Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Soluble amyloid composition links proximal APP processing, fibrillar burden and the p-tau181 response

Norberto Rodriguez-Espinosa

nrodesp@ull.edu.es

Universidad de La Laguna

Article

Keywords: amyloid-beta, APP processing, BACE1, cerebrospinal fluid, p-tau181, amyloid PET

Posted Date: October 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10955863/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Amyloid biomarkers sample distinct levels of an open, kinetically coupled biochemical system, yet scalar summaries may obscure biologically relevant configuration. We asked whether separating shared soluble-peptide variation from Aβ42-selective depletion preserves information and clarifies the amyloid context of the soluble tau response. In 380 Alzheimer's Disease Neuroimaging Initiative participants, CSF Aβ38, Aβ40 and Aβ42 mass-spectrometry measurements were used to derive two peptide-only analytical constructs: variation shared by Aβ38 and Aβ40 (T), and Aβ42-selective depletion relative to their joint profile (D42). BACE activity and sAPPβ preferentially aligned with T (incremental-R^2 separation 0.463; 95% bootstrap interval 0.361–0.535). D42 was preferentially associated with later fibrillar burden: among 91 participants with PET a median 4.98 years later, the D42–T contrast was 31.21 Centiloids/SD (95% CI 11.12–51.30). In 375 participants, p-tau181 was independently associated with both constructs (standardized coefficients 0.518 and 0.472), mapping across the T–D42 state space and revealing a supported region of elevated soluble tau. Information-retention analyses showed that tested scalar compressions discarded readout-relevant information. Together, these findings support a two-coordinate representation of soluble amyloid state that preserves biologically relevant information obscured by scalar summaries and locates the soluble tau response within the joint peptide configuration.

Introduction

Amyloid-β (Aβ) measurements interrogate a biochemical system rather than a single quantity. β-site APP-cleaving enzyme 1 (BACE1) cleavage of amyloid precursor protein (APP) releases soluble APPβ (sAPPβ) and a membrane-bound substrate for γ-secretase. Processive γ-secretase cleavage generates a distribution of Aβ peptide lengths rather than a single terminal product. [1, 2] Once released, peptides undergo transport, binding, assembly and clearance, and human kinetic studies show that CSF concentrations reflect the balance between synthesis and removal rather than production alone. [3] The measurable amyloid system is therefore open and kinetically coupled: concentrations describe aspects of its state but do not establish thermodynamic equilibrium or quantify the underlying molecular fluxes.

Measurements obtained at different levels of this system provide complementary biological reference points. CSF BACE activity reflects proximal enzymatic APP processing, sAPPβ is a released β-cleavage product, and same-assay Aβ38, Aβ40 and Aβ42 characterize soluble peptide composition. Amyloid PET measures tracer-accessible fibrillar burden, providing an independently acquired distal readout rather than a direct measure of plaque formation or accumulation rate. These measurements are related but not interchangeable. CSF and PET abnormalities can emerge asynchronously, demographic and genetic factors modify their relationship, [4, 5] and studies incorporating BACE1, sAPP fragments and multiple Aβ species identify distinct biomarker groupings and relationships with soluble and fibrillar amyloid. [6, 7] This motivates asking which components of soluble Aβ variation align with proximal processing and which with fibrillar expression.

Within the soluble compartment, Aβ42 alone cannot distinguish variation shared with other peptides from variation relatively selective to Aβ42. The Aβ42/Aβ40 ratio addresses part of this problem and has established diagnostic value, [8] but compresses numerator and denominator information into one scalar. Experimental differences between Aβ40 and Aβ42 oligomerization and nucleation provide a further reason why common soluble abundance and Aβ42-selective change need not vary together. [9, 10] For biological state characterization, we therefore define two complementary analytical coordinates: T, representing common soluble variation shared by Aβ38 and Aβ40, and D42, representing Aβ42-selective depletion relative to that shorter-peptide background. They are not intended as new molecular compartments or diagnostic biomarkers, but as a representation that preserves complementary aspects of soluble peptide variation. Only the peptide measurements define these coordinates; proximal processing, tau and PET measurements provide independent biological reference points against which the representation can be interpreted.

Soluble phosphorylated tau raises a distinct question. CSF p-tau increases in association with Aβ pathology before overt tau-PET positivity, and individual phosphorylation sites follow different trajectories during dominantly inherited Alzheimer disease. [11, 12] These observations support considering p-tau181 as a response-related readout of the amyloid-associated biochemical environment rather than equating it with mature neurofibrillary tangles. Its position relative to common soluble variation and Aβ42-selective depletion may therefore characterize the amyloid configuration associated with the soluble tau response.

Here, we used complementary biomarker measurements to examine the organization of the measured amyloid system from the partial information carried by each readout. We defined T and D42 to distinguish common shorter-peptide variation from Aβ42-selective depletion, and used proximal APP-processing markers and subsequent fibrillar amyloid burden as independently acquired biological reference points for this representation. We then asked how p-tau181 is positioned within the empirically anchored T-D42 state space. Finally, we assessed whether compressing the two-coordinate representation into the tested scalar summaries or projections discarded readout-relevant information. Our aim was to characterize the organization of measured amyloid state, not to infer the intrinsic dimensionality or temporal trajectory of Alzheimer disease.

Results

Peptide-only construction defines T and D42 independently of external readouts

The parent cohort comprised 380 participants with contemporaneous CSF BACE activity, sAPPβ, mass-spectrometry Aβ38/Aβ40/Aβ42 and complete covariates (Table 1). All 380 contributed to construction of the soluble coordinates. Elecsys p-tau181 was available in 375 participants, whereas first subsequent quality-controlled florbetapir PET was available in 91 (47 cognitively normal and 44 mild cognitive

impairment participants), at a median 4.98 years after CSF. Availability of p-tau181 or PET determined only the corresponding downstream analyses and did not enter coordinate construction.

Table 1
Cohort characteristics

Characteristic	Parent cohort (N = 380)	Later-PET subset (N = 91)
Age, years, mean (SD)	74.5 (7.1)	74.1 (6.2)
Female, n (%)	155 (40.8%)	38 (41.8%)
Education, years, mean (SD)	15.6 (3.0)	16.0 (3.1)
APOE ε4 carrier, n (%)	189 (49.7%)	31 (34.1%)
CN, n (%)	106 (27.9%)	47 (51.6%)
MCI, n (%)	182 (47.9%)	44 (48.4%)
AD dementia, n (%)	92 (24.2%)	0 (0.0%)
BACE activity, pM, median [IQR]	43.3 [32.7, 57.5]	44.1 [34.2, 61.5]
sAPPβ, pM, median [IQR]	3938.2 [3107.1, 4916.3]	3995.3 [3113.1, 5330.4]
MS Aβ38, pg/mL, median [IQR]	1748 [1426, 2190]	1780 [1554, 2302]
MS Aβ40, pg/mL, median [IQR]	7356 [6281, 9083]	7489 [6644, 9866]
MS Aβ42, pg/mL, median [IQR]	726 [537, 1156]	995 [686, 1494]
p-tau181, pg/mL, median [IQR]	27.2 [19.0, 37.4]	22.1 [16.9, 32.0]
Time to first subsequent PET, years, median [IQR]	-	4.98 [4.08, 5.16]

p-tau181 summaries refer to the 375 available participants. The p-tau181 analysis sample comprised 103 CN, 181 MCI and 91 dementia participants. All 380 contributed to eligible proximal analyses, and the 91-person PET subset was nested without defining eligibility for those analyses. CN, cognitively normal; MCI, mild cognitive impairment; SD, standard deviation; IQR, interquartile range.

