparing the proportion of progressors, APOE4 carriers, females, Aβ-positive, and tau-positive subjects in these null replications to the actual observed values in the HPS subgroup, we estimated a p-value (49) (one sided for increase). A p-value smaller than 0.05 was interpreted as evidence of a significant enrichment, and 0.001 as strong evidence.

One-way ANOVAs were used to evaluate differences between the HPS groupings with respect to age. Post-hoc Tukey's HSD tests were done to assess pairwise differences among the three classes (HPS+, Non-HPS+, Negative). These tests were implemented in Python with the SciPy library (50) version 0.19.1 and StatsModels library (51) version 0.8.0.

To explore the impact of HPS grouping on cognitive trajectories, linear mixed effects models were performed to evaluate the main effects of and interactions between the HPS groups and time on ADAS13 scores up to 36 months of follow-up. The models were first fit with a random effect of participant and then were fit with random slopes (time | participant) if ANOVAs comparing the likelihood ratio suggested a significant improvement in model fit. All tests were performed separately on the ADNI1 and ADNI2 datasets. These tests were implemented in R version 3.3.2 with the library nlme version 3.1.128 (52).

Public code and data. The code used in this experiment is available on a Github repository: https://github.com/SIMEXP/vcog_hps_ad.

ACKNOWLEDGMENTS. We would like to thank Sylvia Villeneuve, Alexa Pichet-Binette and Jacob Vogel for providing us with data to help start our preliminary analyses for this project. We would like to thank Hien Nguyen for advising us on parts of the statistical analyses.

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

The computational resources used to perform the data analysis were provided by Compute Canada (www.computecanada.org). This project was funded by NSERC grant number RN000028 and the Canadian Consortium on Neurodegeneration in Aging (CCNA, www.ccna-ccnv.ca), through a grant from the Canadian Institutes of Health Research and funding from several partners including SANOFI-ADVENTIS R&D. AT is supported by a bursary from the Centre de recherche de l'institut universitaire de gériatrie de Montréal and the Courtois foundation. CD is supported by a salary award from the Lemaire foundation and Courtois foundation. PB is supported by a salary award from "Fonds de recherche du Québec – Santé" and the Courtois foundation.

