Received: 3 May 2024 | Revised: 12 September 2024 | Accepted: 19 September 2024

DOI: 10.1002/dad2.70026

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

A data-driven, multi-domain brain gray matter signature as a powerful biomarker associated with several clinical outcomes

Evan Fletcher[1,2] | Brandon Gavett[1,2] | Sarah Tomaszewski Farias[1,2] | Keith Widaman[3] | Rachel Whitmer[1,2] | Audrey P. Fan[1,2] | Maria Corrada[4] | Charles DeCarli[1,2] | Dan Mungas[1,2]

[1]Department of Neurology, University of California, Davis, California, USA

[2]Alzheimer's Disease Research Center, UC Davis School of Medicine, Davis, California, USA

[3]School of Education, University of California, Riverside, California, USA

[4]Institute for Memory Impairments and Neurological Disorders, Hewitt Hall, Irvine, California, USA

Correspondence
Evan Fletcher, Department of Neurology, University of California, Davis, CA, USA.
Email: emfletcher@ucdavis.edu

Abstract

INTRODUCTION: Characterizing pathological changes in the brain that underlie cognitive impairment, including Alzheimer's disease and related disorders, is central to clinical concerns of prevention, diagnosis, and treatment.

METHODS: We describe the properties of a brain gray matter region ("Union Signature") that is derived from four behavior-specific, data-driven signatures in a discovery cohort.

RESULTS: In a separate validation set, the Union Signature demonstrates clinically relevant properties. Its associations with episodic memory, executive function, and Clinical Dementia Rating Sum of Boxes are stronger than those of several standardly accepted brain measures (e.g., hippocampal volume, cortical gray matter) and other previously developed brain signatures. The ability of the Union Signature to classify clinical syndromes among normal, mild cognitive impairment, and dementia exceeds that of the other measures.

DISCUSSION: The Union Signature is a powerful, multipurpose correlate of clinically relevant outcomes and a strong classifier of clinical syndromes.

KEYWORDS
brain signatures of cognition, clinical measurements, cognitive aging, computationally derived biomarkers

Highlights

- Data-driven brain signatures are potentially valuable in models of cognitive aging.
- In previous work, we outlined rigorous validation of signatures for memory.
- This work demonstrates a signature predicting multiple clinical measures.
- This could be useful in models of interventions for brain support of cognition.

1 | BACKGROUND

Recent years have seen heightened interest in computational approaches to capturing brain atrophy patterns or "brain signatures" that most strongly and/or specifically characterize outcomes such as cognitive function[1] or diagnostic determination.[2–4] The current availability of very large data sets and high computational power has led to active exploration of the properties and generalizability of such data-driven methods.[5,6] The motivating idea is that data-driven computations have the potential to discover brain-behavior relations that explain more variance in clinical outcomes than other approaches guided by theoretical considerations. However, this approach raises issues of reproducibility and validation. The discovered relations and explanatory ability must generalize across separate data sets before they can be used as robust variables.[5,6]

In previous publications, we developed statistically based computational methods for discovering and validating robust brain gray matter (GM) substrates or signatures from T1-weighted magnetic resonance imaging (MRI)[7,8] of episodic memory measured both by neuropsychological testing and informant-rated measures of everyday cognition. These works incorporated principles to support generalizability,[9] including the use of multiple cohorts for independent discovery and validation. We showed that the brain signatures robustly generalized to other non-related cohorts. We also found that episodic memory and third-person informant-rated memory in everyday life were associated with very similar brain signatures,[8] indicating that common brain GM regions, particularly in the medial temporal lobes and caudate, were shared by both of these differing memory measures.

The convergence of underlying signature regions for two clinically relevant outcomes suggested that similar GM substrates might underlie other behavioral outcomes. In addition, recently, a publication from our group[10] using related techniques aimed at predicting clinical status rather than continuous outcome, identified a brain biomarker for participants at increased risk of Alzheimer's disease and related dementias (ADRD). Combined, these findings suggested that it may be possible to identify a generalized brain signature as a useful marker for multiple clinical outcomes. This motivated the current study.

The purpose of the current study is, therefore, to explore the clinically relevant properties of a generalized GM brain signature based on these methods. First, we use these techniques to discover two new domain-specific signatures for executive functions that are neuropsychological and informant based.[11] Next, we compare the spatial GM extents of four signatures—neuropsychological and informant-rated memory + neuropsychological and informant-rated executive function—and evaluate the association of each with all four of those outcomes. We then test whether a common brain signature ("Union Signature"), based on the spatial union of the four signature GM regions, performs as well as the individual signatures in modeling each outcome. From there, we investigate whether the Union Signature has useful clinical properties of strong associations with several relevant measures, including clinical diagnosis and concurrent and change measures of episodic memory, executive performance, and the Clinical Dementia Rating (CDR) scale.[12]

RESEARCH IN CONTEXT

1. **Systematic review**: A PubMed search of "brain signature" produced extensive literature on computational, data-driven approaches to brain associations with outcomes. Earlier algorithmic approaches used straightforward computation to discover brain features that best fit outcomes. Recently, deep learning approaches have used convolutional neural nets, often with very large data sets, to produce powerful models. However, work is ongoing to improve their interpretability.
2. **Interpretation**: Our findings indicate that techniques using algorithmic computation and relatively modest-sized data sets can discover brain substrates that are strongly associated with a range of clinically relevant measures. These afford insight into shared substrates of cognitive measures and provide brain phenotypes that outperform theory-based measures in models of relevant outcomes.
3. **Future directions**: Future work should continue refining signatures and exploring relations to deep-learning methods. Meanwhile, signature phenotypes could be used fruitfully to investigate factors that impact cognitive aging via changes in the brain.

2 | METHODS

2.1 | Imaging cohorts

For discovery of the four domain-specific GM signatures, we used 815 participants from the Alzheimer's Disease Neuroimaging Initiative Phase 3 cohort (ADNI 3) obtained on the ADNI website. These discovery cohorts were used previously for generating memory-specific signatures.[8] ADNI was launched as a public–private partnership by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the U.S. Food and Drug Administration (FDA), private pharmaceutical companies, and non-profit organizations in 2003. Its primary goal 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 MCI and early AD. The principal investigator is Michael Weiner, MD. Current information on ADNI may be obtained at their website, https://adni.loni.usc.edu/. Participants in our ADNI 3 cohort had neuropsychological and informant-rated daily function (everyday cognition or ECog) data along with one structural T1 MRI scan taken close to the time of evaluation (mean time difference, 0.43 years).