T and D42 were derived exclusively from the three same-assay soluble Aβ peptides; BACE activity, sAPPβ, p-tau181 and PET were not used at any stage of their construction. T retained common standardized variation in log-Aβ38 and log-Aβ40. Repeated out-of-fold prediction of log-Aβ42 from the joint Aβ38/Aβ40 profile explained 26.8% of its variation (R^2=0.268). D42 was defined from the sign-reversed standardized out-of-fold residual, so higher D42 denotes less measurable Aβ42 than expected for the participant's shorter-peptide profile. The resulting coordinates therefore separate common shorter-peptide variation from Aβ42-selective departure before any external biological readout is examined. Despite cross-fitted residualization, T and D42 were empirically near-orthogonal (Pearson r ≈ 0.00 in the parent cohort, − 0.015 in the p-tau181 sample and − 0.016 in the PET subset; Spearman correlations were likewise near zero), with essentially unchanged estimates across 101 alternative

cross-fitting sequences (Supplementary Table S12). D42 remained closely aligned with conventional Aβ42/Aβ40 depletion (Spearman ρ = 0.992), indicating that the decomposition preserves the established relative-Aβ42 signal rather than creating a diagnostically unrelated biomarker. The coordinate origin represents the cohort mean and has no pathological interpretation; higher T alone was not classified as abnormal. Independently measured proximal, soluble-tau and fibrillar readouts were then used to orient this peptide-defined state space; subsequent information-retention analyses tested whether the resulting organization was preserved after scalar compression.

BACE activity and sAPPβ anchor the T direction

In simultaneous covariate-adjusted models, BACE activity and sAPPβ were strongly associated with T: standardized coefficients 0.516 (95% CI 0.433–0.598; $P = 1.67\times10^{-34}$) and 0.313 (0.220–0.406; $P = 3.76\times10^{-11}$), respectively. The proximal block contributed incremental R^2=0.468 for T and 0.006 for D42. The resulting separation was 0.463, with complete-pipeline bootstrap median 0.454 and interval 0.361–0.535. BACE-only and sAPPβ-only specifications preserved the separation (0.385 and 0.243). Thus the proximal anchor was supported by both an activity assay and a cleavage-product measurement, rather than either assay alone.

Later fibrillar burden preferentially maps onto the D42 direction

In the 91-person PET model, D42 was associated with 36.92 Centiloids/SD (95% CI 24.25–49.59; $P = 1.12\times10^{-8}$). The direct D42-minus-T contrast was 31.21 Centiloids/SD (95% CI 11.12–51.30; P = 0.0023), establishing differential association rather than relying on the significance of one coefficient and non-significance of another (Fig. 2). Complete-pipeline bootstrap intervals supported both the D42 coefficient and the contrast. Participant-level influence did not account for the contrast: across 101 alternative D42 cross-fitting sequences, all 91 leave-one-out refits retained a positive D42-minus-T contrast in every sequence (overall range 29.5–39.6 Centiloids/SD; Supplementary Table S12). Adding the relative Aβ38:Aβ40 balance preserved the D42 association at 37.53 Centiloids/SD (25.51–49.55; $P = 9.49\times10^{-10}$). The selective fibrillar association therefore extended beyond the common-soluble component and the relative balance of the two shorter peptides.

The marked proximal separation could not, by itself, establish complementary biological information. A fitted linear-Gaussian one-factor model passed through the same T/D42 construction pipeline reproduced the preferential association of BACE and sAPPβ with T, showing that residualization can generate this asymmetry even when the measured biomarkers share a single latent factor. A further distinction appeared in the PET ordering: the specified one-factor model predicted preferential association with T, whereas the observed data showed the opposite ordering, with fibrillar burden preferentially related to D42. None of 5,000 complete-pipeline simulations reproduced the observed positive D42-minus-T contrast (corrected directional P = 0.00020). A conditional PET wild-bootstrap preserving the observed biomarker geometry gave concordant support (P = 0.00035). These checks

apply to the specified linear measurement models and do not exclude all possible one-dimensional disease processes (Supplementary Table S2).

p-tau181 maps across both soluble coordinates

On equal-SD coordinates, standardized log-p-tau181 was independently associated with T and D42: βT = 0.518 (95% CI 0.454–0.581; P = 3.69×10^{-57}) and βD42 = 0.472 (0.394–0.550; P = 1.35×10^{-32}). Together the coordinates added R^2=0.425 beyond covariates, yielding total R^2=0.614. Both relationships remained supported when BACE and sAPPβ were added; the proximal block contributed a further R^2=0.030. The soluble tau association consequently included common-soluble and Aβ42-selective information while retaining an additional relationship with proximal processing (Supplementary Table S4).

These joint coefficients define a p-tau181 gradient at 42.4 degrees from + T toward +D42. Proximal BACE and sAPPβ directions were 4.6 and 6.9 degrees, whereas later PET was oriented at 88.7 degrees. Paired bootstrap intervals separated p-tau181 from BACE (angular difference 37.8 degrees; 95% CI 28.4–47.1) and sAPPβ (35.5 degrees; 22.0-48.1). Within the same 91 PET participants, the PET-minus-p-tau181 difference was 45.5 degrees (30.9–58.2). The diagonal p-tau181 response is therefore an observed feature of its joint associations, not a coordinate used to construct the plane (Supplementary Table S5).

The corresponding covariate-standardized response is displayed over the T-D42 plane in Fig. 3. The colour field and dashed 27 pg/mL contour are descriptive Gaussian-kernel summaries; the solid boundary is defined separately by the adjusted linear log-p-tau181 model. The response-region analysis identified locations where the lower simultaneous 95% bound for the geometric-mean p-tau181 response exceeded the published Elecsys reference level of 27 pg/mL. [13] This region occupied 44.5% of the adequately supported plotting grid (Supplementary Table S6); this is geometric grid area, not participant prevalence. The region describes a conditional population response, not individual diagnostic status, and its boundary is not a biological activation threshold. The main continuous result is the contribution of both soluble coordinates to the p-tau181 gradient.

When D42 and p-tau181 were considered together as predictors of later PET, both retained independent associations: 25.90 Centiloids/SD for D42 (95% CI 11.71–40.09; P = 3.47×10^{-4}) and 23.58 Centiloids/SD for p-tau181 (9.91–37.25; P = 7.20×10^{-4}). Their unique incremental R^2 contributions were 0.116 and 0.061, respectively. Thus the two predictors contributed non-redundant later-PET information, while the directional analysis remained distinct: fibrillar PET itself was predominantly D42-aligned (Table 2).

Table 2
Biological associations supporting the amyloid framework

Relationship	Estimate (95% interval)	P / inference	N
BACE activity with T	0.516 (0.433–0.598)	1.67×10^{-34}	380
sAPPβ with T	0.313 (0.220–0.406)	3.76×10^{-11}	380
Proximal-block R^2 separation	0.463 [0.361–0.535]	Bootstrap interval	380
D42 with later PET	36.92 (24.25–49.59)	1.12×10^{-8}	91
D42 minus T, later PET	31.21 (11.12–51.30)	0.0023	91
T with p-tau181	0.518 (0.454–0.581)	3.69×10^{-57}	375
D42 with p-tau181	0.472 (0.394–0.550)	1.35×10^{-32}	375
Joint PET model: D42	25.90 (11.71–40.09)	3.47×10^{-4}	91
Joint PET model: p-tau181	23.58 (9.91–37.25)	7.20×10^{-4}	91

Proximal and p-tau181 coefficients are standardized associations; PET coefficients are Centiloids/SD. The proximal-block entry is an incremental-R^2 contrast, with its complete-pipeline bootstrap interval in brackets. Other intervals are HC3 confidence intervals. Surface p-tau181 coefficients use equal-SD coordinates in the 375-person sample. Full counterpart estimates and sensitivities, including unsupported comparisons, are given in Supplementary Information.

