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.

Distinguishing discrete subtypes from continuous variation in Alzheimer's disease brain proteomes

Vijay Kumar

vijay.kumar@agra.sharda.ac.in

Department of Biotechnology, Anand School of Engineering and Technology, Sharda University Agra

Bhuvaneshwar Yarlagadda

GloNeuro

Article

Keywords:

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11503653/v1

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

Additional Declarations: There is **NO** Competing Interest.

Distinguishing discrete subtypes from continuous variation in Alzheimer's disease brain proteomes

Bhuvaneshwar Yarlagadda[1], Vijay Kumar[2,*]

[1] GloNeuro, Sector 107, Vishwakarma Road, Noida 201301, Uttar Pradesh, India.

[2] Department of Biotechnology, Anand School of Engineering and Technology, Sharda University Agra, Agra, Uttar Pradesh 282007, India.

[*]Correspondence: vijay.kumar@agra.sharda.ac.in

Abstract

Molecular subtypes of Alzheimer's disease (AD) are increasingly derived by clustering brain proteomes, and their replication in independent cohorts is taken as evidence that they are distinct groups of patients. Clustering, however, divides any dataset into groups, including continuous data, and how well replication distinguishes discrete subtypes from continuous variation is unknown. We addressed this question with the three AD subtypes of the Pan-Neurodegeneration Atlas (PanNDA), a proteomic resource with three cohorts and open data. Data simulated without subtypes met the PanNDA replication criterion in 37% of datasets; meeting the criterion was therefore at most 2.7 times more likely with subtypes than under a continuum shared by the cohorts. The agreement between the observed cohorts was consistent with continuous variation. Subtypes common to all cohorts and as separated as the published ones were unlikely, whereas weaker subtypes and subtypes whose profiles differ between cohorts remain possible. The subtypes corresponded largely to positions on two continuous axes: disease severity, and a contrast between astrocyte and parvalbumin-interneuron proteins that was also present in control brains. Replication therefore needs calibration against continuous data to support discrete subtypes, and positions on continuous axes provide a complementary description of molecular heterogeneity in AD.

Introduction

Patients with Alzheimer's disease (AD) share amyloid-β plaques and neurofibrillary tangles of hyperphosphorylated tau, but differ in the regional distribution of pathology, the rate of cognitive decline and the cellular responses that accompany neurodegeneration. This heterogeneity complicates clinical trials, as a treatment that modifies one disease process may fail in patients in whom another predominates. Dividing patients into subtypes is one proposed solution, and subtypes of AD have been derived from the spread of tau pathology[1], the sequence of regional brain atrophy[2], bulk brain transcriptomes[3], brain proteomes[4] and cerebrospinal fluid proteomes[5,6]. If patients fall into discrete molecular groups, each subtype could correspond to a distinct disease process with its own treatment.

Whether the groups are discrete is rarely tested. Subtypes are usually found by unsupervised clustering, for example consensus clustering[7,8], which assigns every sample to one of a chosen number of groups whether or not separate groups exist[9,10]. In continuous data, clustering cuts a gradient into segments. These segments can still agree between cohorts: cohorts that share the same correlation structure among their molecules are cut at similar places, and differences between clusters estimated from the same data are inflated by the clustering itself[11]. Agreement between cohorts is therefore not, by itself, evidence of discrete groups. In earlier work, some atrophy subtypes of AD partly reflected disease stage[12], proposed subtypes of depression were indistinguishable from a single continuous population[13], subtyping methods that agreed at the level of groups disagreed at the level of individuals[14,15], and cancer subtypes largely reflected the composition of the sampled tissue[16–18].

The Pan-Neurodegeneration Atlas (PanNDA) of Shrestha and colleagues is a recent and openly available example[19]. In three AD cohorts, PanNDA selected 340 proteins whose abundance differed between AD and control brains and divided the AD cases of a discovery cohort into “three distinct molecular subtypes” (ST1 to ST3), which PanNDA associated with synaptic signalling, cytoskeletal regulation and blood-brain barrier transport[19]. The subtypes were derived by consensus clustering, which repeats clustering on random subsets of samples and records how often each pair of samples is assigned to the same cluster (the consensus matrix). They were supported by the consensus matrix, by separation in principal component analysis (PCA), which summarises many correlated proteins by a few

principal components (PCs), and by replication in two validation cohorts[19]. Replication here means that clustering each validation cohort on its own gives three groups that each correspond to one, and only one, discovery subtype. We refer to this criterion as one-to-one replication. It is expected if the same groups exist in every cohort, but it is strong evidence for discrete groups only if data without groups rarely meet it. The clustering algorithm and the number of subtypes were also chosen partly by these criteria, and choosing settings by their output can increase the apparent agreement between cohorts[20]. How often data without subtypes meet such criteria has not been tested.