For validation and exploration of the signature properties, we used a sample that combined several cohorts including the University of California Davis Alzheimer's Disease Research Center (ADRC)

Longitudinal Cohort, the Kaiser Healthy Aging and Diverse Life Experiences Study (KHANDLE), the Study of Healthy Aging in African Americans (STAR), and Life After 90 (LA90) cohorts, consisting of 1874 participants each having repeated behavioral measurements and a single concurrent MRI scan, selected to be closest to the earliest behavioral evaluation (mean time difference, 0.92 years). We refer to this as the UC Davis (UCD) sample. Our UCD sample was racially/ethnically diverse, consisting of 185 of Asian origin, 618 African Americans, 404 Latinx/Hispanics, and 660 Whites, with 7 unclassified. For clinical diagnoses, the UCD sample consisted of 946 cognitively normal (CN), 418 mild cognitive impairment (MCI), and 140 with dementia. We omitted 370 without a diagnosis.

2.2 | Cognitive and everyday function assessments

Assessments of neuropsychological episodic memory and executive function used the Spanish and English Neuropsychological Assessment Scales (SENAS)[13,14] in the UCD cohort and ADNI-Mem, the ADNI measure of episodic memory[15] and ADNI-EF, the measure of executive function[16] in ADNI 3. The SENAS has undergone extensive development as a battery of cognitive tests relevant to cognitive aging that allow for valid comparisons across racial, ethnic, and language groups.[13,14,17,18] Development and validation of the SENAS used modern psychometric methods to ensure highly reliable measurement across a diverse range of abilities and ages.

The ECog third-person informant-rated scales for memory and executive function were used in both ADNI 3 and UCD data sets.[11] The ECog measures cognitively relevant everyday abilities covering multiple domains. For this study, we used the domains of Everyday Memory and Everyday Executive Functioning. For each item of the ECog, informants compare the participant's current level of everyday functioning with how they functioned at their own baseline. The ECog has been shown to have excellent psychometric properties including good test–retest reliability, as well as evidence of various aspects of validity including content and construction.[11,19]

The CDR[12] was used as a measure of global disease severity. It is a clinician rating based on a structured interview with the participant and knowledgeable informant across multiple areas of functioning including memory, orientation, judgment, problem-solving, community affairs, home and hobbies, and personal care. The sum of domains or "Sum of Boxes" (CDR-SB) was used as a continuous measure of clinical status.

2.3 | MRI image processing

Single T1-weighted MRI scans for each participant in the UCD test set were processed by in-house pipelines developed at the IDeA Laboratory, Department of Neurology, University of California Davis, and described previously.[20,21] Steps relevant to the current project included affine transformation followed by nonlinear, deformable B-spline registration to a common structural MRI template space.[22] This registration enabled native space automatic segmentation into GM, white matter (WM), and cerebrospinal fluid (CSF) tissue classes. Thus the segmented tissues of the template were transformed to native space by inverting the registration and then used to initialize an iterative Bayesian algorithm for optimizing estimates of native tissue classes.[23] Our template was constructed as an age-appropriate (i.e., cognitively normal, mean age 65 years) minimal deformation synthetic template (MDT).[24]

We quantified brain GM using thickness measures computed at voxel-level in native space by the DiReCT diffeomorphic algorithm[25] applied to segmented images. DiReCT is a three-dimensional (3D), voxel-based algorithm that is amenable to the voxel aggregation at the base of our method, and thus preferable to other common approaches, such as the surface-based (i.e., calculating distances between inner and outer GM polygonal mesh surfaces) approach used in Freesurfer.[26] Native GM thickness maps were deformed to MDT space via the nonlinear deformations in our standard pipeline. All subsequent analyses for signature region discovery and validation were performed over the deformed GM regions of participants in common MDT space.

2.4 | Signature variables

The discovery and validation technique for GM signature regions has been explained in detail previously.[8] To summarize here, the discovery phase uses 40 randomly selected subsets, each of 400 samples from the full discovery cohort, to compute regions of interest (ROIs) that are significantly associated with behavioral outcome, separately in each subset. Discovery is followed by a consolidation phase. Clusters from the 40 discovery sets are tested for voxelwise overlaps. Voxels contained in at least 70% of the 40 discovery subset clusters are retained as "consensus" regions. We performed the steps leading to consensus regions in parallel for three significance thresholds, corresponding to three levels of *A*-association ($t = 3, 5, 7$). Thus consensus regions were formed independently for each level of association. This is useful for delineating significant associations at different strength levels. However, for simplicity in our further analyses, we used signatures consisting of all consensus regions derived from $t \geq 3$.

Our recent article[8] validated consensus regions derived independently in two disjoint discovery cohorts, consistent with recent recommendations for brain–behavior associations.[9]

In the current study, we followed the same discovery techniques to generate two new signature ROIs from the ADNI 3 cohort related to neuropsychological and ECog measures of executive function. With the signatures previously computed for neuropsychological and ECog memory,[8] this brought our total to four domain-specific signatures: memory and executive function in both neuropsychological and everyday ECog measures. We did not do a complete validation for the new signature ROIs, relying on the previous validation results to support their use as robust variables. However, we still performed a partial validation by comparing the performance of all four cognitive-specific signature variables generated by the discovery in ADNI 3 in both our current UCD and the ADNI 3 samples.

Finally, we formed the union of all four regions (i.e., the set of voxels belonging to at least one region). We created separate unions for