The normalized directions in Fig. 4 integrate these relationships across the measured system: BACE and sAPPβ are predominantly T-aligned, p-tau181 incorporates both soluble coordinates, and later amyloid PET is predominantly D42-aligned. This directional summary complements the response surface in Fig. 3; it does not represent a measured sequence of transitions between biomarkers.

Information-retention analyses support the two-coordinate representation

Adding T to a conventional ratio-depletion model increased p-tau181 R^2 from 0.414 to 0.613, closely matching the T+D42 model (R^2=0.614); the T coefficient remained 0.467 (95% CI 0.402–0.532). Thus, preserving the common-soluble component adds p-tau181-associated information beyond the established relative Aβ42-depletion signal. A reduced-rank analysis likewise rejected a rank-one model in which BACE, sAPPβ and p-tau181 were all related to the same linear combination of T and D42 (restricted-null wild-bootstrap P = 0.000050). The corresponding mean joint cross-validated R^2 gain was modest (0.372 to 0.394), so this analysis is interpreted as an information-retention check rather than a major improvement in prediction.

To test whether such compression also discarded fibrillar information, a single linear score was learned from BACE, sAPPβ and p-tau181 in the 284 participants without PET. In the held-out 91-participant PET subset, the direction in T-D42 space complementary to that score remained strongly associated with

later PET (30.8 Centiloids/SD; HC3 95% CI 17.5–44.1; $P = 5.67\times10^{-6}$). Thus, within ADNI, a common score optimized for proximal-processing and soluble-tau readouts failed to retain information relevant to subsequent fibrillar burden. These analyses concern information retained by the measured coordinates and do not establish two latent biochemical causes. Full linear and nonlinear single-index comparisons, prediction results, negative controls and sensitivity estimates are reported in Supplementary Tables S7-S11.

Discussion

The central finding is that common soluble Aβ variation and Aβ42-selective depletion are not informationally interchangeable. Retaining them as separate peptide-derived coordinates preserves biologically relevant structure that is obscured by the tested scalar compressions. Because T and D42 were constructed without BACE, sAPPβ, p-tau181 or PET, their biological orientation was supplied by independently measured readouts: BACE activity and sAPPβ preferentially aligned with T, later fibrillar PET with D42, and p-tau181 mapped across both, between these orientations. The claim therefore concerns information retained within a measured amyloid state space, not proof that Alzheimer biology has two intrinsic dimensions or two molecular mechanisms.

The choice of coordinates follows the measurement problem. Isolated Aβ42 conflates shared peptide variation with species-selective change, while a ratio retains relative depletion but compresses the soluble background. T preserves that background; D42 represents Aβ42 departure after accounting for the joint shorter-peptide profile. These are explicit, biologically motivated analytical constructs rather than discovered molecular compartments. Their usefulness is established by the external-readout relationships and by the information lost under the tested scalar reductions, not by the algebraic fact that two variables can be plotted on a plane. This representation addresses biological state characterization rather than diagnostic classification, which remains the role of validated biomarker frameworks. [14]

A one-dimensional disease-progression trajectory and a two-coordinate measurement state space are not mutually exclusive. Jack and colleagues described nonlinear biomarker trajectories on a common progression axis and later refined that framework to accommodate variability and more complex amyloid-tau relationships. [15, 16] Subsequent models estimated latent progression scores and multivariate trajectories, [17, 18] accommodated subtype-specific sequences, [19] and represented disease progression geometrically. [20] A scalar trajectory may trace a path through a higher-dimensional observable space; cross-sectional T-D42 geometry does not determine that path. The present contribution is therefore to preserve complementary measured information within the amyloid domain before temporal order is assigned, not to replace progression models.

BACE activity and sAPPβ are especially informative because they provide complementary proximal reference measurements. An activity assay and a released cleavage product are not interchangeable with intracellular APP-cleavage flux, and their CSF levels can be influenced by compartmental exchange

and clearance. Nevertheless, their concordant T alignment anchors the common-soluble coordinate in proximal-processing-related variation rather than leaving it as an abstract statistical component. Later PET provides an independent distal reference: the information most closely associated with common soluble variation is not the information most strongly associated with tracer-accessible fibrillar burden. This organization is consistent with human studies reporting distinct proximal/soluble biomarker groupings and asynchronous soluble-fibrillar relationships. [4–7]

The observed organization is biologically plausible in a system with coupled generation, partitioning and removal. γ-secretase product allocation and the distinct assembly properties of Aβ42 can change measurable peptide composition without proportionately scaling every species. [1, 2, 9, 10] D42 may therefore contain information arising from product selection, binding, oligomerization, fibrillization, compartmental exchange or species-specific clearance, but it cannot be assigned exclusively to any of these mechanisms. Likewise, sustained synthesis and removal can coexist with nearly stable concentrations. [3] The state space describes complementary macroscopic measurements compatible with this kinetic complexity; it does not measure reaction rates, molecular structures or thermodynamic equilibrium.

p-tau181 reveals a further aspect of this organization. Its response was not captured by common-soluble variation or Aβ42-selective depletion alone; instead, it mapped across both coordinates, between the predominantly T-aligned proximal readouts and the predominantly D42-aligned fibrillar readout. This pattern is compatible with an amyloid-associated change in soluble tau metabolism, as suggested by observations of rising p-tau before overt tau-PET positivity and by phosphorylation-site-specific disease trajectories. [11, 12] p-tau181 is treated here as an integrative response-related readout of the measured amyloid environment, not as evidence that one peptide compartment directly causes tau pathology. Concentrations cannot separate phosphorylation, release and clearance, and the response surface does not identify the responsible cellular mechanism. Figures 3 and 4 provide complementary views of this result.

The distinction also clarifies the role of conventional amyloid biomarkers. A validated Aβ42/Aβ40 measure can be sufficient for detecting AD-related pathology while leaving common-soluble variation unresolved. Diagnostic criteria appropriately address detection and staging with validated biomarkers; [14] this study instead asks what configurations of measured amyloid accompany proximal processing, soluble tau and fibrillar expression. The independent T contribution alongside ratio depletion makes that distinction quantitative. We do not propose a diagnostic replacement, infer clinical resilience from the map, or establish treatment-selection utility.

The information-retention analyses provide formal support for keeping both coordinates visible, but their scope is deliberately limited. Rank two in the conditional linear model means that a common linear projection of T and D42 loses information across the measured readouts; it does not imply two latent biochemical causes. The nonlinear single-index comparison broadens that check but does not exhaust all possible scalar disease models. Likewise, the held-out PET analysis shows that a score learned from

proximal and soluble-tau readouts can omit PET-relevant information within ADNI, not that every one-dimensional summary is universally inferior.

Model-compatibility checks provide a complementary guardrail. Residualization alone can reproduce the large proximal T-D42 separation under a fitted linear one-factor measurement model, so that separation is not sufficient evidence of distinct biological structure. What that model did not reproduce was the observed reversal in external PET information, and the conditional PET wild bootstrap gave concordant support. These diagnostics concern specified linear generating models rather than every nonlinear, time-dependent or latent one-dimensional process. Internal bootstrap and cross-validation assess stability within ADNI and do not substitute for independent replication.

Important limitations concern the selected ADNI cohort, the small CN/MCI PET subset, the approximately five-year CSF-to-PET interval and the absence of contemporaneous PET. The distal association concerns later-measured burden rather than newly accumulated plaque. BACE technical reproducibility was limited, and covariate adjustment does not identify causal effects. Directional angles depend on scaling and adjustment; equal-SD normalization facilitates geometric comparison but does not equalize analytical reliability of T and D42. Response-region area depends on the specified grid and reference level and is not participant prevalence. The analyses were iterative and not preregistered, and independent-cohort replication is required.