If the PanNDA subtypes are not discrete, they still describe real differences between patients, and the question becomes which processes order the patients along continuous variation. Bulk brain tissue is a mixture of cell types whose proportions and states vary between individuals, even without disease[21], and between tissue blocks. Protein co-expression networks of the AD brain contain glial, neuronal, synaptic and mitochondrial modules[22,23], astrocyte states differ between individuals[24,25], inhibitory interneurons and excitatory neurons differ in their vulnerability in AD[26–28], and dysfunction of parvalbumin interneurons disrupts network activity in mouse models of AD[29]. Variation among patients could therefore reflect disease stage, a disease-specific process or differences that exist independently of disease. Control brains can separate these possibilities, as variation that is also present in controls is not specific to AD.

Here we asked how well cross-cohort replication distinguishes discrete subtypes from continuous variation, using the PanNDA subtypes as a case with three cohorts, openly available data and a documented clustering pipeline. We also asked which continuous variation orders the patients. We compared the replication in the three PanNDA cohorts with data simulated without subtypes, estimated which subtype separations the replication statistics can detect, and related the main axes of variation to pathology, tissue composition, an independent proteomic cohort[30] and single-nucleus RNA sequencing data[28].

Results

The published subtypes overlap on two principal axes

We analysed the three PanNDA AD cohorts after the outlier removal applied by the PanNDA authors (Methods): discovery (frontal gyrus; 275 cases, 46 controls and 77 early-stage or mild cognitive impairment (MCI) samples), validation 1 (perirhinal cortex; 130 cases, 64 controls) and validation 2 (frontal gyrus; 94 cases, 77 controls). The PanNDA clustering pipeline reproduced all 499 published subtype assignments (Supplementary Table 1).

We asked whether the published subtypes form separate groups in the space of the subtyping proteins. We measured separation as the distances between subtype centroids, the average protein profiles of the subtypes (Methods). Subtypes simulated at the published separation overlapped in the plane of the two leading principal components, PC1 and PC2 (Fig. 1a), and the observed subtypes occupied neighbouring regions of this plane (Fig. 1b, c). Any partition of continuous data occupies adjacent sectors of this plane; the separation in PCA reported by PanNDA[19] therefore also arises without subtypes (Fig. 1a). We named PC1 the severity axis and PC2 the astrocyte-parvalbumin axis (astrocyte-PV axis; PV, parvalbumin), after the proteins at its two ends; in discovery cases, they explained 12.5% and 8.3% of the variance. In every cohort, ST1 lay at the low end of the severity axis and ST3 at the astrocyte end, whereas the positions of ST2 and of the most severe subtype differed between cohorts. The proteins behind the blood-brain barrier transport annotation of ST3 (KCNC1 to KCNC4, SCN4B and SLC1A2) lie at the PV end of the axis, consistent with their lower abundance in ST3. Positions on PC1 and PC2 alone assigned 73% to 88% of cases to their published subtype (Cohen's kappa 0.59 to 0.82; Supplementary Table 2), similar to clusters of data simulated without subtypes (median kappa 0.77 to 0.79).

