Whole-brain functional connectivity predicts groupwise and sex-specific tau PET in preclincal Alzheimer's disease

Hamid Abuwarda[1,2], Anne Trainer[1], Corey Horien[2], Xilin Shen[4], Suyeon Ju[1], R. Todd Constable[2,3,4], Carolyn Fredericks[1,2]

[1]Department of Neurology, School of Medicine, Yale University, Street, New Haven, 100190, CT, US.
[2]Interdepartmental Neuroscience Program, School of Medicine, Yale University, Street, New Haven, 10587, CT, US.
[3]Department of Radiology, School of Medicine, Yale University, Street, New Haven, 10587, CT, US.
[4]Department of Biomedical Engineering, School of Medicine, Yale University, Street, New Haven, 10587, CT, US.

Abstract

Preclinical Alzheimer's disease, characterized by the initial accumulation of amyloid and tau pathologies without symptoms, presents a critical opportunity for early intervention. Yet, the interplay between these pathological markers and the functional connectome during this window remains understudied. We therefore set out to elucidate the relationship between the functional connectome and amyloid and tau, as assessed by PET imaging, in individuals with preclinical AD using connectome-based predictive modeling (CPM). We found that functional connectivity predicts tau PET, outperforming amyloid PET models. These models were predominantly governed by linear relationships between functional connectivity and tau. Tau models demonstrated a stronger correlation to global connectivity than underlying tau PET. Furthermore, we identify sex-based differences in the ability to predict regional tau, without any underlying differences in tau PET or global connectivity. Taken together, these results suggest tau is more closely coupled to functional connectivity than amyloid in preclinical disease, and that multimodal predictive modeling approaches stand to identify unique relationships that any one modality may be insufficient to discern.

Keywords: Alzheimer's disease, tau, amyloid, functional connectivity, PET, predictive modeling

1 Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative process characterized by the buildup of amyloid and tau pathology. Amyloid spreads quickly and diffusely throughout the neocortex, a process which can begin decades prior to symptom onset [1] [2]. In amnestic AD, tau spread parallels clinical symptomatology and occurs in a highly stereotyped manner, following Braak staging[3] . Tau buildup typically begins in the entorhinal cortex, then spreads to the medial and lateral temporal cortices and cingulate gyrus, corresponding with the first overt cognitive changes. The later stages of the disease are subsequently characterized by tau spread to the remaining neocortex, when AD dementia begins [3] [4] [5].

AD is also associated with characteristic changes in functional connectivity. The default mode network (DMN), which subserves short-term memory and self-referential processing, is specifically targeted in amnestic AD [6] [7][8] [9]. Preclinical disease is also characterized by regional connectivity changes in DMN and early Braak stage regions [10] [11][12] However, while correlation based studies that focus exclusively on the DMN are insightful and can advance our knowledge of the disease, they may also overlook important features of disease due to limited scope of analysis and the tendency of regression-based models to overfit the data. Therefore, an unbiased, data-driven approach that captures features of the entire functional connectome is needed to better grasp the complexity of AD, particularly in the preclinical phase of disease. Another limitation has been that functional connectivity changes associated with the earliest preclinical stages of disease are more challenging to establish, in part because of the lack of large datasets and resulting limited power.

Sex differences are also a major characteristic of the disease. Females have a higher risk of developing AD, and a more severe course of the disease once they do [13] [14]. In healthy aging, females are known to demonstrate posterior DMN hyperconnectivity and may rely on posterior DMN connectivity for memory performance, suggesting a possible role for posterior DMN as a candidate for female-specific susceptibility [15] [16]. Yet, whether significant sex differences exist in the functional connectome in individuals with preclinical Alzheimer's disease is unknown.

Importantly, the relationship between the amyloid and tau pathology in preclinical disease and preclinical changes in functional connectivity is poorly understood. Evidence from animal models and emerging evidence from human studies suggest amyloid-beta induces a state of cellular hyperexcitability and ultimately network hyperconnectivity[17] [18] [19] [20]. While amyloid deposition has been noted to overlap significantly with the DMN especially in very early stages, tau demonstrates a closer coupling to functional connectivity and spreads along functional networks [21] [22]. Using a robust and data-driven machine learning technique, we therefore sought to determine the relationship of changes in the functional connectome to both amyloid and tau pathology in preclinical disease. We focus on the preclinical phase of AD, as it

is a prime window for intervention. Therefore, gaining a deeper understanding of how amyloid, tau, and the functional connectome interrelate within this critical preclinical window is highly important.

We therefore asked, does functional connectivity have predictive information for tau and amyloid PET signal, and if so, can we characterize that relationship across regions and between sexes? To address these fundamental questions, we used connectome-based predictive modeling (CPM), a supervised machine learning technique that identifies edges which correlate with outcomes-of-interest [23]. CPM has been commonly used in relating connectivity to cognitive [24][25][26][27] and behavioral/symptomatic measures [28][29][30][16][31]. However, to our knowledge, functional predictive modeling has not been implemented in a cross modal manner to study biologically validated variables, such as contemporaneous PET imaging. This multimodal approach can be particularly insightful - by adding additional complexity into our predictive models, we can identify relationships that any one modality may be unable to identify [32]. We therefore sought to establish these pathologically associated connectivity models to better inform our understanding of the underlying relationship between pathology, assessed through PET, and the functional connectome in preclinical disease. Given female-risk in the disease, we also set out to determine how these models may vary by sex.

We hypothesize that because amyloid is less relevant to functional connectivity changes than tau, it would be minimally predicted by connectivity. Because tau is more intrinsically linked to connectivity, we hypothesize it would be better modeled by functional connectivity (FC). We also hypothesized sex-specific models would demonstrate different predictive powers between males and females for various regions in preclinical disease.

2 Results

Connectome-based predictive modeling. To determine the relationship between whole brain functional connectivity and amyloid and tau pathology, we used connectome-based predictive modeling (CPM). CPM is a data-driven technique for developing connectivity-based models of phenotypes-of-interest. Specifically, CPM identifies edges from the functional connectome significantly associated with outcomes-of-interest, and uses these edges to generate a linear model between connectivity and the outcome [33]. One major advantage of CPM is it allows one to incorporate information from the functional connectome, rather than a specific ROI/network, the latter of which limits the scope of analysis. This is especially important, as recent work has shown tau covariance is associated with whole-brain functional connectivity, invariant to the network assessed [34]. Furthermore, predictive modeling approaches such as CPM minimize in-sample inflation through a cross-validation procedure, thereby improving generalizability [35] [36], a crucial step for studies with potential clinical applications [37].