In conclusion, retaining common soluble Aβ variation and Aβ42-selective depletion as separate peptide-derived coordinates preserves biologically relevant information that the tested scalar reductions do not fully retain. Independently measured readouts orient this state space differently: proximal BACE/sAPPβ preferentially toward T, subsequent fibrillar PET toward D42, and soluble p-tau181 across both. This framework characterizes configurations of measured amyloid state without assigning them to separate mechanisms or a temporal sequence, and provides a basis for longitudinal investigation of how those configurations and the tau response evolve together.

Methods

Study design, measurements and analytical hierarchy

This secondary observational study used de-identified ADNI data under the ADNI Data Use Agreement. The ADNI study protocols were approved by the University of Southern California Health Sciences Campus Institutional Review Board (USC HSC IRB), and written informed consent was obtained from all participants and/or their legally authorized representatives. Eligibility required contemporaneous BACE activity, sAPPβ and mass-spectrometry Aβ38/Aβ40/Aβ42 and complete age, sex, education, APOE ε4 dosage and diagnosis. The parent cohort contained 380 unique participants; missing p-tau181 and missing PET defined their respective 375- and 91-person analysis sets without reducing upstream eligibility. The analytical framework had three principal stages. First, T and D42 were constructed exclusively from soluble Aβ peptide measurements. Second, independently measured proximal APP-

processing readouts, p-tau181 and later fibrillar PET were used to biologically orient the resulting state space. Third, information-retention analyses tested whether compressing the T-D42 representation into the evaluated scalar summaries or projections discarded readout-relevant information. Response-surface, model-compatibility, conditional and selection analyses were treated as supporting or sensitivity analyses. Analyses were developed iteratively, not preregistered; supplementary estimates retain the exploratory and sensitivity results.

BACE activity and sAPPβ came from the Biomarkers Consortium CSF project: a two-step enzyme/substrate assay with immunodetection and a specific product immunoassay, respectively, both reported in pM. [21, 22] Joint technical replicates were combined on the log scale. Aβ38, Aβ40 and Aβ42 were quantified within the University of Pennsylvania mass-spectrometry assay and reported in pg/mL. [23, 24] Elecsys p-tau181 provided the soluble-tau measurement. BACE/sAPPβ and mass-spectrometry measurements were matched by collection date. Subsequent PET used the first quality-controlled florbetapir scan strictly after CSF, expressed in Centiloids. [25] Duplicate scan-date/tracer records retained the latest processing record. CSF collection defined time zero; PET was not used to construct the soluble coordinates.

Peptide-only soluble-coordinate construction

T and D42 were constructed exclusively from same-assay CSF Aβ38, Aβ40 and Aβ42 measurements. External readouts - including BACE activity, sAPPβ, p-tau181 and amyloid PET - did not enter coordinate construction and were introduced only in subsequent analyses. Peptide concentrations were log-transformed to represent proportional differences and then standardized. With z denoting centring and unit-SD scaling, the common-soluble coordinate was:

$$T_i \;=\; z\,\{[\,z(log\ A\beta\,38_i)\;+\;z(log\ A\beta\,40_i)]/2\}. \quad (1)$$

Expected log-Aβ42 was estimated from log-Aβ38 and log-Aβ40 using 50 repeated participant-level five-fold splits. The 50 held-out predictions per participant were averaged; D42 was the sign-reversed standardized residual:

$$D42_i \;=\; -z\{log\ A\beta\,42_i \;-\; predicted\,log\ A\beta\,42_i, OOF\}. \quad (2)$$

The conditional expectation uses both shorter peptides rather than T alone, thereby preserving relative Aβ38:Aβ40 composition when estimating expected Aβ42. Positive D42 therefore denotes less measurable Aβ42 than expected for the participant's joint shorter-peptide profile. Cross-fitting reduces self-prediction of the generated coordinate. [26] T represents common soluble Aβ38/Aβ40 variation and is neither total Aβ nor a measure of production rate; its notation is unrelated to the tau category of AT(N). For state-space analyses, T and D42 were placed on an equal-SD metric in the 375-participant p-tau181 sample, and the same scaling was applied to the PET subset. Equal-SD scaling facilitates directional comparison but does not imply equal analytical reliability of the two coordinates. Complete-pipeline resampling and predictive analyses re-estimated transformations within the corresponding resample or training set as specified below. Empirical T–D42 dependence was assessed by Pearson and

Spearman correlation in the parent, p-tau181 and PET samples, with stability evaluated across alternative cross-fitting sequences.

Proximal and later-PET association models

Parallel ordinary least-squares models related T and D42 to standardized log-BACE and log-sAPPβ simultaneously, adjusting for age, sex, APOE ε4 dosage, education and diagnosis. HC3 heteroskedasticity-consistent covariance provided coefficient uncertainty. [27] Proximal separation was the difference between block incremental R^2 values, each calculated as full-model minus covariate-only R^2. Stability used 1,000 participant bootstraps rebuilding standardization, T and cross-fitted D42, and 200 repeated five-fold predictive comparisons. [28] A complementary reparameterization separated common Aβ38/Aβ40 variation from their relative balance, G38:40, to evaluate whether the selective PET association reflected shorter-peptide composition.

The primary PET model included D42, T, BACE, sAPPβ, CSF-to-PET interval and the clinical/genetic covariates. The signed βD42-minus-βT contrast directly compared soluble-coordinate associations on their reference scales. A separate joint model added standardized log-p-tau181. Incremental R^2 and partial R^2 were distinguished: the latter is the reduction in residual sum of squares relative to the reduced model. The PET outcome was later burden, not a rate of formation or accumulation. Participant-level influence was assessed by omitting each PET participant in turn, refitting the full model and recalculating the D42-minus-T contrast. Robustness of this audit to coordinate construction was evaluated across 101 alternative D42 cross-fitting sequences.

Model-compatibility and operational controls

A maximum-likelihood linear-Gaussian one-factor model was fitted to covariate-residualized, unit-variance log-BACE, log-sAPPβ and log-Aβ38/40/42. Five thousand datasets generated from its loadings and independent indicator-specific residuals underwent the complete coordinate and association pipeline. Covariate means were reintroduced, the PET-observation pattern was retained, and PET was generated from the same factor with an effect calibrated to the observed biomarker-block PET variance. A separate reduced PET mean excluding D42 generated 20,000 leverage-adjusted Rademacher wild-bootstrap outcomes while holding observed biomarker geometry fixed. [29] Corrected empirical probabilities used (exceedances + 1)/(replicates + 1). [30] These are diagnostics of specified generating models; fitted factor parameters were treated as plug-in estimates.

Operational sensitivities included alternative T definitions, quadratic Aβ42 mapping, ratio depletion, Elecsys Aβ42 substitution, a D38 negative-control residual and technical-replicate choice. PET-availability weighting was restricted to the 288-person CN/MCI target population because no baseline dementia participant had PET. Stabilized weights were truncated at the 1st and 99th percentiles, with effective sample size and balance diagnostics reported. [31] These controls address different measurement and selection vulnerabilities; their complete results appear in the supplement.

p-tau181 associations, geometry and response-region inference

Log-p-tau181 was related to T and D42 with the standard covariates; an expanded model also included BACE and sAPPβ. For directional comparisons, each readout was regressed on both equal-SD coordinates and covariates. Direction was θ = atan2(βD42,βT). PET directional models additionally included time to PET; unlike fully adjusted PET-association models, they did not include proximal markers. Ten thousand participant bootstraps estimated angular uncertainty. Paired p-tau-versus-proximal contrasts used the same 375 participants; the paired PET-versus-p-tau contrast used the same 91. Angles describe adjusted association orientation, not molecular geometry or disease time.