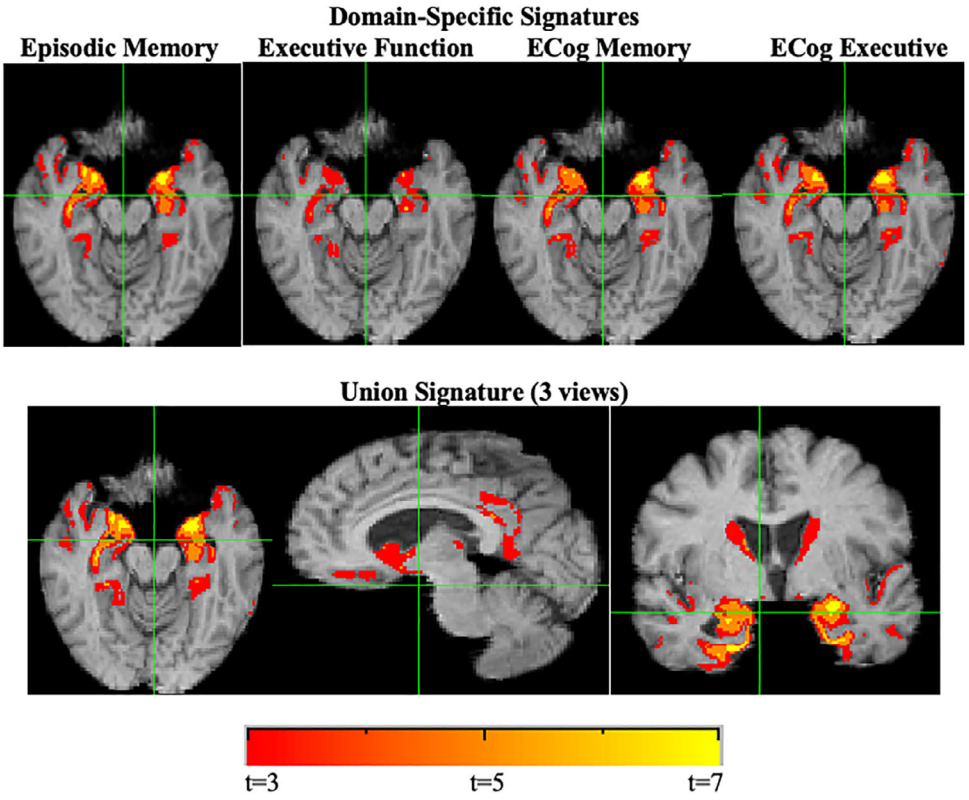

FIGURE 1 Top row: Four domain-specific signatures computed from discovery sets in ADNI 3. Left to right: episodic memory, executive, ECog memory, ECog executive. Bottom row: The Union Signature (three coordinate views). Color codings (red, orange, and yellow) indicate differential strengths of associations (i.e., minimum *t*-value threshold of regression β coefficient for contiguous clusters of voxels) with each outcome, but each color-coded region is statistically significant ($p < 0.001$). The color-coded regions of the Union Signature are the respective spatial unions across the four corresponding regions in the cognitive-specific signatures. These colored regions are useful in delineating relative strengths of associations of brain locations to outcome. Levels are regions of voxel association strength represented by regression coefficient *t*-values (red = 3; orange = 5; yellow = 7).

each *t*-level of association (i.e., the union of four domain-specific $t = 3$ regions, and so on). The "Union Signature" consists of these three *t*-levels. It was motivated by two observations. First, the spatial extents of the four individual signature regions appeared congruent (Figure 1, top row), suggesting strongly overlapping brain substrates for separate behavioral domains (thus replicating the substrate similarity we earlier found for two domains[8]). Second, we tested the performance of each signature for predicting all four behavioral domains, finding that each domain was explained well by not only its "own" signature region but also by all the other signatures. This suggested that a Union Signature might be a powerful single-purpose brain substrate. Furthermore, because of random training set variability in outcome and imaging data and the risk of overlearning, individual signatures might be less generalizable than desired, but a union of all four signatures, by combining the random variabilities of each and thus dampening them, might be less sensitive to training set noise than any individual signature.

We then compared the Union Signature performance in UCD and ADNI 3 with those of the domain-specific signatures. From the foregoing considerations, we hypothesized that the Union signature would perform as well or better than the domain-specific signatures in explaining all four behavioral outcomes.

In the figures, color-coded spatial extents depict the union of corresponding levels from domain-specific signatures. However, in the rest of this article, the Union Signature variable is the mean GM thickness across the combined color-coded regions without differentiating by sub-regions of level of association.

2.5 | Statistical analyses testing the performance of the Union Signature

We tested the ADNI 3–derived Union Signature in models explaining episodic memory, executive function, CDR-SB[27] (continuous measures), and clinical diagnosis in our UCD validation sample. It is important to note that all model performance evaluations occurred in the UCD cohort, which was separate from the ADNI discovery cohort.

We performed mixed-effect modeling of sequential measures for episodic memory, executive function, and CDR-SB in each participant. The time variable was the time difference of each evaluation from the single MRI scan of a participant ("TimeSinceScan"). Models examined the effects of brain predictor variables (i.e., Union Signature and other comparison brain variables) on the intercept and slope of each

outcome. Effects on longitudinal outcome change (slope) were modeled by the interaction TimeSinceScan × brain measure. All brain measures consisted of mean GM voxel-based thickness measures over specific regions. Each brain measure is identified by the names of its region (e.g., Union Signature, Hippocampus, and so on).

Control variables in all models included age at first evaluation (centered at 75), education (centered at 12 years), categorical variables for gender and race/ethnicity (with four categories of Asian, African American, Latinx/Hispanic, and White), and whether cognitive exams were in Spanish or English.

To account for practice effects contributing to scores for episodic memory and executive function, we performed preliminary analyses without brain variables, using indicator variables P_1 coding for one test and P_2 for two or more tests. Thus, at the earliest exam, $P_1 = P_2 = 0$; at the second exam, $P_1 = 1$ but $P_2 = 0$. At subsequent exams, both variables are positive. We found that P_1 contributed to estimates of memory slope, and P_1 and P_2 both contributed to estimates of executive slope (analyses not shown). We, therefore, incorporated P_1 in the full models of episodic memory and P_1 and P_2 in models of executive outcome. Practice effect variables were included as independent predictors and in interactions with brain measures and time × brain measures. Models of CDR-SB did not require controlling for practice effects. Because some participants in LA90, the oldest-old cohort in our UCD sample, were administered shortened forms of the cognitive exams, we also included a control indicator variable for exam formats.

Models were evaluated by the function lmer in the R lme4 package.[28] Model fit performance for continuous outcomes was measured by marginal R^2 (i.e., R^2 fit of the fixed effects; computed by the performance package in R[29]) and effect sizes for brain variables on intercept and slope. To enable direct comparison of effect sizes for different brain measure predictions of an outcome, all brain measures were scaled to Z-scores (mean = 0, SD = 1).

For clinical diagnosis outcome, we performed binary logistic modeling using the glm function (R stats package)[30] for baseline diagnosis (CN, MCI, or dementia). Binary categories were CN versus dementia, CN versus the rest, and dementia versus the rest. Predictor variables included age, gender, education, and race/ethnicity, as in the mixed-effects models above, but no variables pertaining to test scores or administration and no time variable. Model performance was measured by the area under the curve (AUC) of receiver-operating characteristic (ROC) separating binary classifications.

2.6 | Comparing Union Signature with other model performances

Using the UCD cohort as a test set, we compared model fit performances of the Union Signature with brain GM variables known to be associated with our cognitive outcomes. For each variable except hippocampal volume, we sampled GM thickness as described earlier for our signature ROIs. These variables were total cortical GM thickness, temporal lobe GM thickness, hippocampal volume, and entorhinal cortex GM thickness. In addition, in recent years, other research efforts have generated data-driven ROIs associated with the outcomes, including clinical classification and associations with continuous cognitive measures. We selected three for comparison, based on the availability of explicitly described ROIs from which the signatures could be constructed, and on relevance to continuous cognitive outcome or clinical diagnosis. First was Dickerson, a signature of cortical GM thinning that distinguished CN from AD[4] and stable MCI from converters to AD.[2] This set of GM regions was also subsequently shown to be a marker for episodic memory.[31] Second was Epelbaum, a data-driven selection of brain GM regions from the Desikan-Killiany-Tourville (DKT) atlas[32] which together explained episodic memory.[1] And third was Schwarz, an AD signature region made from a composite of entorhinal, fusiform, parahippocampal, middle-temporal, inferior-temporal, and angular gyri.[33] All brain models besides our Union Signature were formed from appropriate GM ROIs of the Mindboggle atlas (https://mindboggle.info/), a second-generation version of the DKT atlas of GM parcellation.[32,34]

Metrics of model performance are reported as means over 10,000 iterations for each model, using the boot package in R.[35] Significance was computed by 95% confidence intervals (CIs) derived from these iterations.