We applied CPM to the baseline resting-state functional magnetic resonance imaging (rsfMRI) and tau and amyloid PET images of the Anti-Amyloid in Asymptomatic Alzheimer's disease (A4) study, a clinical trial of cognitively unimpaired adults (aged

65-85). Whole brain connectivity matrices were derived by correlation of the mean time series for each pairwise region defined by a functional atlas (Shen-268). Amyloid standard uptake value ratios (SUVRs) were summarized regionally and as a cortical composite, the latter of which is commonly used to assess for amyloid positivity [38]. Tau SUVRs were assessed regionally given tau's spatial/regional relationship to symptoms and atrophy[39]. PET SUVRs were summarized using the Desikan-Killiany atlas (Supplementary Figure 1). Models were trained to predict these regional tau, regional amyloid and cortical amyloid metrics using whole-brain functional connectivity (Fig. 1A). We implemented a 5-fold cross-validation scheme, and reiterated each model 1000 times, where each time participants were randomly reassigned to new folds, ensuring an unbiased estimation for the cohort. Model performances was assessed by Spearman's correlation of predicted and true amyloid/tau pet binding (r_s). Model significance was determined against permuted models, in which PET metrics were randomly shuffled between participants to assess for spurious correlations. All reported p-values are corrected for the false discover rate (FDR) [40]. Models were rerun with different edge-selection thresholds, with minimal difference in predictive power (Supplementary Table 2 and 3). We additionally ran two sets of models, one with only amyloid-elevated participants (n=342) and one with amyloid-elevated and non-elevated participants (n=394). Because the larger cohort predictive accuracies were better, we present findings for the larger cohort here; we report full results for both groups in Supplemental Tables 2 and 3.

CPM minimally predicts local and global amyloid binding. Given the utility of composite cortical amyloid in determining amyloid positivity, we developed models to predict cortical amyloid. To avoid artificially constraining our predictive range by developing models in amyloid-positive individuals, we randomly subsetted 394 individuals (to match the number of individuals in our tau PET models) from the full cohort of amyloid positive and negative participants. Using this subsampled cohort, we found CPM minimally predicted cortical amyloid SUVR and centiloid (SUVR $r_s = 0.09$, centiloid $r_s = 0.12$; Fig. 1b). To determine whether CPM preferentially predicts regional amyloid SUVR over cortical amyloid SUVR, we developed models to predict amyloid binding in the posterior cingulate and precuneus, which are some of the first regions to demonstrate a buildup of amyloid in preclinical disease [41]. CPM continued to demonstrate minimal predictive power for amyloid binding in these regions (PCC $r_s = 0.12$, Precuneus $r_s = 0.13$; Fig. 1b).

CPM better predicts tau than amyloid. For tau, we chose an initial set of 8 regions (entorhinal, parahippocampal, fusiform, amygdala, inferior temporal, middle temporal, precuneus, and posterior cingulate regions) that would likely best sample changes in early AD, encompassing Braak stages I - V. The hippocampus was excluded due to non-specific binding of the tau ligand AV-1451 to the choroid plexus, making hippocampal binding uninterpretable [42]. To broaden our scope, we also developed models in an exploratory set of 12 other regions involved in the later stages of AD, representing Braak IV-VI (Supplementary Table 2). We also generated a model to assess tau binding in a temporal meta-ROI, a composite of six AD regions used commonly in the literature to assess tau in early temporal ROIs [43].

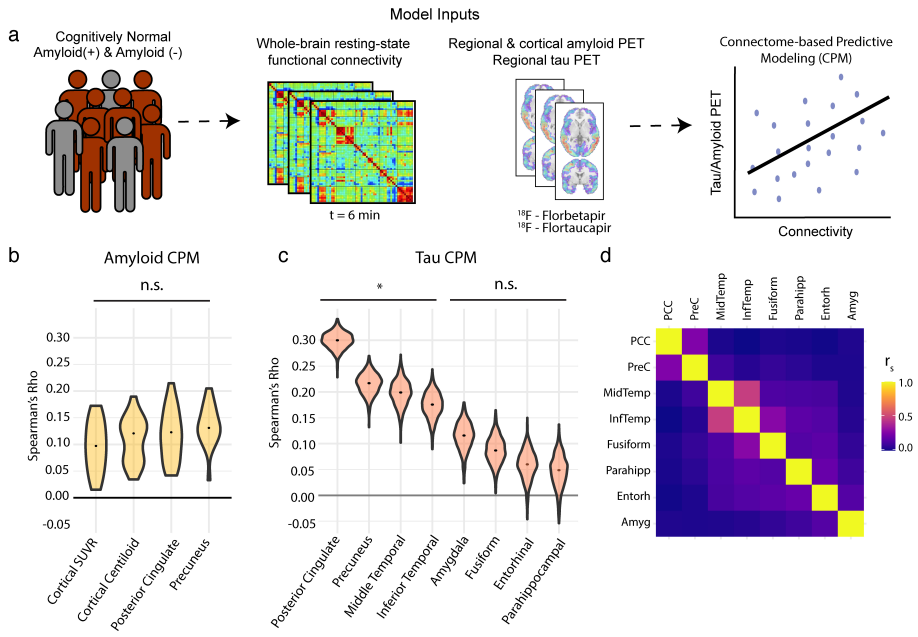

Fig. 1 Functional connectivity predicts regional tau PET in preclinical Alzheimers disease. (A) Study overview: Resting-state functional connectivity and tau and amyloid PET metrics from amyloid(+) (red) and amyloid(-) subjects (gray) were used as inputs into CPM. (B) Violin plots of cortical amyloid (SUVR or centiloids), and regional amyloid CPM in the posterior cingulate and precuneus regions. n=394 subjects. (C) Regional tau CPM violin plots in amyloid positive individuals of early stage regions in AD (Braak I - V). Plots represent distribution over 100 model permutations, with the median model performance depicted by the middle black dot. Model performance measured by Spearman's rho of predicted versus observed regional tau PET. n=394 subjects. (D) Spearman's correlation of predictive edges between models. Regions are ordered from top-to-bottom and left-to-right starting based off its distance from posterior cingulate cortex. PCC = Posterior cingulate cortex, PreC = Precuneus, MidTemp = Middle Temporal, InfTemp = Inferior Temporal, Fusi = Fusiform, Parahipp = Parahippocampal , Entorh = Entorhinal, Amyg = Amygdala. *Significant p<0.05 against permuted models, FDR-adjusted (21 tests). n.s. = not significant against permuted models

CPM predicted tau PET in the posterior cingulate cortex, precuneus and the lateral temporal cortex (median CPM: posterior cingulate $r_s = 0.30$, p=0.021; precuneus $r_s = 0.22$, p=0.021; middle temporal $r_s = 0.20$, p = 0.021; inferior temporal $r_s = 0.18$, p = 0.021, FDR-adjusted for 21 tests) but minimally predicted regions in the medial temporal cortex, amygdala, and fusiform gyrus (Fig 1b; median CPM: amygdala $r_s = 0.12$, p = 0.11; fusiform $r_s = 0.09$, p = 0.14; entorhinal $r_s = 0.06$, p = 0.19; parahippocampal $r_s = 0.05$, p = 0.19; FDR-adjusted for 21 tests). When comparing the correspondence of the predictive edges between these models, we found that these tau models were spatially distinctive (Fig. 1d). That is, predictive edges for each region were unique and demonstrated minimal overlap between one another, with the exception of some neighboring ROIs (e.g. PCC-PreC, Middle temporal - Inferior temporal). In summary, amyloid and early-stage tau models demonstrated minimal predictive power, while later-stage tau models performed moderately well. Because

tau PET models outperformed amyloid models for the same number of subjects, we focus on tau models for the remainder of this study.