DRAFT

1. Ward A, Tardiff S, Dye C, Arrighi HM (2013) Rate of conversion from prodromal Alzheimer's disease to Alzheimer's dementia: a systematic review of the literature. *Dementia and Geriatric Cognitive Disorders Extra* 3(1):320–332.
2. Visser PJ, Scheltens P, Verhey FRJ (2005) Do MCI criteria in drug trials accurately identify subjects with predementia Alzheimer's disease? *Journal of neurology, neurosurgery, and psychiatry* 76(10):1348–1354.
3. Dong A, et al. (2017) Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. *Brain* 140(3):735–747.
4. Belleville S, et al. (2017) Neuropsychological Measures that Predict Progression from Mild Cognitive Impairment to Alzheimer's type dementia in Older Adults: a Systematic Review and Meta-Analysis. *Neuropsychology Review* 27(4):328–353.
5. Jack CR, et al. (1997) Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology* 49(3):786–794.
6. McDonald CR, et al. (2009) Regional rates of neocortical atrophy from normal aging to early Alzheimer disease. *Neurology* 73(6):457–465.
7. 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(4):1415–1422.
8. Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ (2011) Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *NBA* 32(12):2322.e19–2322.e27.
9. Eskildsen SF, et al. (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.
10. Querbes O, et al. (2009) Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain* 132(8):2036–2047.
11. Wee CY, Yap PT, Shen D, for the Alzheimer's Disease Neuroimaging Initiative (2013) Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. *Human Brain Mapping* 34(12):3411–3425.
12. Devanand DP, et al. (2007) Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. *Neurology* 68(11):828–836.
13. Korolev S, Safiullin A, Belyaev M, Dodonova Y (2017) Residual and plain convolutional neural networks for 3D brain mri classification. pp. 1–4.
14. Moradi E, et al. (2015) Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage* 104(C):398–412.
15. Mathotaarachchi S, et al. (2017) Identifying incipient dementia individuals using machine learning and amyloid imaging. *Neurobiology of Aging* 59:80–90.
16. Dansereau C, et al. (2017) A brain signature highly predictive of future progression to Alzheimer's dementia. *arXiv preprint arXiv:1712.08058*.
17. Jansen WJ, et al. (2015) Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. *JAMA : the journal of the American Medical Association* 313(19):1924–1938.
18. Ossenkoppele R, et al. (2015) Prevalence of amyloid PET positivity in dementia syndromes: a meta-analysis. *JAMA : the journal of the American Medical Association* 313(19):1939–1949.
19. 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. *Journal of Neuropathology & Experimental Neurology* 71(4):266–273.
20. Bennett DA, et al. (2006) Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* 66(12):1837–1844.
21. Noh Y, et al. (2014) Anatomical heterogeneity of Alzheimer disease Based on cortical thickness on MRIs. *Neurology* 83(21):1936–1944.
22. Hwang J, et al. (2016) Prediction of Alzheimer's disease pathophysiology based on cortical thickness patterns. *Alzheimer's & dementia (Amsterdam, Netherlands)* 2:58–67.
23. Park JY, et al. (2017) Robust Identification of Alzheimer's Disease subtypes based on cortical atrophy patterns. *Scientific Reports* 7:43270.
24. Varol E, Sotiras A, Davatzikos C, Initiative ftADN (2017) HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. *NeuroImage* 145(Part B):346–364.
25. Petersen RC (2003) Mild cognitive impairment clinical trials. *Nature Reviews Drug Discovery* 2(8):646–653.
26. Petersen RC, et al. (2014) Mild cognitive impairment: a concept in evolution. *Journal of internal medicine* 275(3):214–228.
27. Aggarwal NT, Wilson RS, Beck TL, Bienias JL, Bennett DA (2005) Mild cognitive impairment in different functional domains and incident Alzheimer's disease. *Journal of neurology, neurosurgery, and psychiatry* 76(11):1479–1484.
28. Busse A, Hensel A, Gühne U, Angermeyer MC, Riedel-Heller SG (2006) Mild cognitive impairment: long-term course of four clinical subtypes. *Neurology* 67(12):2176–2185.
29. Tabert MH, et al. (2006) Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. *Archives of General Psychiatry* 63(8):916–924.
30. Mitchell J, Arnold R, Dawson K, Nestor PJ, Hodges JR (2009) Outcome in subgroups of mild cognitive impairment (MCI) is highly predictable using a simple algorithm. *Journal of Neurology* 256(9):1500–1509.
31. Zhang X, et al. (2016) Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America* 113(42):E6535–E6544.
32. Dong A, Honnorat N, Gaonkar B, Davatzikos C (2016) CHIMERA: Clustering of Heterogeneous Disease Effects via Distribution Matching of Imaging Patterns. *Medical Imaging, IEEE Transactions on* 35(2):612–621.
33. Orban P, et al. (2017) Subtypes of functional brain connectivity as early markers of neurodegeneration in Alzheimer's disease. *bioRxiv* p. 195164.
34. Mitchell AJ, Shiri-Feshki M (2009) Rate of progression of mild cognitive impairment to dementia - meta-analysis of 41 robust inception cohort studies. *Acta Psychiatrica Scandinavica* 119(4):252–265.
35. Shaw LM, et al. (2009) Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of Neurology* 65(4):403–413.
36. Crane PK, et al. (2012) Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging and Behavior* 6(4):502–516.
37. Gibbons LE, et al. (2012) A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging and Behavior* 6(4):517–527.
38. Dickerson BC, Sperling RA, Hyman BT, Albert MS, Blacker D (2007) Clinical prediction of Alzheimer disease dementia across the spectrum of mild cognitive impairment. *Archives of General Psychiatry* 64(12):1443–1450.
39. Korolev IO, Symonds LL, Bozoki AC, for the Alzheimer's Disease Neuroimaging Initiative (2016) Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification. *PLoS ONE* 11(2):e0138866–25.
40. Belleville S, Gauthier S, Lepage É, Kergoat MJ, Gilbert B (2014) Predicting decline in mild cognitive impairment: A prospective cognitive study. *Neuropsychology* 28(4):643–652.
41. Braak H, Braak E (1991) Neuropathological stageing of Alzheimer-related changes. *Acta neuropathologica* 82(4):239–259.
42. Spillantini MG, Goedert M (2013) Tau pathology and neurodegeneration. *The Lancet Neurology* 12(6):609–622.
43. Bellec P, et al. (2012) The pipeline system for Octave and Matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows. *Frontiers in neuroinformatics* 6:7.
44. Fonov V, et al. (2011) Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage* 54(1):313–327.
45. Ad-Dab'bagh Y, et al. (2006) The CIVET image-processing environment: a fully automated comprehensive pipeline for anatomical neuroimaging research. *Proceedings of the 12th annual meeting of the organization for human brain mapping* p. 2266.
46. Ashburner J (2007) A fast diffeomorphic image registration algorithm. *NeuroImage* 38(1):95–113.
47. Pedregosa F, et al. (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* pp. 2825–2830.
48. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ (2008) LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research* 9(Aug):1871–1874.
49. Phipson B, Smyth GK (2010) Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Statistical applications in genetics and molecular biology* 9(1).
50. Jones E, Oliphant T, Peterson P (2001) SciPy: open source scientific tools for Python, http://www.scipy.org/.
51. Seabold S, Perktold J (2010) Statsmodels: Econometric and statistical modeling with python in *Proceedings of the 9th Python in Science Conference*. (SciPy society Austin).
52. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2018) nlme: linear and nonlinear mixed effects models. R package version 3.1-117, https://CRAN.R-project.org/package=nlme.

DRAFT