TABLE 1 Participant characteristics for discovery data set (ADNI 3) and test data set (UCD) comprising UCD longitudinal cohort, KHANDLE, STAR, and LA90.

	ADNI 3 discovery set	UCD test set
N	815	1874
Age (mean, SD)	71.4 (7.3)	75.8 (9.1)
Male (%)	48	40
Education, years	16.5 (2.5)	13.9 (4.1)
Race/ethnicity (%)	Hispanic/Latino 4 Non-Hispanic 95 NA 1	Asian 9.9 African American 33.1 Latinx/Hispanic 21.6 White 35.4
Clinical diagnosis (percent)	CN 50 MCI 27 AD 10 NA 13	CN 50.4 MCI 22.3 Dementia 7.5 NA 19
Scanner field strength (% 3T)	99.9	47.6

Note: The two cohorts differed by age, education, ethnoracial diversity, and mix of scanner strengths.
Abbreviations: AD, Alzheimer's disease; CN, normal control; dementia, all cause dementia; MCI, mild cognitive impairment; NA, not available.

3 | RESULTS

3.1 | Participant characteristics

Participant characteristics for the discovery and test data sets are shown in Table 1. The ADNI 3 cohort was younger and showed higher education than UCD, whereas UCD had greater race/ethnic diversity.

Distributions of clinical diagnosis were similar, but it is notable that the dementia category in the UCD includes all-cause dementia, whereas in ADNI 3, the dementia diagnosis is based on AD. The UCD cohort was highly diverse racially/ethnically, in contrast to the homogeneous nature of the ADNI 3.

3.2 | The Union Signature

Figure 1 displays four individual domain signatures and the Union Signature. Figure 1 (top row) shows strong spatial resemblances across all individual signatures. We then formed a Union Signature consisting of the spatial union of all four regions, shown in Figure 1 (bottom row).

3.3 | Verification of Union Signature versus cognitive-specific signature model fits

Next, we compared the Union Signature model fit for each outcome with the four domain-specific signature fits in each of our ADNI 3 and UCD cohorts. Thus, we applied the signature regions computed from the ADNI 3 discovery cohort in that set itself and in our UCD data set. Table 2 shows the mean bootstrapped adjusted R^2 model fits for each model.

The table indicates that the Union Signature R^2 fit value to each outcome was higher than any of the individual signature domain fits with two exceptions (neuropsychological memory outcome, which was best predicted by the Everyday Memory signature in the UCD cohorts, and executive in ADNI, which was best predicted by the executive signature). In many cases, the Union fit was significantly higher than the domain-specific signature fits; it was significantly less in only one instance, for the executive domain signature and executive domain in ADNI.

The results displayed in Figure 1 and Table 2 suggest, first, that there is remarkable consistency in spatial extent and configuration for all domain-specific signatures, although the executive shows a lesser extent than the others. Second, the performances of four domain-specific signatures in predicting each outcome are quite consistent across outcomes in each cohort. Compared to these, the Union Signature is the best or near-best predictor of each behavioral outcome (bottom rows) regardless of cohort.

3.4 | Brain region characteristics of the Union Signature

To ascertain the spatial distribution of the Union Signature relative to standard cortical parcellations, we tabulated the percent overlaps of the Union Signature region with GM regions in the Mindboggle atlas. Table 3 shows overlaps exceeding 10% of each region in descending order. The table indicates heavy overlap of the Union Signature with

TABLE 2 Comparison of outcome model fits (adjusted R^2 coefficient of determination) by domain-specific signature predictors and the Union Signature.

A. Comparisons in ADNI 3				
Signature	**Mem**	**Exec**	**ECogMem**	**ECogExec**
Mem	0.247	0.213	0.205†	0.206†
Exec	0.223†	0.235*,†	0.196†	0.201†
ECogMem	0.210†	0.192†	0.211†	0.210†
ECogExec	0.201†	0.188†	0.199†	0.208†
Union	0.251*	0.215	0.235*	0.235*
B. Comparisons in UCD				
Signature	**Mem**	**Exec**	**ECogMem**	**ECogExec**
Mem	0.306	0.240	0.168†	0.166†
Exec	0.255†	0.220†	0.102†	0.121†
ECogMem	0.315*	0.246	0.175*	0.173
ECogExec	0.313	0.248	0.171†	0.171†
Union	0.311	0.249*	0.175*	0.175*

Note: Models were simple regressions of the form outcome ~ signature variable. We employed bootstrapping to evaluate mean model fit and mean pairwise difference between Signature Union and other model fit, for each cognitive outcome. Confidence intervals derived from each bootstrap gave estimates of significance. Bootstraps consisted of 10,000 iterations for each individual signature model fit and for pairwise differences of the union versus each of the other signature model fits. Highest value in each column indicated by asterisk. Dagger notation indicates significant difference (95% confidence interval [CI]) with the Union Signature fit value in a column. For ECogMem outcome in UCD, the Union and ECogMem models were tied to three decimal places. In both test sets, the Union Signature has the strongest model fit to outcome in three of four domains: Episodic memory, ECog Memory, and ECog Executive in ADNI 3, and Executive, ECog Memory, and ECog Executive in UCD.

Abbreviations: ECogExec, everyday functional executive; ECogMem, every day functional memory; Exec, neuropsychological executive function; Mem, neuropsychological episodic memory.

TABLE 3 Fractions of atlas ROI parcellations with greater than 10% overlap by the Union Signature region.

Region overlapped	Fraction of region in overlap with Union Signature
Amygdala	0.963
Entorhinal	0.772
Hippocampus	0.657
Caudate	0.544
Parahippocampal gyrus	0.384
Isthmus cingulate	0.281
Superior temporal gyrus	0.215
Fusiform gyrus	0.163
Medial orbitofrontal	0.144
Inferior temporal gyrus	0.112

All regions are bilateral.

medial temporal structures of the amygdala, entorhinal cortex, and hippocampus, followed by the caudate nucleus, all with overlaps of over 50% of these structures.

3.5 | Performance comparisons for Union Signature associations to relevant clinical outcomes

Next we evaluated the Union Signature performance, which predicted a variety of clinically relevant outcomes in the UCD data sample. The cognitive models were repeated measures of continuous outcomes for episodic memory, executive function, and CDR-SB. We report the effects of a brain measure on outcome intercept and of time × brain measure on outcome slope. Models were evaluated by these two effect sizes and by marginal R^2. For clinical outcome status, we used single-time models of binary categories. Thus, logistic regression of clinical status outcomes included binary variables for CN versus Dementia, CN versus the rest (i.e., MCI and Dementia), and Dementia versus the rest (i.e., CN and MCI). In logistic models, we measured performance by ROC with their areas under the curve (AUC).