Tau-functional connectivity relationship is primarily linear. To determine whether we can improve predictive power by incorporating non-linear dependencies into our models, we used a complementary predictive modeling approach, kernel ridge regression (KRR). KRR is a powerful machine learning technique that optimizes a linear function in high dimensional space, allowing one to incorporate both linear and nonlinear associations into the predictive model. This approach has been shown to outperform linear models with comparable prediction accuracies to deep neural networks when predicting cognitive/behavioral measures, highlighting the complex confluence of linear and nonlinear relationships when assessing these abstract metrics [44].

When applying KRR, we found that tau predictions in these regions were stable and invariant to the modeling choice across all tested regions (Fig. 2a, KRR median model: posterior cingulate $r_s = 0.30$; precuneus $r_s = 0.16$, middle temporal $r_s = 0.19$; inferior temporal $r_s = 0.17$; amygdala $r_s = 0.12$; fusiform $r_s = 0.08$; parahippocampal $r_s = 0.08$; entorhinal r_s =0.05; all p >0.05, CPM<KRR by one-sided Steiger's Z-test, FDR-adjusted for 8 tests). We overlay CPM model results from Fig. 1c for comparison. As a standard of comparison, we compared CPM and KRR in predicting a memory score commonly used in AD cognitive assessments, the Rey Auditory and Verbal Learning Test (RAVLT) in a separate dataset, Human Connectome Project - Aging. We found a significant increase in predictive power in KRR models compared to CPM, aligning with others' work using KRR in predicting cognitive variables (CPM median model: rest 1 $r_s = 0.26$; rest 2 $r_s = 0.27$; rest 3 $r_s = 0.25$; rest 4 $r_s = 0.26$; KRR median model: rest 1 $r_s = 0.36$; rest 2 $r_s = 0.32$; rest 3 $r_s = 0.31$; rest 4 $r_s = 0.34$). This suggests that tau is primarily linearly associated with resting-state functional connectivity and does not benefit from incorporating nonlinear relationships into the predictive model.

Tau model accuracies are associated with global connectivity. We next sought to identify associations related to the differences in CPM performance through a post-hoc analysis. We considered whether underlying tau PET differences may be associated with predictive power across regions. Additionally, as some of our best-performing models (posterior cingulate and precuneus) are major hubs in graph theoretic models of brain connectivity, we also considered whether global connectivity of a region might relate to our ability to predict tau signal there [45]. We therefore calculated a metric of global connectivity, intrinsic connectivity distribution (ICD), for each region and visualized the distributions of regional tau PET and regional global connectivity (Fig. 3a).

Median model accuracies were poorly correlated with median tau PET SUVR for each region ($r_s = 0.08$, $R^2 = 0.006$, p=0.72 by student t-test, FDR-adjusted for 2 tests) but well correlated with global connectivity (Fig. 3b r_s= 0.52, $R^2 = 0.27$, p <0.05 by student t-test, FDR-adjusted for 2 tests). This highlights that CPM performance is more highly associated with the global connectivity of a region than it is with underlying tau PET metrics.

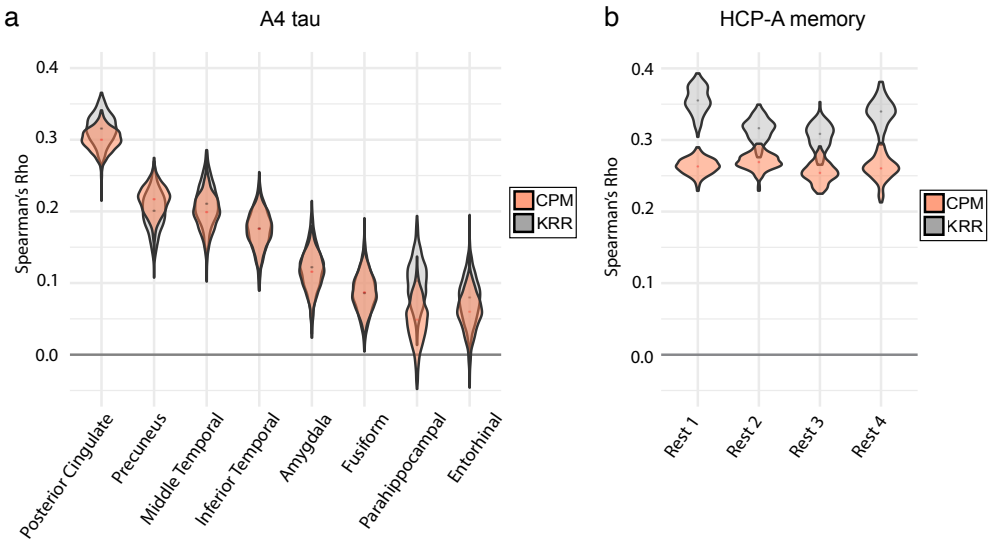

Fig. 2 Tau is predominantly linearly associated with resting-state functional connectivity in preclinical AD. (A) CPM and KRR tau model performances for the 8 regions in A4. Median performance depicted by black (CPM) or red (KRR) dots. n.s. = not significant p<0.05 by Steiger's one-sided z-test (CPM<KRR, FDR-adjusted for 8 tests). (B) CPM and KRR memory model performances across the four resting state scans in HCP-A. Models were developed to predict RAVLT-L.

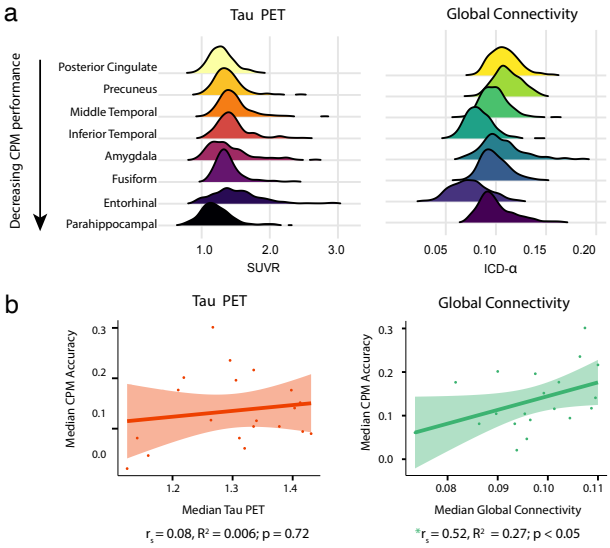

Fig. 3 Higher global connectivity is associated with improved tau model accuracy. (A) Distributions of tau PET and global connectivity (ICD - α) for each region. Regions are ordered in descending order of CPM accuracy. (B) Spearman's correlation (r_s) between median tau PET (left) and median global connectivity (right) against the median CPM accuracy for the expanded set of 20 regions; *Significant (p<0.05), FDR-adjusted (2 tests).