For visualization, a two-dimensional Gaussian kernel smoother displayed covariate-standardized p-tau181 over the T-D42 plane, using a bandwidth of 0.38 SD selected by leave-one-out prediction. This descriptive colour field and its dashed 27 pg/mL contour were not used for statistical inference. The covariate-adjusted linear log-p-tau181 model supplied the gradient arrow and the response-region confidence boundary. A quadratic extension was assessed by an HC3 joint test, AIC, BIC and 100 repeated ten-fold cross-validation, with comparison results retained in Supplementary Table S6. Response-region inference exponentiated predictions at fixed cohort-mean covariates, yielding geometric rather than arithmetic means. A 150-by-150 grid covered coordinate 1st-99th percentiles; support required the convex hull and Gaussian-kernel Kish effective N at least 20 with bandwidth 0.38 SD. Five thousand participant coefficient bootstraps gave an approximate simultaneous 95% band by the maximum standardized excursion over supported points. The region criterion was a lower bound above the published 27 pg/mL Elecsys reference. [13] Coordinates, support and covariate reference were fixed for this inference. A confidence region for a conditional mean is not an individual diagnostic classification.

Information-retention and conditional analyses

Reduced-rank regression compared Y = ZA+XB + E with rank(B) = 1 or rank(B) = 2, where X contained T/D42, Y contained separately standardized log-BACE, log-sAPPβ and log-p-tau181, and Z contained an intercept and covariates. [32] A was unrestricted; equal response weighting defined the loss. The statistic N log(SSE1/SSE2) was calibrated with 20,000 restricted-null wild-bootstrap replicates using the same random sign for each participant's whole residual vector. A separate 2,000-replicate parent-cohort bootstrap rebuilt coordinates while keeping bootstrap copies of each participant in the same fold. The main predictive comparison used 100 repeated five-fold participant splits with training-only scaling and coordinate construction; fold errors were pooled within repetitions. Repeated-split variation was not interpreted as sampling confidence intervals. Rank two here refers to information retained by the measured coordinates and readouts, not to two latent biochemical causes.

A nested sensitivity allowed response-specific polynomials of one common linear score, with degree and curvature regularization selected by inner cross-validation; 30 repeated five-fold outer evaluations

compared this family with the two-coordinate model. For PET, the rank-one score was learned exclusively in 284 non-PET participants; the perpendicular direction in that training-standardized metric was then tested in the other 91, alongside the retained score, proximal markers, clinical covariates and time to PET. Five thousand two-stratum bootstraps refitted score learning and PET regression. These PET checks used inherited parent-cohort coordinates and constitute held-out-participant score validation, not independent-cohort replication. Orthogonality does not imply statistical independence.

Conditional p-tau analyses examined D42 relationships before and after including p-tau181 and evaluated the T-aligned portion of proximal-marker information. They characterize changes in the conditional estimand, not mediation or causal inhibition. Ordinary coefficient tests were two-sided; directional simulation probabilities were model checks. Exploratory comparisons were not globally multiplicity-adjusted, and their complete estimates are retained in the supplement. The reference-region band controls simultaneous uncertainty over its specified plotting domain only. Analyses used Python with NumPy, pandas, SciPy, statsmodels and scikit-learn.

Declarations

Acknowledgments.
Data collection and sharing for the Alzheimer's Disease Neuroimaging Initiative (ADNI) is funded by the National Institute on Aging (National Institutes of Health Grant U19 AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health (FNIH), including 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 author also acknowledges the ADNI participants and study partners.

Author contributions
Norberto Rodriguez-Espinosa: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing – original draft; Writing – review and editing. The author read and approved the final manuscript.

Ethics approval, informed consent and regulatory compliance

This secondary observational study used de-identified data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) under the ADNI Data Use Agreement. The ADNI study protocols were approved by the

University of Southern California Health Sciences Campus Institutional Review Board (USC HSC IRB), and written informed consent was obtained from all participants and/or their legally authorized representatives. All methods in the present secondary analysis were carried out in accordance with relevant guidelines and regulations.

Data availability

The participant-level dataset used in this study is not publicly available as unrestricted/open-access data. The data are owned and distributed by the Alzheimer's Disease Neuroimaging Initiative (ADNI) under controlled-access conditions.

Researchers may obtain the same de-identified participant-level data by applying through the LONI Image and Data Archive (IDA), accepting the ADNI Data Use Agreement, and submitting an online Data Use Application describing their institutional affiliation and proposed research use. Once approved by ADNI, investigators can download the required study files through the IDA Study Files interface. Under the ADNI Data Use Agreement, participant-level ADNI data cannot be publicly redistributed by individual investigators; for this reason, we cannot deposit the participant-level dataset used for this study with the article or in a public repository.

For reproduction of the present analyses, the required ADNI source datasets are:

- Biomarkers Consortium CSF Targeted Proteomics Project — BACE1 activity and sAPPβ measurements;
- UPENNMSMSABETA — University of Pennsylvania mass-spectrometry CSF Aβ38, Aβ40 and Aβ42 measurements;
- UPENNBIOMK_ROCHE_ELECSYS — CSF Elecsys p-tau181 measurements;
- UCBERKELEY_AMY_6MM — UC Berkeley 6-mm amyloid-PET summary data, including Centiloid values and quality-control information;
- the corresponding ADNI harmonized clinical/demographic/genetic tables (ADSL/ADNIMERGE) for age, sex, education, APOE ε4 and clinical diagnosis.

The exact manuscript-matched source releases and variables required for reconstruction are specified in the reproducibility package/data manifest accompanying the analysis code. The public code repository contains the complete analysis specifications and aggregate outputs, but no participant-level ADNI data.

Thus, the study data are best described as controlled-access third-party data available to approved investigators from ADNI, rather than openly publicly available data.

Code availability

Analysis code for cohort construction, peptide-only T/D42 derivation, primary association models, resampling, reduced-rank and single-index analyses, model-compatibility checks, influence audits and

figure-generation support is publicly available on Zenodo at https://doi.org/10.5281/zenodo.22639080. The repository can also be accessed directly at https://zenodo.org/records/22639080. The repository contains no participant-level ADNI data; full numerical reproduction requires authorized access to the manuscript-matched ADNI source releases listed in its data manifest.

Competing interests

The author declares no competing interests related to this work.

References

1. Takami, M. et al. γ-Secretase: successive tripeptide and tetrapeptide release from the transmembrane domain of β-carboxyl terminal fragment. *J. Neurosci.* **29**, 13042–13052 (2009).
2. Chávez-Gutiérrez, L. et al. The mechanism of γ-secretase dysfunction in familial Alzheimer disease. *EMBO J.* **31**, 2261–2274 (2012).
3. Bateman, R. J. et al. Human amyloid-β synthesis and clearance rates as measured in cerebrospinal fluid in vivo. *Nat. Med.* **12**, 856–861 (2006).
4. Sala, A. et al. Longitudinal pathways of cerebrospinal fluid and positron emission tomography biomarkers of amyloid-β positivity. *Mol. Psychiatry*. **26**, 5864–5874 (2021).
5. Cacciaglia, R. et al. Age, sex and APOE-ε4 modify the balance between soluble and fibrillar β-amyloid in non-demented individuals: topographical patterns across two independent cohorts. *Mol. Psychiatry*. **27**, 2010–2018 (2022).
6. Lehmann, S. et al. Clarifying the association of CSF Aβ, tau, BACE1, and neurogranin with AT(N) stages in Alzheimer disease. *Mol. Neurodegener*. **19**, 66 (2024).
7. Gao, F. et al. Associations of CSF BACE1 with amyloid pathology, neurodegeneration, and cognition in Alzheimer's disease. *Acta Neuropathol.* **147**, 97 (2024).
8. Janelidze, S. et al. CSF Aβ42/Aβ40 and Aβ42/Aβ38 ratios: better diagnostic markers of Alzheimer disease. *Ann. Clin. Transl Neurol.* **3**, 154–165 (2016).
9. Bitan, G. et al. Amyloid beta-protein (Aβ) assembly: Aβ40 and Aβ42 oligomerize through distinct pathways. Proc. Natl Acad. Sci. USA 100, 330–335 (2003). 10.1073/pnas.222681699
10. Meisl, G. et al. Differences in nucleation behavior underlie the contrasting aggregation kinetics of the Aβ40 and Aβ42 peptides. Proc. Natl Acad. Sci. USA 111, 9384–9389 (2014). 10.1073/pnas.1401564111
11. Mattsson-Carlgren, N. et al. Aβ deposition is associated with increases in soluble and phosphorylated tau that precede a positive Tau PET in Alzheimer's disease. *Sci. Adv.* **6**, eaaz2387. 10.1126/sciadv.aaz2387 (2020).
12. Barthélemy, N. R. et al. A soluble phosphorylated tau signature links tau, amyloid and the evolution of stages of dominantly inherited Alzheimer's disease. *Nat. Med.* **26**, 398–407. 10.1038/s41591-020-0781-z (2020).