3.5.1 | Models of continuous outcomes

Longitudinal models compared the Union Signature model performance against those of other brain measures that are commonly accepted as strongly associated with our outcomes of interest. All brain variables were converted to Z-scores (mean = 0, SD = 1) to enable comparison of effect sizes across models.

We evaluated model fit metrics and their significance in each model using bootstrapping with 10,000 iterations. Results for brain measure models of each outcome are shown in Table S1 and Figure 2. All marginal R^2 model fits to outcome were significant ($p < 0.05$) based on bootstrapped CIs. Model fits are further depicted in Figure 2, including error bars for 95% CIs derived from the bootstrapped iterations. All brain variable effects for outcome intercept are significant ($p < 0.05$ based on CIs).

Across all models, marginal R^2 and effect sizes on intercept were larger in magnitude for the Union Signature than any other brain variable (top row, Table S1) (effects on CDR are negative because higher CDR indicates a worse score). From Figure 2, Union Signature CIs for marginal R^2 model fit and intercept effect size were non-overlapping with any other model, indicating that these fit values were significantly higher than those in all other models at the level of $p < 0.05$ (top and middle panels). The same holds true for slope effect size in CDR-SB outcome models (Figure 2, bottom) but not in other outcomes. The Union Signature is thus significantly better than all other models for model fit and estimate size of outcome intercept, and better than other slope estimates for CDR-SB. Slope estimates for executive and episodic memory are much smaller, probably because of controlling for practice effects.

3.5.2 | Categorical models of clinical status

We compared the binary classification power of the Union Signature with those of other models based on comparison brain measures: CN versus Dementia, CN versus non-CN (CN vs Rest), and Dementia versus Rest.

Figure 3 shows higher AUCs for Union Signature in all classifications compared to the other brain measures. We further tested the significance of differences for the Union Signature versus the top three other performing measures–Dickerson Signature, hippocampus, and entorhinal GM—in 10,000 pairwise bootstrap iterations. From bootstrapped CIs, the AUC for Union Signature was higher than all other measures in each of the three binary classifications at the 99% level ($p < 0.01$) except for entorhinal GM in Dementia versus Rest, where the difference was significant at the 95% level ($p < 0.05$).

To summarize our results, for multilevel model fits of three continuous outcomes and binary classification of clinical status categories, the Union Signature outperformed all other GM brain markers, including commonly used measures that are strongly associated with memory and cognitive status (hippocampus, entorhinal cortex) and three other signature ROIs computed by data-driven approaches tailored to these outcomes.

4 | DISCUSSION

The Union Signature was constructed as the spatial union of the GM substrates for four domain-specific signatures. Surprisingly, these four spatial extents were highly congruent (Figure 1, top row) and, furthermore, each was highly associated not only with its own outcome, but all domains of neuropsychological and informant-rated memory and executive function (Table 2). These findings supported the idea of combining them via the spatial union of all their substrates. We then tested the performance of the Union Signature. Although generated in an ethnoracially homogeneous ADNI 3 cohort, it performed better than other brain measures when tested in the highly diverse UCD cohort, indicating robustness beyond the discovery cohort, and corroborating the validity of a generalized signature substrate for multiple outcomes. These results suggest that among many comparison brain GM measures, the Union Signature is a powerful, multi-purpose correlate of clinically relevant cross-sectional and longitudinal outcomes and a strong classifier of clinical syndromes.

4.1 | Implications

This has theoretical and practical implications. First, it reinforces the notion that computationally derived brain ROIs can perform better than theory-based brain structures in models of relevant outcomes.[7,36] The Union Signature outperformed several other brain measures that are known to have strong associations with cognition and cognitive change, including medial temporal structures and three other signature