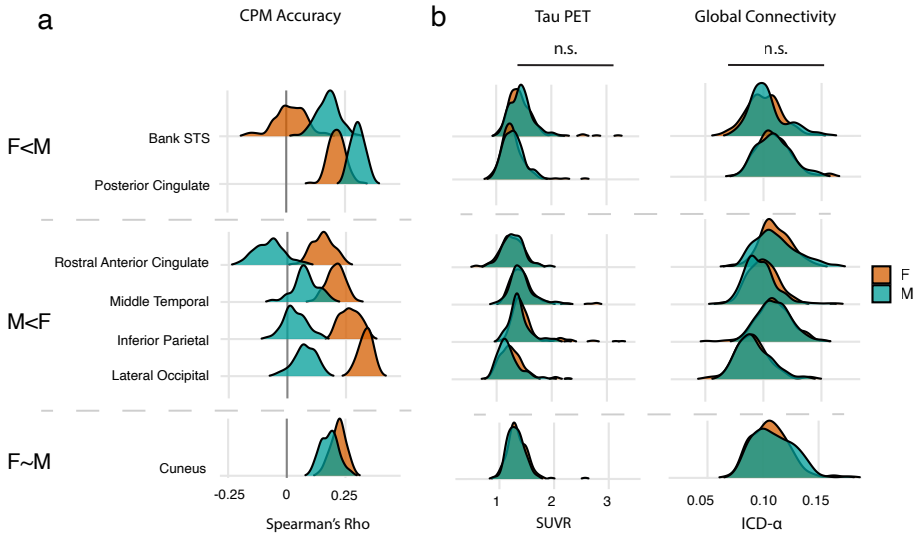

Fig. 4 Sex differences are detectable in preclinical disease (A) CPM accuracies of female and male specific models for each region. Regions shown are those in which males outperformed females, females outperformed males, and models which had about equal performance. (B) Sex-specific distributions of tau PET (SUVR) and global connectivity (ICD-α) for each region. n.s. = not significant by student t-test, FDR-adjusted (7 tests)

Detectable sex differences in preclinical disease. Given the importance of sex in symptomatic AD, we set out to determine whether sex-specific CPM models differentially impact prediction performances in this preclinical asymptomatic cohort. We randomly subsampled females to match males (n=142) to ensure power does not play a role–recent work has suggested sex differences is simply a difference in power[46]. We assessed underlying covariate differences between the two groups that may influence our ability to predict tau, namely age, education, cognitive metrics (MMSE and PACC), and cortical amyloid. We found that they differed significantly with regards to age, education, and preclinical Alzheimer's cognitive composite (PACC), a cognitive metric that has been used to discern early cognitive changes[47]. We therefore include these as covariates into the models to ensure they do not influence model prediction differences.

We generated male and female-specific models and found that males and females performed differently for different regions (Fig 4A). There were some regions where males outperformed females (M¿F, bank STS $\Delta r_s = 0.16$, posterior cingulate $\Delta r_s = 0.09$), some regions where females outperformed males (F¿M rostral anterior cingulate $\Delta r_s = 0.22$, middle temporal $\Delta r_s = 0.14$, inferior parietal $\Delta r_s =$, lateral occipital Δr_s =0.24) and a region where performances were both high (F M, cuneus $\Delta r_s = 0.04$). These differences are not attributable to underlying tau PET and global connectivity differences in each region (p>0.05 by student t-test for all male vs. female tau PET and ICD distributions)(Fig 4b). These results suggest predictive models such as CPM can differentially model tau predictions in males and females in preclinical disease, without underlying differences in tau burden or global connectivity.

3 Tables

	A4 Aβ(+) (n=342)	**HCP-A** (n=566)
Age, yrs (μ ± s.d)	72.2±4.9	57.3±
Females, n(%)	200 (58.5%)	330 (58.3%)
Education, yrs(μ ± s.d.)		
MMSE (μ ± s.d.)	28.6 ±1.3	
Cortical Amyloid, Centiloids (μ ± s.d)	64.0±32.7	–
APOE, n (%)		
$\epsilon2/\epsilon2$	15 (4.3%)	–
$\epsilon2/\epsilon3$	132 (38.5%)	–
$\epsilon3/\epsilon3$	158 (46.2%)	–
$\epsilon3/\epsilon4$	24 (7.0%)	–
$\epsilon4/\epsilon4$	12 (3.5%)	–
$\epsilon2/\epsilon4$	1 (0.003%)	–

Table 1 Dataset Demographics.

	Females (n=142)	Males (n=142)	p-value
Age (μ ± s.d)	71.4± 4.8	73.4±4.9	p<0.001
Education(μ ± s.d.)	15.7 ± 2.7	16.8±2.7	p<0.01
MMSE (μ ± s.d.)	28.8 ± 1.2	28.5 ± 1.4	p=0.12
PACC Composite (μ ± s.d.)	-0.1± 2.5	-1.2 ± 2.6	p<0.001
Cortical Amyloid, Centiloids (μ ± s.d)	62.7±33.7	64.0 ± 32.5	p = 0.44
APOE			
$\epsilon2/\epsilon2$	0 (0%)	0 (0%)	
$\epsilon2/\epsilon3$	6 (4.2%)	8 (5.6%)	
$\epsilon3/\epsilon3$	57 (40.1%)	56 (39.4%)	
$\epsilon3/\epsilon4$	67 (47.2%)	65 (45.8%)	
$\epsilon4/\epsilon4$	10 (7.0%)	6 (4.2%)	
$\epsilon2/\epsilon4$	2 (1.4%)	7 (4.9%)	

Table 2 Sex Demographics for male and female A4 subgroups. Significance determined by Wilcoxon Rank Sum Test, FDR-adjusted for 5 tests.

4 Methods

4.1 Datasets

The Anti-Amyloid in Asymptomatic Alzheimer's disease (A4) study is a clinical trial study of cognitively unimpaired adults (aged 65-85). Specific study inclusion and exclusion criteria is described elsewhere [48] Of those with baseline rsfMRI, 1491 participants met the requisite motion threshold (¡0.3 mm mean frame-to-frame displacement). Of those with available baseline rsfMRI, 394 subjects had tau PET images (n=51 healthy controls, n = 342 amyloid-elevated individuals).