13. Blennow, K. et al. Predicting clinical decline and conversion to Alzheimer's disease or dementia using novel Elecsys Aβ(1–42), pTau and tTau CSF immunoassays. *Sci. Rep.* **9**, 19024. 10.1038/s41598-019-54204-z (2019).
14. Jack, C. R. Jr et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimers Dement.* **20**, 5143–5169. 10.1002/alz.13859 (2024).
15. Jack, C. R. Jr et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* **9**, 119–128. 10.1016/S1474-4422(09)70299-6 (2010).
16. Jack, C. R. Jr et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* **12**, 207–216. 10.1016/S1474-4422(12)70291-0 (2013).
17. Jedynak, B. M. et al. A computational neurodegenerative disease progression score: method and results with the Alzheimer's disease Neuroimaging Initiative cohort. *Neuroimage* **63**, 1478–1486. 10.1016/j.neuroimage.2012.07.059 (2012).
18. Donohue, M. C. et al. Estimating long-term multivariate progression from short-term data. *Alzheimers Dement.* **10** 10.1016/j.jalz.2013.10.003 (2014). S400-S410.
19. Young, A. L. et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nat. Commun.* **9**, 4273. 10.1038/s41467-018-05892-0 (2018).
20. Koval, I. et al. AD Course Map charts Alzheimer's disease progression. *Sci. Rep.* **11**, 8020. 10.1038/s41598-021-87434-1 (2021).
21. Wu, G. et al. Decrease in age-adjusted cerebrospinal fluid β-secretase activity in Alzheimer's subjects. *Clin. Biochem.* **41**, 986–996 (2008).
22. Wu, G. et al. Decrease in brain soluble amyloid precursor protein β (sAPPβ) in Alzheimer's disease cortex. *J. Neurosci. Res.* **89**, 822–832 (2011).
23. Korecka, M. et al. Qualification of a surrogate matrix-based absolute quantification method for Aβ1–42 in human cerebrospinal fluid using 2D UPLC-tandem mass spectrometry. *J. Alzheimers Dis.* **41**, 441–451 (2014).
24. Lame, M. E., Chambers, E. E. & Blatnik, M. Quantitation of amyloid beta peptides Aβ1–38, Aβ1–40, and Aβ1–42 in human cerebrospinal fluid by ultra-performance liquid chromatography-tandem mass spectrometry. *Anal. Biochem.* **419**, 133–139 (2011).
25. Klunk, W. E. et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimers Dement.* **11**, 1–15e4 (2015).
26. Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. R Stat. Soc. B*. **36**, 111–133 (1974).
27. MacKinnon, J. G. & White, H. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *J. Econom*. **29**, 305–325 (1985).
28. Efron, B. Bootstrap methods: another look at the jackknife. *Ann. Stat.* **7**, 1–26 (1979).

29. Davidson, R. & Flachaire, E. The wild bootstrap, tamed at last. *J. Econom*. **146**, 162–169 (2008).
30. Phipson, B. & Smyth, G. K. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* **9**, Article 39 (2010).
31. Cole, S. R. & Hernán, M. A. Constructing inverse probability weights for marginal structural models. *Am. J. Epidemiol.* **168**, 656–664 (2008).
32. Izenman, A. J. Reduced-rank regression for the multivariate linear model. *J. Multivar. Anal.* **5**, 248–264. 10.1016/0047-259X(75)90042-1 (1975).

Figures

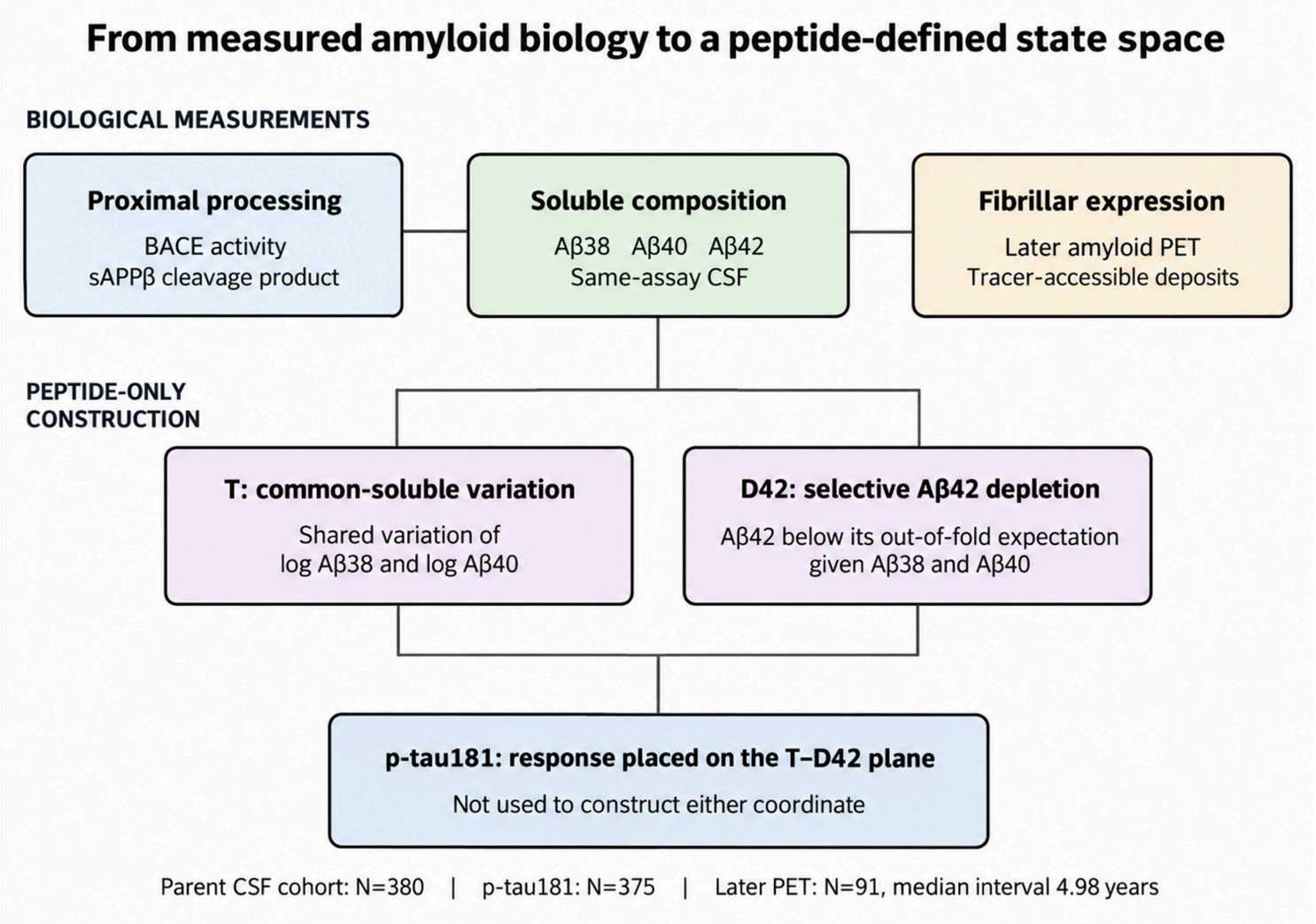


Figure 1

Rationale and structure of the peptide-defined amyloid state space. Upper boxes summarize the biological measurement levels examined in this study: proximal APP processing (CSF BACE activity and sAPPβ), soluble amyloid composition measured from same-assay CSF Aβ38, Aβ40, and Aβ42, and subsequent fibrillar expression indexed by later amyloid PET. The analytical coordinates T and D42 were derived only from the soluble peptide measurements. T captures variation shared by Aβ38 and Aβ40,