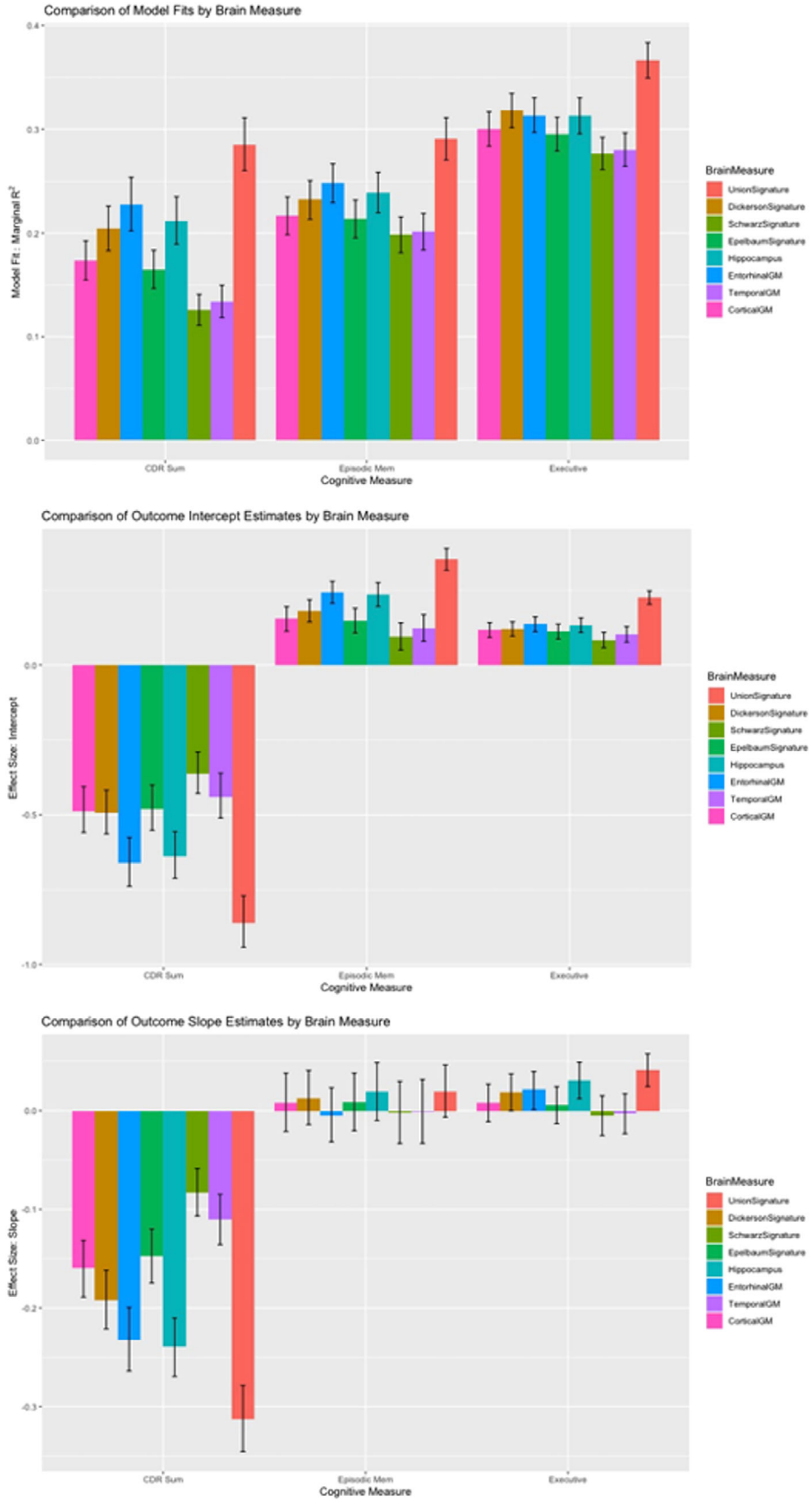

FIGURE 2 Graphical depiction of model fit metrics, including 95% confidence interval (CI) error bars for metrics in models of all outcomes. Top panel: marginal R^2 model fits. Middle: effect estimates for outcome intercept. Bottom: effect estimates for outcome slopes. All values are means over 10,000 bootstrap iterations for each model, from which each 95% CI was also computed. Values (height of bars) are those in Table S1. Error bars not crossing y = 0 indicate significance at $p < 0.05$. Non-overlapping error bars of different models indicate significant differences between them. Union Signature error bars are non-overlapping with all others except for executive and memory slope estimates.

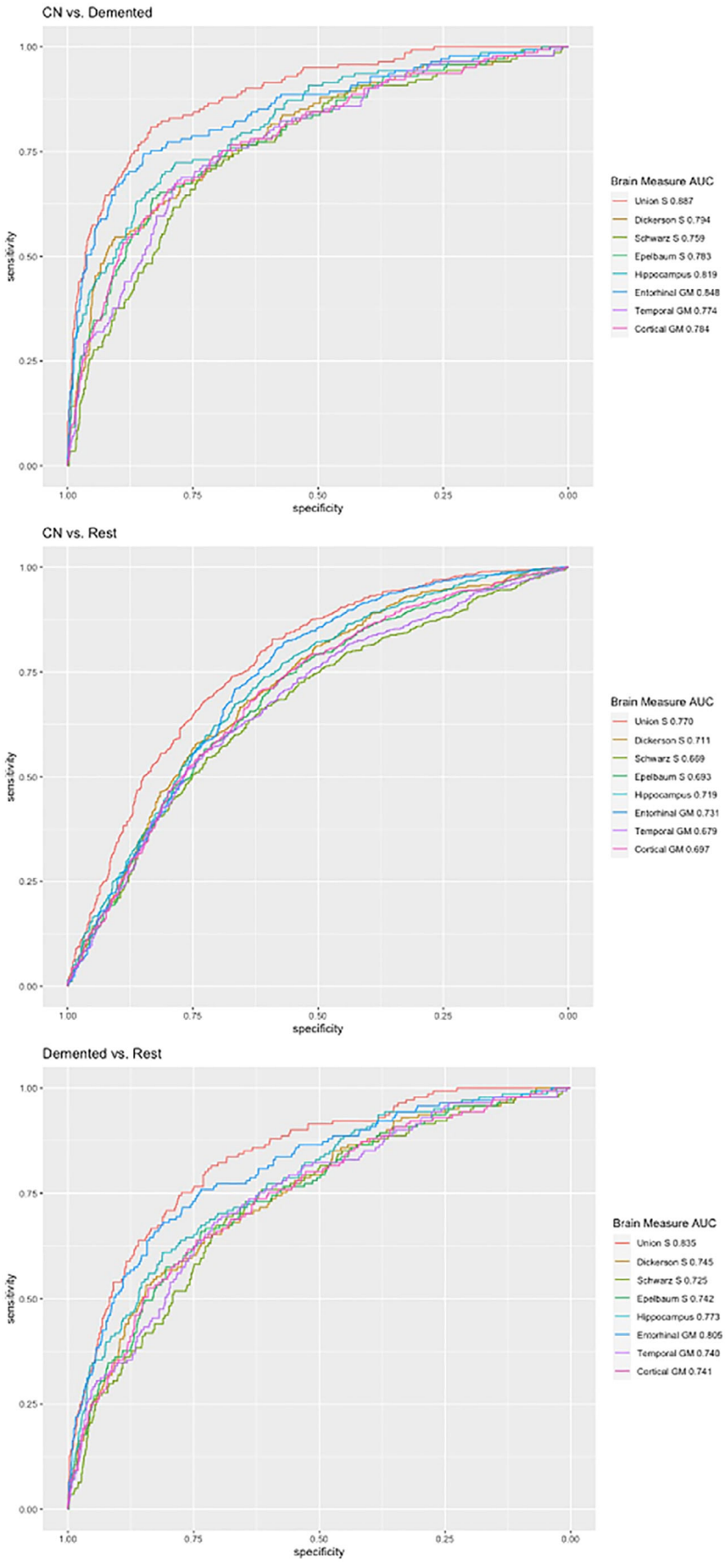

FIGURE 3 Top panel: Plots of normal vs dementia ($N = 1091$) classification by eight brain measures in the UCD test set. AUCs are shown in the legend. Middle panel: Plots of classification for CN vs the rest (top). Bottom panel: Dementia vs the rest (bottom) ($N = 1511$ for both).

variables developed previously to distinguish clinical stages of cognitive impairment or find GM associations with episodic memory. This lends support to wider use of brain markers computationally adapted to outcomes of interest. Second, it supports the idea that separate domains of cognitive performance relevant to successful aging and cognitive impairment share an overlapped set of GM substrates. The Union Signature could, therefore, be a useful multi-purpose biomarker for modeling an array of cognitive trajectories and diagnostic outcomes, potentially simplifying the models by standing in for multiple brain measures that are frequently used.

4.2 | Potential applications of the Union Signature

The notion of a shared brain substrate for relevant outcomes suggests studies of ways in which life exposures specifically impact this substrate, how these impacts may differ by race/ethnicity, and how life exposures may contribute, therefore, to differences in outcome trajectories via their impacts on the Union Signature. These could lead to tailored recommendations of interventions for supporting brain health and consequently improved cognitive aging.

The Union Signature may also be relevant to studies of brain reserve (BR), cognitive reserve (CR), and brain maintenance (BM). BR is commonly characterized as "brain capital" or the amount of physical brain resources,[37] meaning that with a larger amount of relevant substrate, the brain can absorb more loss due to aging and pathology before CI is manifest. CR is characterized as the ability to recruit resources to compensate for losses already incurred, thereby maintaining cognitive performance in the face of pathology.[38] And BM is the concept that rates of cognitive decline may be minimized by "maintaining the brain" in a youthful-like state.[39,40] Each of these definitions entails ambiguity in the form of which brain measures should be taken to characterize BR or to maintain the brain in a relatively youthful state (BM). CR is often measured as the residual of cognitive performance after accounting for cognitive variance explained by brain measures,[41,42] but depending on the brain measures used, this residual might contain variance due to unaccounted-for brain effects.[43] It could, therefore, be made tighter and more precise by using a brain variable that accounts for more of the brain contribution to the outcome than accepted measures like hippocampal volume, WM hyperintensities, or total GM volume. Similarly, BM could be brought into sharper focus by using a region that has been computed to maximize its association with changes in cognitive measures.