Human Connectome Project-Aging (HCP-A) study consists of 689 healthy subjects aged 36-100. Imaging data were from the 1.0 release of HCP-A, while neuropsychological data was derived form the 2.0 release. We excluded individuals with missing data, anatomical abnormalities (such as meningioma, vascular abnormalities or ventriculomegaly significant enough to distort cortical anatomy, as well as for significant

scanner artifacts), and motion ¿0.3 mm mean frame-to-frame displacement. After exclusion, the remaining sample size was n = 566.

4.2 Image Parameters

A4. Participants underwent scanning using a 3T MRI and included one resting state fMRI scan (t=6.5 mins, (TR/TE = 3000/30 ms, flip angle = 80°), and a T1-weighted anatomical (MPRAGE) scan.

HCP-A. All participants in the HCP-A study underwent scanning using a Siemens 3T Prisma MRI equipped with 80 mT/m gradients and a 32-channel head coil. The data collected included four resting-state fMRI (rsfMRI) scans, three task-fMRI (tfMRI) scans, and a T1-weighted (T1w) anatomical scan (as detailed in Harms et al., 2018). Each rsfMRI scan lasted 6.5 minutes, totaling 26 minutes. The rsfMRI scans were divided into two sessions on the same day, with each session including one rsfMRI scan using an anterior to posterior (AP) phase encoding direction and another using a posterior to anterior (PA) direction. A multi-echo MPRAGE sequence, detailed in Harms et al. (2018), was utilized for the T1w scans. All fMRI scans employed a 2D multiband (MB) gradient-recalled echo (GRE) echo-planar imaging (EPI) sequence (MB8, TR/TE = 800/37 ms, flip angle = 52°).

4.3 MRI Preprocessing

Detailed description of data preprocessing is outlined in previous work [49]. Briefly, MPRAGE scans were skull stripped using optiBET [50] and nonlinearly aligned to the MNI-152 template using BioImage Suite (BIS; [51]). Functional images were slice-time and motion corrected using SPM12, and scans with a maximum mean frame-to-frame displacement (FFD) greater than 0.3 mm were excluded to minimize motion-related artifacts [25] [49]. This resulted in removal of 1 subject in Prevent-AD, and x subjects in A4. Linear, quadratic, and cubic drifts; a 24-parameter motion model [52]; and signals from cerebrospinal fluid, white matter, and global regions were statistically removed from the data as detailed previously [15]. Connectivity matrices were generated using BIS using the Shen-268 atlas. MRIs across datasets were processed similarly.

4.4 PET Preprocessing

A4. Tau PET ([18F]flortaucipir/FTP) SUVR measurements were obtained from A4, using the 90-110 minute window post-injection (4x5-minute frames). Preprocessing and analysis were performed using PETSurfer, an implementation offered in FreeSurfer (v. 6.0+). Each of the five-minute tau PET frames were motion-corrected and then summed. These composite PET images were registered to corresponding MRI images, segmented according to the Desikan-Killiany Atlas (DSK), and corrected for partial-volume effects using FreeSurfer. This pipeline is also available online (https://surfer.nmr.mgh.harvard.edu/fswiki/PetSurfer). For each brain region defined by the DSK atlas, average tracer uptake values were calculated and standardized against the whole cerebellar cortex as the reference regions, generating standardized uptake value ratios (SUVR).

Cortical and regional amyloid PET (florbetapir-PET/FBP) SUVRs were obtained from the A4 online data release (Version: 2021-04-01). SUVR metrics were derived using 50-70 minute (4x5-minute frames) post-injection images. Images were summed and realigned into a single 3D image. SUVRs were computed using the whole cerebellum as the reference region.

4.5 Connectome-Based Predictive Modeling (CPM)

To identify network of edges associated with tau and amyloid buildup, we employed connectome-based predictive modeling (CPM), the specifics of which can be found in Shen *et al.*, 2017[33]. Briefly, connectivity matrices (predictors) and measures-of-interest (outcomes) are compiled for each participant. For each functional connection, the connection strength of all participants is aggregated and then correlated with the measure-of-interest. This process is repeated for every connection in the brain, resulting in a single correlation matrix for the study. We use a partial correlation approach in which scanner and motion are covariates. The connections that demonstrate a significant correlation to the outcome are likely important brain-based contributors to that outcome and are therefore selected for. For each participant, the selected connections from their connectivity matrix are summed to generate a patient-specific brain network summary score. Each participant's summary network score is related to their respective outcome measure, and a linear regression is generated to model the "brain-outcome measure" relationship. To assess model accuracy, the linear regression model is validated internally on a group of novel participants that was not used during the model-building process. This process is repeated, where each time, individuals are randomly re-shuffled into new groups; this generates an unbiased estimation of the study cohort.

Using the rsfMRI (t = 6.5 min) from the 394 subjects who had tau-PET in A4, we sought to predict contemporaneous FTP binding (standardized uptake value ratio (SUVR)) in individual temporal and parietal regions implicated in early-stage AD, and a meta-ROI (comprised of entorhinal, parahippocampal, amygdala, fusiform, middle temporal, and inferior temporal regions) previously described in the literature [43]. We additionally explored other regions for later stages of AD. This resulted in a total of 21 individual tau models. We ran two sets of models (394 subjects with 51 amyloid negative individuals, and 342 subjects of only amyloid-positive individuals). We also generated 3 amyloid models to predict cortical amyloid, precuneus and posterior cingulate SUVRs, which were provided in the A4 dataset. We used a 5-fold cross-validation, and reiterated this process 1000 times, each time randomly shuffling individuals into different folds. We used a feature selection cut-off of $p = 0.05$ to select edges. Variables such as A4 motion and scanner type (encoded as a discrete variable) were treated as covariates within our model to account for in-scan head movement and inter-scan variability.

4.5.1 Model Assessment

Model performances were assessed by using a Spearman's correlation of predicted vs. observed variables.

In addition, we tested the robustness of our models by comparing them to models where the participant labels were randomly shuffled before predicting pathology. This can identify other underlying relationship that may spuriously drive predictions. After conducting this random permutation 1000 times, we calculated how often the accuracy of the randomly permuted predictions surpassed the median accuracy of the regular predictions. This was used to calculate a non-parametric p-value, described in Scheinost *et al.* (2021)[53].

$$P = \frac{\#\{\rho_{\text{null}} \geq \rho_{\text{median}}\}}{1000} \tag{1}$$

We adjusted these non-parametric p-values for multiple comparisons using the Benjamini-Hochberg procedure.