whereas D42 captures Aβ42-selective depletion relative to its out-of-fold expectation given Aβ38 and Aβ40. BACE activity, sAPPβ, p-tau181, and amyloid PET were excluded from coordinate construction and were used only as external readouts to evaluate the biological information retained by the two-coordinate representation. p-tau181 was then mapped onto the T–D42 state space. The scheme summarizes measurement levels and analysis structure and does not imply a demonstrated temporal or causal sequence.

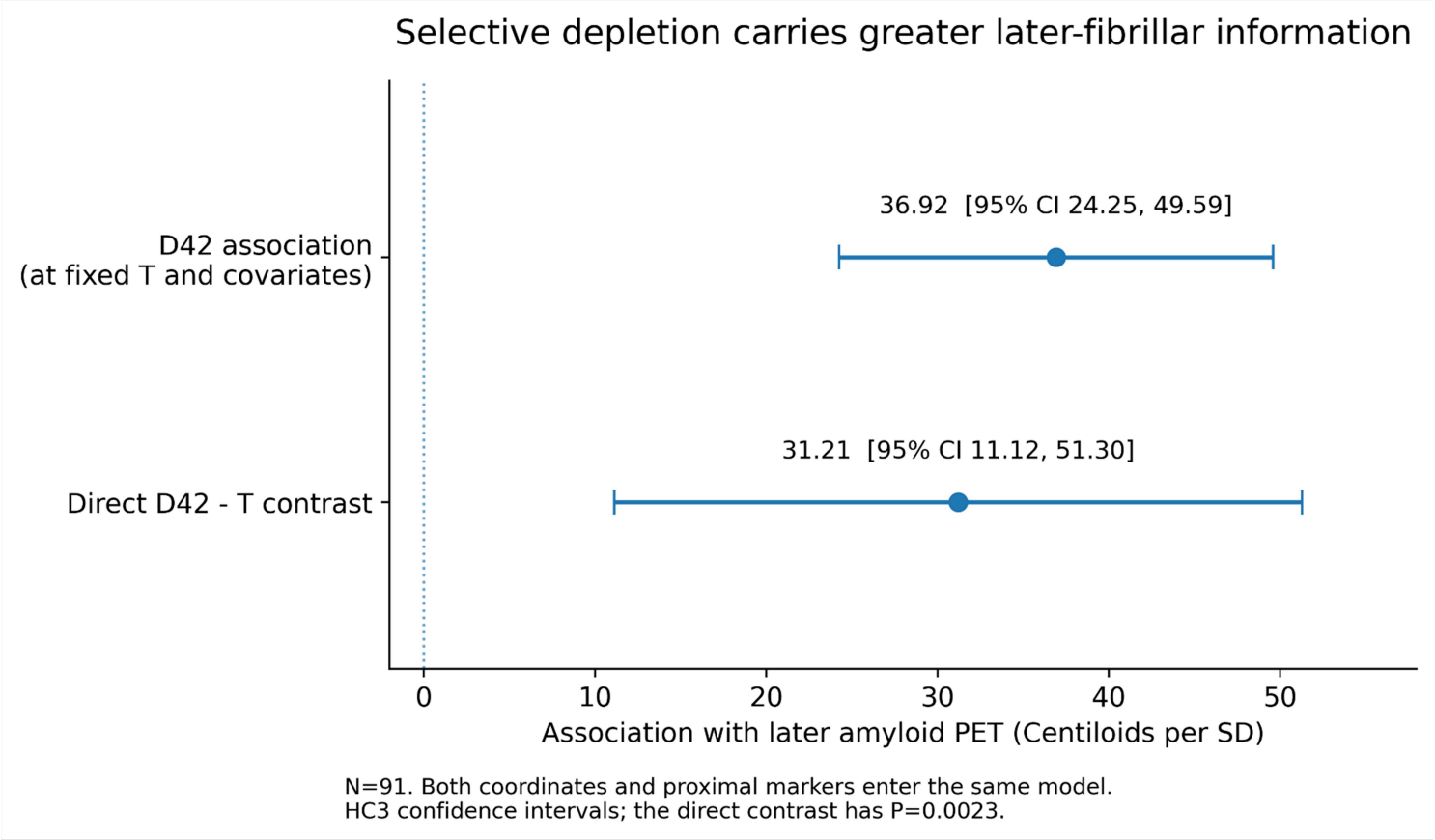


Figure 2

Aβ42-selective depletion carries greater later-fibrillar information. Points and bars are HC3 estimates and 95% confidence intervals for D42 and for the directly tested D42-minus-T contrast in the same 91-person PET model. The model includes both coordinates, proximal markers, clinical/genetic covariates and time to PET. Displaying the contrast avoids inferring a difference solely from the significance of one coefficient and non-significance of another. Units are Centiloids per coordinate SD.