4.3 | Relations to deep learning signature approaches

Recently there have been notable data-driven approaches exploring brain associations with cognition,[1] aging,[44,45] and clinical diagnosis.[46] A PubMed search shows that the vast majority of recent efforts have been aimed at developing models for diagnostic classification or prediction of conversion.[46,47] Deep learning (DL) techniques are used in most of these efforts, and these can achieve excellent classification results.[48–50]

DL techniques implicitly use more imaging features in more complex models than explicit algorithmic approaches such as those used in our project. DL thus has potentially greater classification power. For example, in a recent study using a convolutional neural net designed to account for contextual information, the classification of CN versus AD attained an ROC AUC of 0.926 based on training with 416 ADNI participants. Results of nine other DL approaches, for comparison, were in the mid to high 0.8–0.9 range.[51] Our results for CN versus AD (Figure 3, top panel) would therefore rank competitively, but not as the highest, among those approaches. Another example[48] compares a DL approach with an ROI-based feature approach using ventricles and GM parcellations. The DL model produced ROC AUC classifications for CN versus other categories of ≈0.87, depending on the validation method. These are higher than any of our GM measure clinical syndrome classifications for CN versus the rest, including the Union Signature (Figure 3, middle panel). These examples demonstrate the potential power of DL models.

On the other hand, substantial issues remain for machine learning approaches. First, DL models entail difficulties of interpretation. There is currently no standard approach for relating model features to human-comprehensible brain characteristics,[47] although this issue is being addressed (see, e.g., Ref.[49]). Models are most useful when they generate not only predictions, but insight based on interpretability. Second, current machine learning approaches to brain signatures of continuous metrics for cognition, as opposed to diagnosis, are rare. The approach that we describe here could, therefore, fill a current need, combining straightforward application and strong predictive power with interpretability, thereby leading to insights into common brain GM substrates for a range of cognitive performance and clinical diagnoses.

4.4 | Strengths, limitations, and future work

In this study, we demonstrated a data-driven, generalizable brain signature based on structural MRI GM, and showed that it outperforms other accepted brain measures, including several other signature approaches. Further gains might accrue from incorporating additional imaging modalities, including measures of WM integrity using diffusion tensor imaging (DTI) and positron emission tomography (PET) images of amyloid and tau accumulation. Preliminary work in our laboratory suggests a ≈5% improvement in model fit by adding DTI-derived fractional anisotropy imaging (data not shown). However, two caveats apply. With current technology, both these modalities lack the resolution of structural MRI, making them less robust for computationally derived signature methods. PET data are currently less available than structural MRI in the number of images needed to compute robust and generalizable signatures.

DL combined with multiple imaging modalities may be a promising approach to generating signatures of increased power. Yet there is necessarily a limit for any signature performance. A common assumption is that no model solely based on brain measures can completely explain behavioral outcomes (this is the basis for the concept of

CR[38]). Increasingly powerful models must, therefore, approach an explanatory upper limit of less than 100%, producing incremental and diminishing improvements the closer they get to this limit. Probing this limit remains important in future work. In the meantime, our method offers a brain measure that is easy to understand and compute, and the superior power of which makes it useful for modeling a variety of clinical outcomes. Our brain GM signature masks, in standard image template spaces, will be available by request.

In conclusion, results suggest that the Union Signature ROI has several important properties. First, it is informative, indicating shared GM regional substrates for a range of cognitive and diagnostic outcomes that also accord with current theory; second, it is performative, achieving better fits to cognitive and diagnostic outcomes than theory-based brain regions or other signature masks recently developed; third, it is robust across ethnoracially diverse data cohorts; and fourth, it is practical, being easily applied in new cohorts to set up modeling in a straightforward way.

ACKNOWLEDGMENTS

The authors acknowledge the use of data from Alzheimer's Disease Neuroimaging Initiative (ADNI) and from several cohorts, including the University of California Davis Alzheimer's Disease Research Center (ADRC), the Longitudinal Cohort and the Kaiser Healthy Aging and Diverse Life Experiences Study (KHANDLE), the Study of Healthy Aging in African Americans (STAR), and Life After 90 (LA90) cohorts. This work was supported by grants from the National Institute on Aging (NIA; P30 AG072972; R01 AG031563). Data collection and sharing for this project were funded by the ADNI (National Institutes of Health [NIH] Grant U01 AG024904) and Department of Defense ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the NIA and 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 NIH (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 funding sources had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

CONSENT STATEMENT

All data were pre-acquired in existing studies, and no consent statements are required.

ORCID

Evan Fletcher https://orcid.org/0000-0002-9761-3131

REFERENCES

1. Epelbaum S, Bouteloup V, Mangin JF, et al. Neural correlates of episodic memory in the Memento cohort. *Alzheimer's Dement Transl Res Clin Interv*. 2018;4:224-233.
2. Bakkour A, Morris JC, Dickerson BC. The cortical signature of prodromal AD: regional thinning predicts mild AD dementia. *Neurology*. 2009;72:1048-1055.
3. Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol Aging*. 2011;32:2322.e19-27.
4. Dickerson BC, Bakkour A, Salat DH, et al. The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cereb Cortex*. 2009;19:497-510.
5. Marek S, Tervo-Clemmens B, Calabro FJ, et al. Reproducible brain-wide association studies require thousands of individuals. *Nature*. 2022;603(7902):654-660.
6. Masouleh SK, Eickhoff SB, Hoffstaedter F, Genon S. Empirical examination of the replicability of associations between brain structure and psychological variables. *Elife*. 2019;8:1-25.
7. Fletcher E, Gavett B, Crane P, et al. A robust brain signature regions approach for episodic memory performance in older adults. *Brain*. 2021;144:1089-1102.
8. Fletcher E, Farias S, DeCarli C, et al. Toward a statistical validation of brain signatures as robust measures of behavioral substrates. *Hum Brain Mapp*. 2023;44:3094-3111.
9. Rosenberg MD, Finn ES. How to establish robust brain – behavior relationships without thousands of individuals. *Nat Neurosci*. 2022;25:835-837.
10. Satizabal CL, Beiser AS, Fletcher E, Seshadri S, Decarli C. A novel neuroimaging signature for ADRD risk stratification in the community. *Alzheimers Dement*. 2023:1-13.
11. Farias ST, Reed BR, Cahn-Weiner D, Jagust W, Baynes K, Decarli C. The Measurement of Everyday Cognition (ECog): scale development and psychometric properties. *Neuropsychology*. 2008;22:531-544.
12. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *Int Psychogeriatrics*. 1997;9:173-176.
13. Mungas D, Reed BR, Marshall SC, González HM. Development of psychometrically matched English and Spanish language neuropsychological tests for older persons. *Neuropsychology*. 2000;14:209-223.
14. Mungas D, Reed BR, Crane PK, Haan MN, Gonzalez H, González H. Spanish and English Neuropsychological Assessment Scales (SENAS): further development and psychometric characteristics. *Psychol Assess*. 2004;16:347-359.