4.6 Kernel Ridge Regression

Kernel ridge regression is a powerful machine learning technique that transforms the input data (connectivity matrix) into a higher-dimensional feature space through a nonlinear function, and then optimizes a linear function within that space. By doing so, this approach can integrate both linear and nonlinear dependencies to make predictions. Specifically, the optimization aims to minimize the cost function:

$$C = \frac{1}{2}\sum ||y_i - w^T\phi(x_i)||^2 + \frac{1}{2}\lambda||w||^2, \tag{2}$$

Here $\phi : x_i \rightarrow \phi(x_i)$ represents the nonlinear transformation that maps the input data into the feature space. For a new input x*, the predicted y* value is calculated using the kernel $K(x_i, x_j) = \phi(x_i)^T\phi(x_j)$ by

$$y* = K(x*, X) * ((K(X, X) + \lambda I)^{-1} * y). \tag{3}$$

The kernel K, which serves as the inner product within the feature space, should be positive-definite; in the present study, we used Gaussian kernels. The hyperparameter for the Gaussian kernel (gamma) was manually selected between 0.001 and 10; the hyperparameter for ridge regression (lambda) was manually selected between 0 and 10.

4.7 Intrinsic Connectivity Distribution

To assess global connectivity, we used intrinsic connectivity distribution (ICD), which has been previously described [54]. This method better captures global connectivity when compared to traditional metrics such as node degree, which requires arbitrarily defining correlation threshold values (cite). ICD is able to capture the full distribution of connections to the seed, thereby eliminating the need for a correlation threshold. Briefly, connections from connectivity matrices are fit to a stretched exponential model

with an unspecified variance parameter, denoted as alpha, which is then optimized:

$$d(\tau, x, \alpha(x), \beta(x)) = e^{\alpha(x)\tau^{\beta(x)}} \tag{4}$$

Where τ = threshold, x = connectivity matrix, α is the unspecified variance parameter, and β is the shape parameter [54]. These alpha values are summarized as a parametric image, with higher values indicating a broader correlation distribution and therefore higher global connectivity.

5 Discussion

In this work, we use a whole-brain, data-driven approach to study how the functional connectome relates to tau and amyloid PET binding. We demonstrate that in asymptomatic AD, functional connectivity is a more robust predictor of tau binding than of amyloid binding. Furthermore, our analyses reveal a predominantly linear association between tau and functional connectivity, supporting the use of linear models in examining their relationship. Notably, the accuracy of tau predictive models correlates with the global connectivity of a region, rather than with tau PET. Our study also sheds light on sex-specific differences in brain regions that show no difference in tau PET binding or global connectivity, highlighting nuanced, sex-dependent aspects of AD not previously appreciated.

We show that CPM preferentially predicts tau PET over amyloid PET, which may suggest a potentially stronger link between tau and functional connectivity. Not only would this converge on existing work that tau is closely associated with disease outcomes such as cortical atrophy and symptomatology [39], but it also complements recent evidence identifying an association between whole-brain tau PET covariance and functional connectivity [34]. Regional and cortical amyloid were minimally predicted, underscoring the limited predictive ability for amyloid through functional connectivity in preclinical disease.

Our findings indicate models were more accurate in regions traditionally associated with later stages of amnestic AD. Interestingly, predictive power for tau burden was not associated with metrics of tau PET (median tau PET). Rather, it correlated with the global connectivity of a region. One interpretation of this finding is that tau affects more globally integrated ROIs, resulting in more disrupted connectivity compared to regions that are less globally connected. Predictive models like CPM then identify these disrupted edges as inputs into the model by having a greater pool of edges significantly associated with the outcome of interest.

This finding also aligns with recent work showing an association between tau progression and global hubness in AD patients [55]. A key distinction from this work is that the authors did not find effects in preclinical AD, but rather only in symptomatic disease. This discrepancy may stem from differences in approach: our analysis focuses on specific ROIs, whereas that study used a whole-brain metric convolving tau and hubness. Our approach may provide a more nuanced analysis, important in studying subtle differences that may arise in preclinical disease.

Our ability to differentially predict tau in different ROIs, despite a lack of notable differences in tau PET, may underscore the potential of multimodal approaches. By

incorporating additional dimensions to our analyses, these approaches may enhance sensitivity to uncover complex, yet subtle, disease characteristics not discernible from unimodal analyses. Furthermore, multimodal work, especially if one modality is well-described and well-validated (as in the case of tau PET), allows for investigations of functional connectivity to be anchored in biologically informed ways. That is, tau and amyloid PET are well-studied, and the pattern of spread has been well-documented in amnestic AD [3] [1]. As such, we can use PET as an indirect way to assess connectivity differences that track with the progression of AD. Such a multimodal approach increases confidence in results, especially in human neuroimaging research, where each technique has inherent strengths and weaknesses with respect to spatiotemporal resolution [32].

We also note that the observed link between tau and connectivity is primarily linear. Importantly, this validates the use of models such as CPM, which maintain a high degree of interpretability (e.g. one can look at edge-level contributions in CPM, but cannot do so in KRR) [56]. While we provide evidence for this linear relationship in preclinical disease, it remains to be seen whether the linear nature between tau and functional connectivity continues throughout the disease course.

Our work here adds to the literature by showing female and male-specific model performance differences across ROIs, without any underlying differences in tau PET or global connectivity. As we note above, our multimodal approach may offer a more sensitive approach for detecting nuanced variations in the early stages of AD. Yet a critical questions emerges: why would sex differences manifest early? One hypothesis is that these regions may be crucial in sex-specific functional networks, indicating we might be identifying vulnerabilities unique to each sex. Further work is required to detail the exact nature of this relationship.

We also recognize that our metric of global connectivity (ICD) may not capture more fine grained connectivity characteristics, such as those with graph theoretical metrics. It will therefore be important to see if other metrics describing FC may better better define these sex differences. While we adjust for known covariates to minimize confounding effects, it is possible that other overlooked factors could influence our results. Thus, these findings should be further explored through quasi-experimental statistical techniques (such as regression discontinuity, difference-in-differences, and instrumental variables) and experimental studies to deepen our understand of how gender/sex differences enhance tau prediction in preclinical AD[57]. Given the impact of sex differences in AD prevalence and progression, it will be important to continue studying the role of sex in brain-phenotype prediction studies in large, well-powered studies.

There are important caveats to the current work. This study used connectivity data from a single six minute resting state fMRI scan. It has been shown that longer scan times allow more reproducible estimates of connectivity[58], and the use of task-based data has also been determined to improve prediction performance [25]. Future studies could examine the impact of using longer, task-based connectivity data to generate predictive models. In addition, it is important to note that the power to detect brain-phenotype relationships may be different with biological variables (such as tau)

versus behavioral variables (such as psychological or cognitive measures); follow-up work could benchmark effect sizes and reproducibility for each.

An additional limitation is the A4 study did not include participants who were at the tails of the distribution with respect to cognitive metrics. It is unclear how this might impact model performance in participants with more extreme cognitive measures. Encouragingly, however, the A4 dataset is comprised of participants from the US, Canada, Australia, and Japan, potentially boosting generalizability .