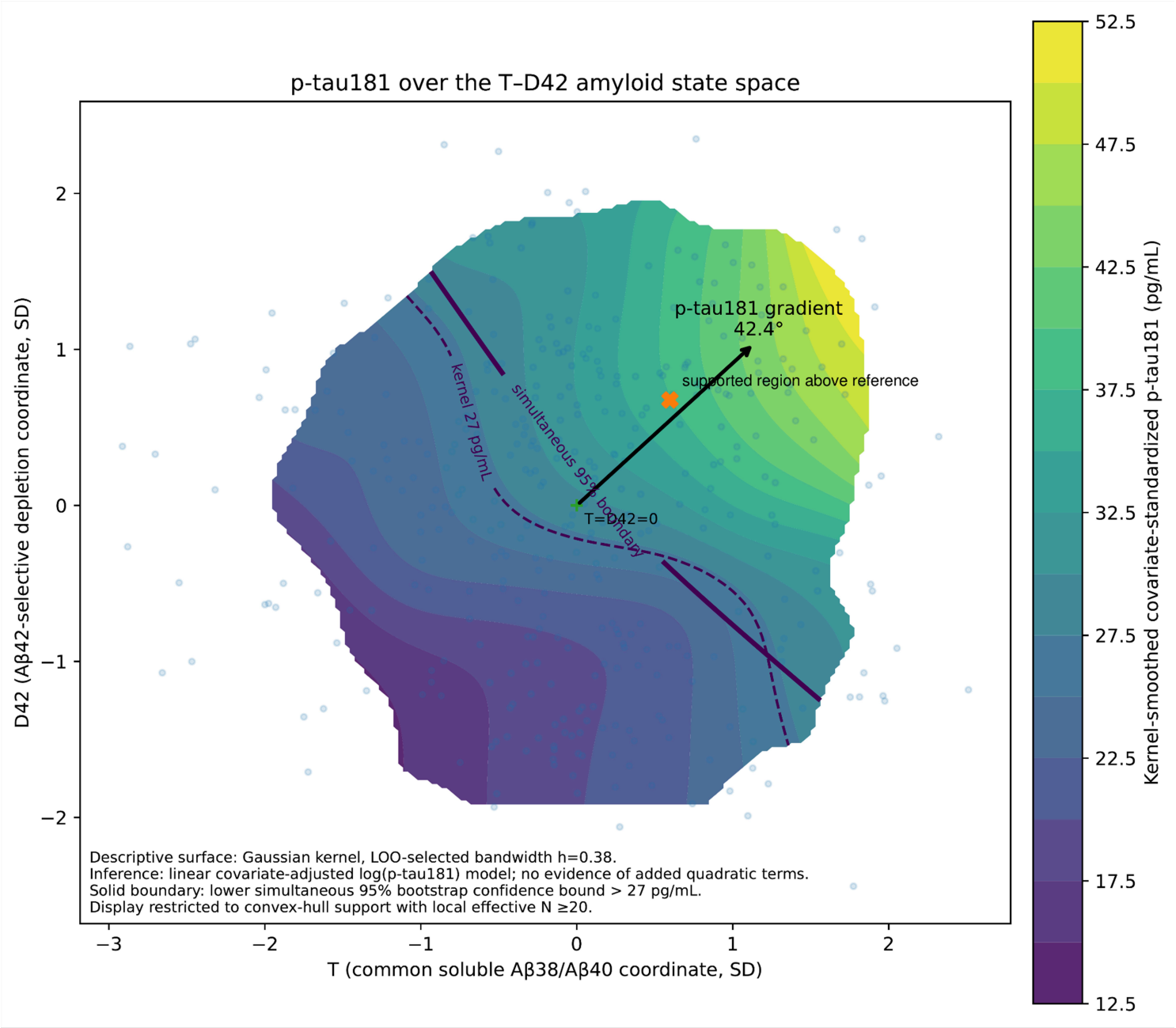


Figure 3

p-tau181 over the T-D42 amyloid state space. The colour field is a descriptive Gaussian-kernel surface of covariate-standardized p-tau181 (leave-one-out-selected bandwidth 0.38 SD); the dashed contour marks its 27 pg/mL level. Separately, the solid boundary marks where the lower simultaneous 95% bootstrap confidence bound from the adjusted linear log-p-tau181 model equals the 27 pg/mL reference. On the higher-response side, the bound exceeds that level. The arrow shows the fitted p-tau181 gradient (42.4 degrees), not a temporal trajectory. The orange marker lies within the displayed region where the lower simultaneous bound exceeds the 27 pg/mL reference; the orange cross marks T=D42=0. The label describes this inferential criterion and does not denote a discrete biological state. Points show observed coordinate locations; shading is restricted to convex-hull support with local effective N at least 20. The origin is the cohort mean. A location outside the lower-bound region is not thereby p-tau-negative. Kernel

curvature is descriptive, and the reference contour is not a biological onset threshold or an individual diagnostic rule (N=375).

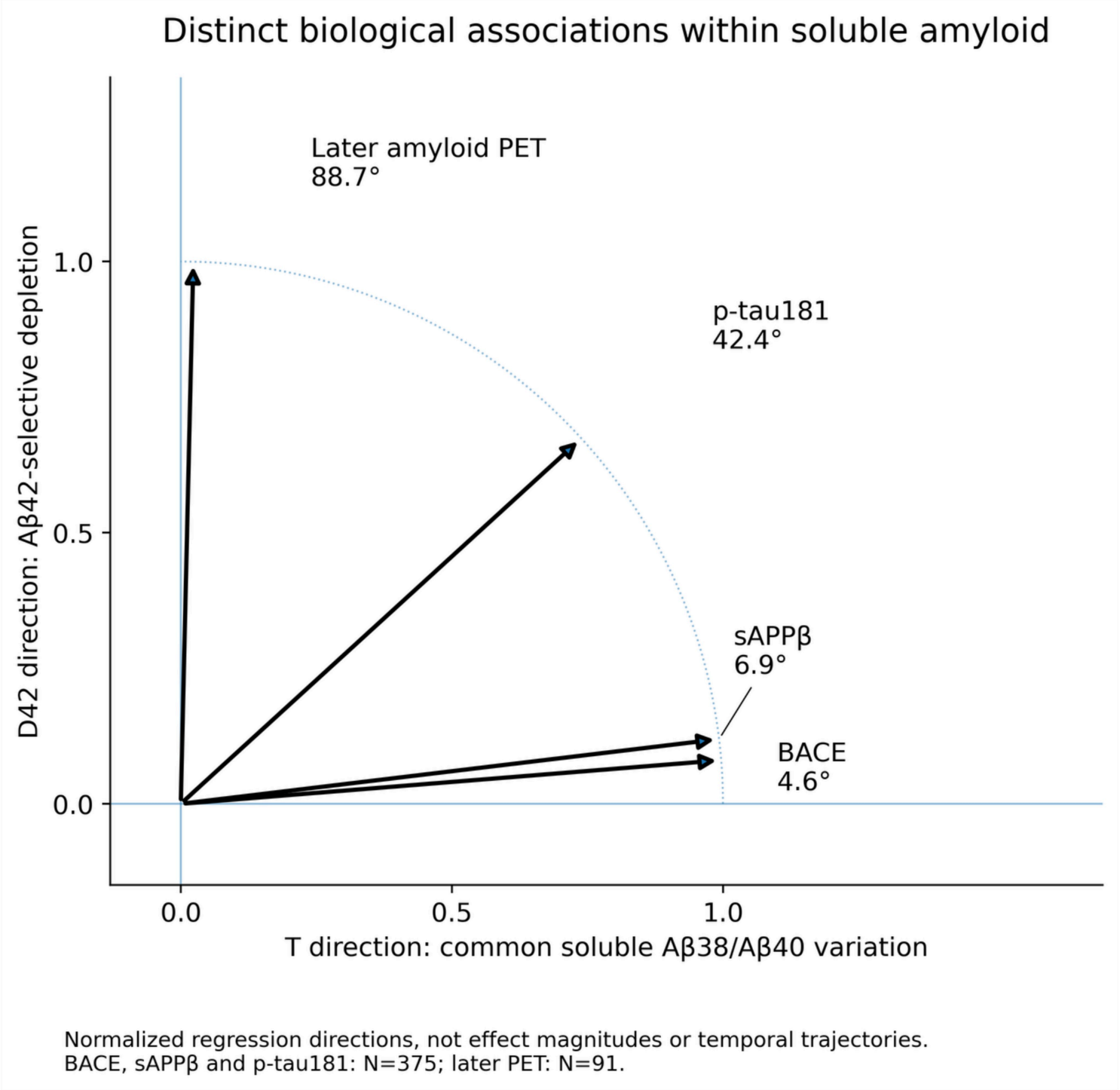


Figure 4

Directional synthesis of the proximal, soluble-tau and fibrillar associations. Arrows show covariate-adjusted regression gradients in the equal-SD T-D42 metric, normalized to equal length. BACE and sAPPβ are predominantly T-aligned; p-tau181 incorporates both coordinates; later amyloid PET is predominantly D42-aligned. T and D42 are peptide-derived coordinates, not variables constructed using these external

readouts. The origin and unit arc display direction vectors, not participant positions or biomarker thresholds. Equal lengths do not imply equal effect sizes, and angular order is not temporal order. Equal-SD normalization facilitates comparison of direction but does not equate analytical reliability. BACE, sAPPβ and p-tau181 models use 375 participants; PET uses 91 and additionally adjusts for the CSF-to-PET interval. Paired angular contrasts, including PET versus p-tau181 within the same 91 participants, are reported in Supplementary Table S5. This comparative summary complements the response surface in Fig. 3.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 02SupplementaryInformation.pdf