15. Crane PK, Carle A, Gibbons LE, et al. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav*. 2012;6:502-516.
16. Gibbons LE, Carle AC, Mackin RS, et al. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging Behav*. 2012;6:517-527.
17. Mungas D, Reed BR, Haan MN, Gonzalez H. Spanish and English Neuropsychological Assessment Scales: relationship to demographics, language, cognition, and independent function. *Neuropsychology*. 2005;19:466-475.
18. Mungas D, Widaman KF, Reed BR, Tomaszewski Farias S. Measurement invariance of neuropsychological tests in diverse older persons. *Neuropsychology*. 2011;25:260-269.
19. Early DR, Widaman KF, Harvey D, et al. Demographic predictors of cognitive change in ethnically diverse older persons. *Psychol Aging*. 2013;28:633-645.
20. Fletcher E, Carmichael O, Pasternak O, Maier-Hein KH, DeCarli C. Early brain loss in circuits affected by Alzheimer's disease is predicted by fornix microstructure but may be independent of gray matter. *Front Aging Neurosci*. 2014;6:1-9.
21. Fletcher E, Decarli C, Fan AP, Knaack A. Convolutional neural net learning can achieve production-level brain segmentation in structural magnetic resonance imaging. *Front Neurosci*. 2021;15:1-15.
22. Rueckert D, Aljabar P, Heckemann RA, et al. *Diffeomorphic Registration Using B-Splines*. In MICCAI 2006 Springer-Verlag; 2006:702-709.
23. Fletcher E, Singh B, Harvey D, Carmichael O, DeCarli C, (2012) Adaptive image segmentation for robust measurement of longitudinal brain change. In 34th Annual Conference of the IEEE EMBs, pp. 5319-5322.
24. Kochunov P, Lancaster JL, Thompson P, et al. Regional spatial normalization: toward and optimal target. *J Comput Assist Tomogr*. 2001;25:805-816.
25. Das SR, Avants BB, Grossman M, Gee JC. Registration based cortical thickness measurement. *Neuroimage*. 2009;45:867-879.
26. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex. *Proc Natl Acad Sci*. 2000;97:11050-11055.
27. Lynch CA, Walsh C, Blanco A, et al. The clinical dementia rating sum of box score in mild dementia. *Dement Geriatr Cogn Disord*. 2006;21:40-43.
28. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models Usinglme4. *J Stat Softw*. 2015:67.
29. Lüdecke D, Ben-Shachar M, Patil I, Waggoner P, Makowski D. Performance: an R package for assessment, comparison and testing of statistical models. *J Open Source Softw*. 2021;6:3139.
30. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2013.
31. Busovaca E, Zimmerman ME, Meier IB, et al. Is the Alzheimer's disease cortical thickness signature a biological marker for memory? *Brain Imaging Behav*. 2016;10:517-523.
32. Klein A, Tourville J. 101 labeled brain images and a consistent human cortical labeling protocol. *Front Neurosci*. 2012:1-12.
33. Schwarz CG, Gunter JL, Wiste HJ, et al. A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer s disease severity. *NeuroImage Clin*. 2016;11:802-812.
34. Klein A, Ghosh SS, Bao FS, et al. Mindboggling morphometry of human brains. *PLoS Comput Biol*. 2017:1-40.
35. Canty A, Ripley BD. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-28. 2021.
36. Jolly AE, Hampshire A. Scientific Commentaries: a robust brain signature region approach for episodic memory performance in older adults. *Brain*. 2021;144:1038-1040.
37. Stern Y. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol*. 2012;11:1006-1012.
38. Stern Y, Arenaza-Urquijo EM, Bartrés-Faz D, et al. Whitepaper: defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's Dement*. 2020;16:1305-1311.
39. Johansson J, Wåhlin A, Lundquist A, Brandmaier AM, Lindenberger U, Nyberg L. Model of brain maintenance reveals specific change-change association between medial-temporal lobe integrity and episodic memory. *Aging Brain*. 2021:100027.
40. Nyberg L, Lövdén M, Riklund K, Lindenberger U, Bäckman L. Memory aging and brain maintenance. *Trends Cogn Sci*. 2012;16:292-305.
41. Zahodne LB, Manly JJ, Brickman AM, Siedlecki KL, Decarli C, Stern Y. Quantifying cognitive reserve in older adults by decomposing episodic memory variance: replication and extension. *J Int Neuropsychol Soc*. 2013;19:854-862.
42. Reed BR, Mungas D, Farias ST, et al. Measuring cognitive reserve based on the decomposition of episodic memory variance. *Brain*. 2010;133:2196-2209.
43. Mungas D, Fletcher E, Gavett BE, et al. Comparison of education and episodic memory as modifiers of brain atrophy effects on cognitive decline: implications for measuring cognitive reserve. *J Int Neuropsychol Soc*. 2021;27:401-411.
44. Bacas E, Kahhalé I, Raamana PR, Pablo JB, Anand AS, Hanson JL. Probing multiple algorithms to calculate brain age: examining reliability, relations with demographics, and predictive power. *Hum Brain Mapp*. 2023;44(9):3481-3492.
45. Cole JH, Poudel RPK, Tsagkrasoulis D, et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *Neuroimage*. 2017;163:115-124.
46. Rathore S, Habes M, Iftikhar MA, Shacklett A, Davatzikos C. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *Neuroimage*. 2017;155:530-548.
47. Vieira S, Pinaya WHL, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. *Neurosci Biobehav Rev*. 2017;74:58-75.
48. Liu S, Masurkar AV, Rusinek H, et al. Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs. *Sci Rep*. 2022;12:1-12.
49. Qiu S, Joshi PS, Miller MI, et al. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. *Brain*. 2020;143:1920-1933.
50. Suk HI, Lee SW, Shen D. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Med Image Anal*. 2017;37:101-113.
51. Tong Y, Li Z, Huang H, Gao L, Xu M, Hu Z. Research of spatial context convolutional neural networks for early diagnosis of Alzheimer's disease. *J Supercomput*. 2023;80(4):5279-5297.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fletcher E, Gavett B, Farias ST, et al. A data-driven, multi-domain brain gray matter signature as a powerful biomarker associated with several clinical outcomes. *Alzheimer's Dement*. 2024;16:e70026. https://doi.org/10.1002/dad2.70026