Finally, AD is a highly heterogenous condition, and it has been shown that models defined across an entire sample might not be relevant for all individuals in the dataset[59]. Although not the focus of the current paper, additional studies could examine how data-driven subtyping could improve brain-phenotype predictions. Integrating a longitudinal design in future studies can also better help tease apart temporal heterogeneity of individual patients. While the present work is dependent on associative relationships, and therefore we cannot define causal relationships here, these models can still be powerful hypothesis generation tools. They should be integrated with studies that can better define causal relationships - i.e. complex causal inference statistical techniques, experimental work in animals, or randomized control trials. We may also explore how these models can be clinically relevant - e.g. are predictive edges from these tau models sufficient to explain cognitive metrics, or can they be a more sensitive metric to assess risk for AD?

In the future, it will be important to more precisely assess the temporal aspects of disease progression. We focus on preclinical AD in this study, but this loose term can refer to a range of up to 10-20 years prior to symptom onset[2]. It will be vital to assess the stability of the predictive models identified here not only in different stages of preclinical disease, but also in symptomatic disease. New techniques such as the subtype and stage inference algorithm (SuStain) may help in this effort, as it can pseudotemporally assign individuals into different stages of the disease[60]. Yet this poses another limitation: we need to increase power that much more to better temporally define this disease. With new large preclinical datasets such as A4, the field can begin to do so. We anticipate studying preclinical disease with new blood-biomarkers with high PPV and NPV will help remove this roadblock to recruit individuals [61].

Our results underscore the importance of a integrating multiple neuroimaging modalities in a predictive modeling framework to understand complex diseases like AD. Sex-differences in preclinical disease remains a vastly unexplored area of research in AD; we highlight the ability to detect these sex differences in preclinical disease using a multimodal approach. Future efforts to continue better understanding preclinical AD will provide insights and possibly allow targeted interventions to prevent or delay progression of the disease.

Declarations

Funding. This work was funded by awards to CF from the National Institutes of Health (5K23AG059919-04), the Alzheimer's Association (2019-AACSF-644153), a Development Award from the Yale Alzheimer's Disease Research Center (P30AG0066508), Women's Health Research at Yale, and the McCance Foundation.

Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under Award Number U01AG052564 and by funds provided by the McDonnell Center for Systems Neuroscience at Washington University in St. Louis. HA and CH were supported by a Medical Scientist Training Program training grant (NIH/NIGMS T32GM007205). HA was supported by National Center for Advancing Translational Science (NCATS) under Grant Number UL1 TR001863.

Conflict of interest. The authors declare no conflicts of interest.

Authors' contributions. HA, CF, TC conceptualized the study, AT preprocessed the A4 data, SJ preprocessed the HCP-A data and generated HCP-A memory CPM models. HA performed analyses with support from AT. HA drafted the manuscript, with contributions from CH, AT, and CF.

References

[1] Thal, D. R., Rub, U., Orantes, M. & Braak, H. Phases of aβ-deposition in the human brain and its relevance for the development of ad. *Neurology* **58**, 1791–1800 (2002).

[2] Li, Y. *et al.* Timing of biomarker changes in sporadic alzheimer's disease in estimated years from symptom onset. *Annals of Neurology* (2024).

[3] Braak, H. & Braak, E. Neuropathological stageing of alzheimer-related changes. *Acta neuropathologica* **82**, 239–259 (1991).

[4] Braak, H., Thal, D. R., Ghebremedhin, E. & Del Tredici, K. Stages of the pathologic process in alzheimer disease: age categories from 1 to 100 years. *Journal of Neuropathology & Experimental Neurology* **70**, 960–969 (2011).

[5] St-Onge, F., Chapleau, M., Breitner, J. C. S., Villeneuve, S. & Pichet Binette, f. t. A. D. N. I., Alexa. Tau accumulation and its spatial progression across the Alzheimer's disease spectrum. *Brain Communications* **6**, fcae031 (2024).

[6] Menon, V. 20 years of the default mode network: A review and synthesis. *Neuron* (2023).

[7] Greicius, M. D., Srivastava, G., Reiss, A. L. & Menon, V. Default-mode network activity distinguishes alzheimer's disease from healthy aging: evidence from functional mri. *Proceedings of the National Academy of Sciences* **101**, 4637–4642 (2004).

[8] Sperling, R. A. *et al.* Amyloid deposition is associated with impaired default network function in older persons without dementia. *Neuron* **63**, 178–188 (2009).

[9] Schultz, A. P. *et al.* Phases of hyperconnectivity and hypoconnectivity in the default mode and salience networks track with amyloid and tau in clinically normal individuals. *Journal of Neuroscience* **37**, 4323–4331 (2017).

[10] Filippini, N. *et al.* Distinct patterns of brain activity in young carriers of the apoe-ε4 allele. *Proceedings of the National Academy of Sciences* **106**, 7209–7214 (2009).

[11] Huijbers, W. *et al.* Amyloid deposition is linked to aberrant entorhinal activity among cognitively normal older adults. *Journal of Neuroscience* **34**, 5200–5210 (2014).

[12] Billette, O. V. *et al.* Novelty-related fmri responses of precuneus and medial

temporal regions in individuals at risk for alzheimer disease. *Neurology* **99**, e775–e788 (2022).

[13] Zhu, D., Montagne, A. & Zhao, Z. Alzheimer's pathogenic mechanisms and underlying sex difference. *Cellular and Molecular Life Sciences* **78**, 4907–4920 (2021).

[14] Filon, J. R. *et al.* Gender differences in alzheimer disease: brain atrophy, histopathology burden, and cognition. *Journal of Neuropathology & Experimental Neurology* **75**, 748–754 (2016).

[15] Ficek-Tani, B. *et al.* Sex differences in default mode network connectivity in healthy aging adults. *Cerebral Cortex* **33**, 6139–6151 (2023).

[16] Ju, Y. *et al.* Connectome-based models can predict early symptom improvement in major depressive disorder. *Journal of affective disorders* **273**, 442–452 (2020).

[17] Kazim, S. F. *et al.* Early-onset network hyperexcitability in presymptomatic alzheimer's disease transgenic mice is suppressed by passive immunization with anti-human app/aβ antibody and by mglur5 blockade. *Frontiers in aging neuroscience* **9**, 71 (2017).

[18] Ghatak, S. *et al.* Mechanisms of hyperexcitability in alzheimer's disease hipsc-derived neurons and cerebral organoids vs isogenic controls. *Elife* **8**, e50333 (2019).

[19] Busche, M. A. *et al.* Tau impairs neural circuits, dominating amyloid-β effects, in alzheimer models in vivo. *Nature neuroscience* **22**, 57–64 (2019).

[20] Jones, D. T. *et al.* Cascading network failure across the alzheimer's disease spectrum. *Brain* **139**, 547–562 (2016).

[21] Ossenkoppele, R. *et al.* Tau covariance patterns in alzheimer's disease patients match intrinsic connectivity networks in the healthy brain. *NeuroImage: Clinical* **23**, 101848 (2019).

[22] Franzmeier, N. *et al.* Tau deposition patterns are associated with functional connectivity in primary tauopathies. *Nature communications* **13**, 1362 (2022).

[23] Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience* **18**, 1664–1671 (2015).

[24] Rosenberg, M. D. *et al.* A neuromarker of sustained attention from whole-brain functional connectivity. *Nature neuroscience* **19**, 165–171 (2016).

[25] Greene, A. S., Gao, S., Scheinost, D. & Constable, R. T. Task-induced brain state manipulation improves prediction of individual traits. *Nature communications* **9**, 2807 (2018).

[26] Horien, C. *et al.* A generalizable connectome-based marker of in-scan sustained attention in neurodiverse youth. *Cerebral Cortex* **33**, 6320–6334 (2023).

[27] Manglani, H. R., Fountain-Zaragoza, S., Shankar, A., Nicholas, J. A. & Prakash, R. S. Employing connectome-based models to predict working memory in multiple sclerosis. *Brain Connectivity* **12**, 502–514 (2022).

[28] Lake, E. M. *et al.* The functional brain organization of an individual allows prediction of measures of social abilities transdiagnostically in autism and attention-deficit/hyperactivity disorder. *Biological psychiatry* **86**, 315–326 (2019).

[29] Yip, S. W., Scheinost, D., Potenza, M. N. & Carroll, K. M. Connectome-based prediction of cocaine abstinence. *American Journal of Psychiatry* **176**, 156–164

(2019).
[30] Rapuano, K. M. *et al.* Behavioral and brain signatures of substance use vulnerability in childhood. *Developmental cognitive neuroscience* **46**, 100878 (2020).
[31] Cao, H. *et al.* A functional connectome-based neural signature for individualized prediction of antipsychotic response in first-episode psychosis. *American Journal of Psychiatry* **180**, 827–835 (2023).
[32] Greene, A. S., Horien, C., Barson, D., Scheinost, D. & Constable, R. T. Why is everyone talking about brain state? *Trends in Neurosciences* (2023).
[33] Shen, X. *et al.* Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *nature protocols* **12**, 506–518 (2017).
[34] Franzmeier, N. *et al.* Functional brain architecture is associated with the rate of tau accumulation in alzheimer's disease. *Nature communications* **11**, 347 (2020).
[35] Marek, S. *et al.* Reproducible brain-wide association studies require thousands of individuals. *Nature* **603**, 654–660 (2022).
[36] Spisak, T., Bingel, U. & Wager, T. D. Multivariate bwas can be replicable with moderate sample sizes. *Nature* **615**, E4–E7 (2023).
[37] Gabrieli, J. D., Ghosh, S. S. & Whitfield-Gabrieli, S. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* **85**, 11–26 (2015).
[38] Pemberton, H. G. *et al.* Quantification of amyloid pet for future clinical use: a state-of-the-art review. *European journal of nuclear medicine and molecular imaging* **49**, 3508–3528 (2022).
[39] Malpetti, M., La Joie, R. & Rabinovici, G. D. Tau beats amyloid in predicting brain atrophy in alzheimer disease: implications for prognosis and clinical trials. *Journal of Nuclear Medicine* **63**, 830 (2022).
[40] Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289–300 (1995).
[41] Insel, P. S., Mormino, E. C., Aisen, P. S., Thompson, W. K. & Donohue, M. C. Neuroanatomical spread of amyloid β and tau in alzheimer's disease: implications for primary prevention. *Brain Communications* **2**, fcaa007 (2020).
[42] Johnson, K. A. *et al.* Tau positron emission tomographic imaging in aging and early a lzheimer disease. *Annals of neurology* **79**, 110–119 (2016).
[43] Jack Jr, C. R. *et al.* Defining imaging biomarker cut points for brain aging and alzheimer's disease. *Alzheimer's & Dementia* **13**, 205–216 (2017).
[44] He, T. *et al.* Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. *NeuroImage* **206**, 116276 (2020).
[45] Oldham, S. & Fornito, A. The development of brain network hubs. *Developmental cognitive neuroscience* **36**, 100607 (2019).
[46] Galinsky, A. D. *et al.* Are many sex/gender differences really power differences? *PNAS nexus* **3**, pgae025 (2024).
[47] Donohue, M. C. *et al.* The preclinical alzheimer cognitive composite: measuring amyloid-related decline. *JAMA neurology* **71**, 961–970 (2014).

[48] Sperling, R. A. *et al.* Association of factors with elevated amyloid burden in clinically normal older individuals. *JAMA neurology* **77**, 735–745 (2020).
[49] Horien, C., Shen, X., Scheinost, D. & Constable, R. T. The individual functional connectome is unique and stable over months to years. *Neuroimage* **189**, 676–687 (2019).
[50] Lutkenhoff, E. S. *et al.* Optimized brain extraction for pathological brains (optibet). *PloS one* **9**, e115551 (2014).
[51] Joshi, A. *et al.* Unified framework for development, deployment and robust testing of neuroimaging algorithms. *Neuroinformatics* **9**, 69–84 (2011).
[52] Satterthwaite, T. D. *et al.* An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* **64**, 240–256 (2013).
[53] Scheinost, D. *et al.* Functional connectivity during frustration: a preliminary study of predictive modeling of irritability in youth. *Neuropsychopharmacology* **46**, 1300–1306 (2021).
[54] Scheinost, D. *et al.* The intrinsic connectivity distribution: a novel contrast measure reflecting voxel level functional connectivity. *Neuroimage* **62**, 1510–1519 (2012).
[55] Frontzkowski, L. *et al.* Earlier alzheimer's disease onset is associated with tau pathology in brain hub regions and facilitated tau spreading. *Nature Communications* **13**, 4899 (2022).
[56] Scheinost, D. *et al.* Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage* **193**, 35–45 (2019).
[57] Marinescu, I. E., Lawlor, P. N. & Kording, K. P. Quasi-experimental causality in neuroscience and behavioural research. *Nature human behaviour* **2**, 891–898 (2018).
[58] Noble, S. *et al.* Multisite reliability of mr-based functional connectivity. *Neuroimage* **146**, 959–970 (2017).
[59] Greene, A. S. *et al.* Brain–phenotype models fail for individuals who defy sample stereotypes. *Nature* **609**, 109–118 (2022).
[60] Vogel, J. W. *et al.* Four distinct trajectories of tau deposition identified in alzheimer's disease. *Nature medicine* **27**, 871–881 (2021).
[61] Barthélemy, N. R. *et al.* Highly accurate blood test for alzheimer's disease comparable or superior to clinical csf tests. *Nature Medicine* 1–1 (2024).