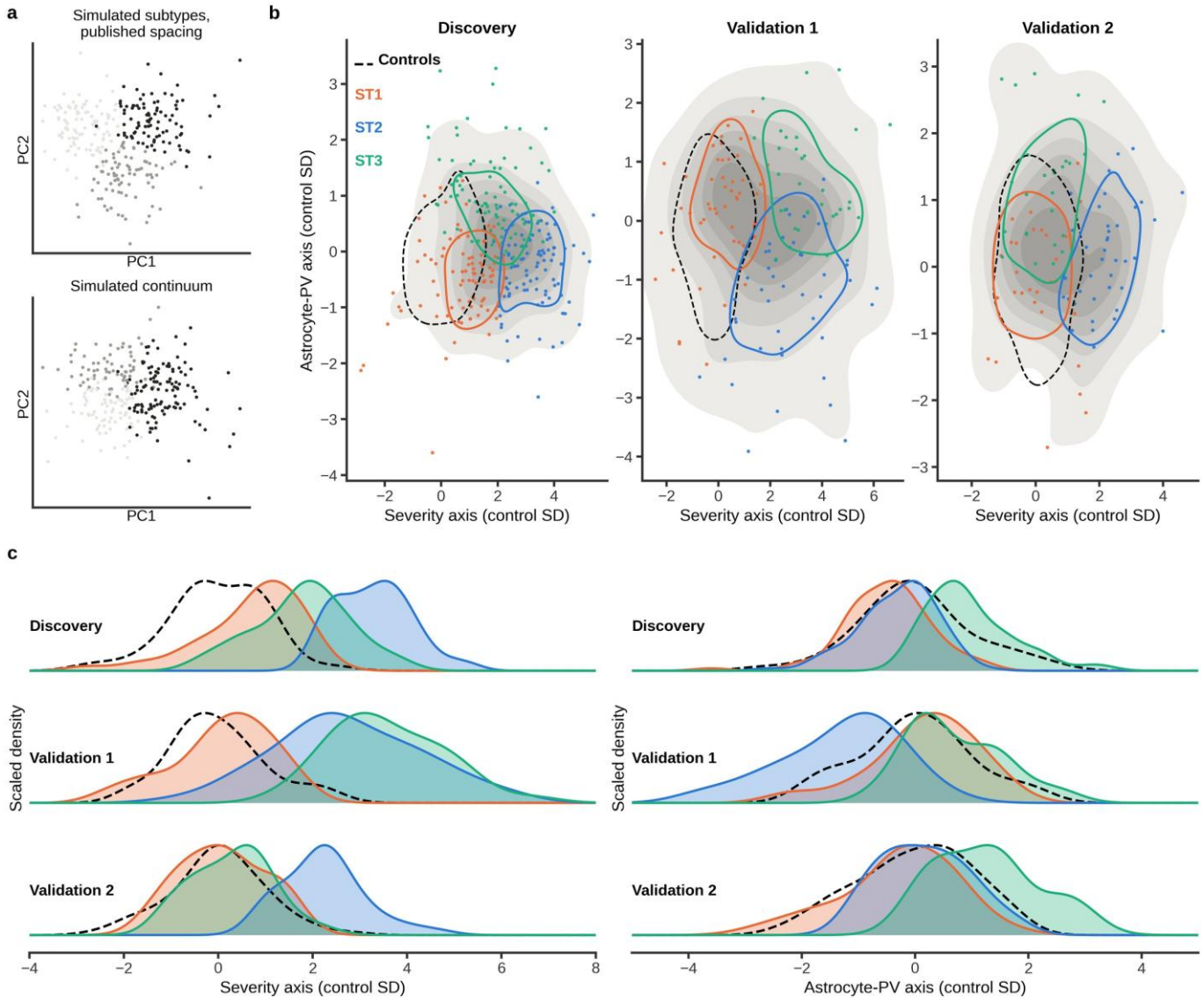


Fig. 1 | Position of Alzheimer's disease cases and controls on the two principal axes of the subtyping proteins. a, One simulated dataset of discovery size (275 cases) with subtypes at the published separation (top) and one without subtypes (bottom), both built on the residual variation of the observed data, clustered with the PanNDA pipeline and projected on principal components 1 and 2 (PC1, PC2). Grey shades, pipeline clusters; component scores in arbitrary units. **b**, Cases and controls on the severity and astrocyte-parvalbumin (astrocyte-PV) axes in discovery (*n* = 275 cases, 46 controls), validation 1 (*n* = 130, 64) and validation 2 (*n* = 94, 77). Points, cases coloured by published subtype (ST1 orange, ST2 blue, ST3 green); grey shading, case density; coloured contours, 40% of the maximum density of each subtype; dashed contour, controls (35%). Axes in control standard deviations (SD). **c**, Density of controls (dashed) and of each subtype along each axis, scaled to equal height.

Data without subtypes meet the replication criterion

We next tested whether these overlapping subtypes replicate across cohorts more often than clusters of continuous data. Clusters of continuous data can replicate when the cohorts share the same correlation structure: clustering then cuts each cohort at similar places, and the resulting clusters have similar centroids (Fig. 2a). We simulated three-cohort datasets without subtypes and with inserted subtypes, analysed them with the PanNDA pipeline and compared them with the observed cohorts (Fig. 2b to d; models in Supplementary Table 3). Data without subtypes kept the correlation structure of the subtyping proteins, either shared by all cohorts (the shared continuum) or specific to each cohort, and gave consensus matrices similar to those of the observed cohorts (Fig. 3a, b). The proportion of ambiguous clustering (PAC), the proportion of sample pairs clustered together in some resamplings and apart in others, was not lower in the observed data than in the shared continuum ($P = 0.21$ to 0.33; Fig. 3c; Supplementary Fig. 1c), whereas subtypes at the published separation gave a PAC at or below the observed value in 62% to 98% of datasets.

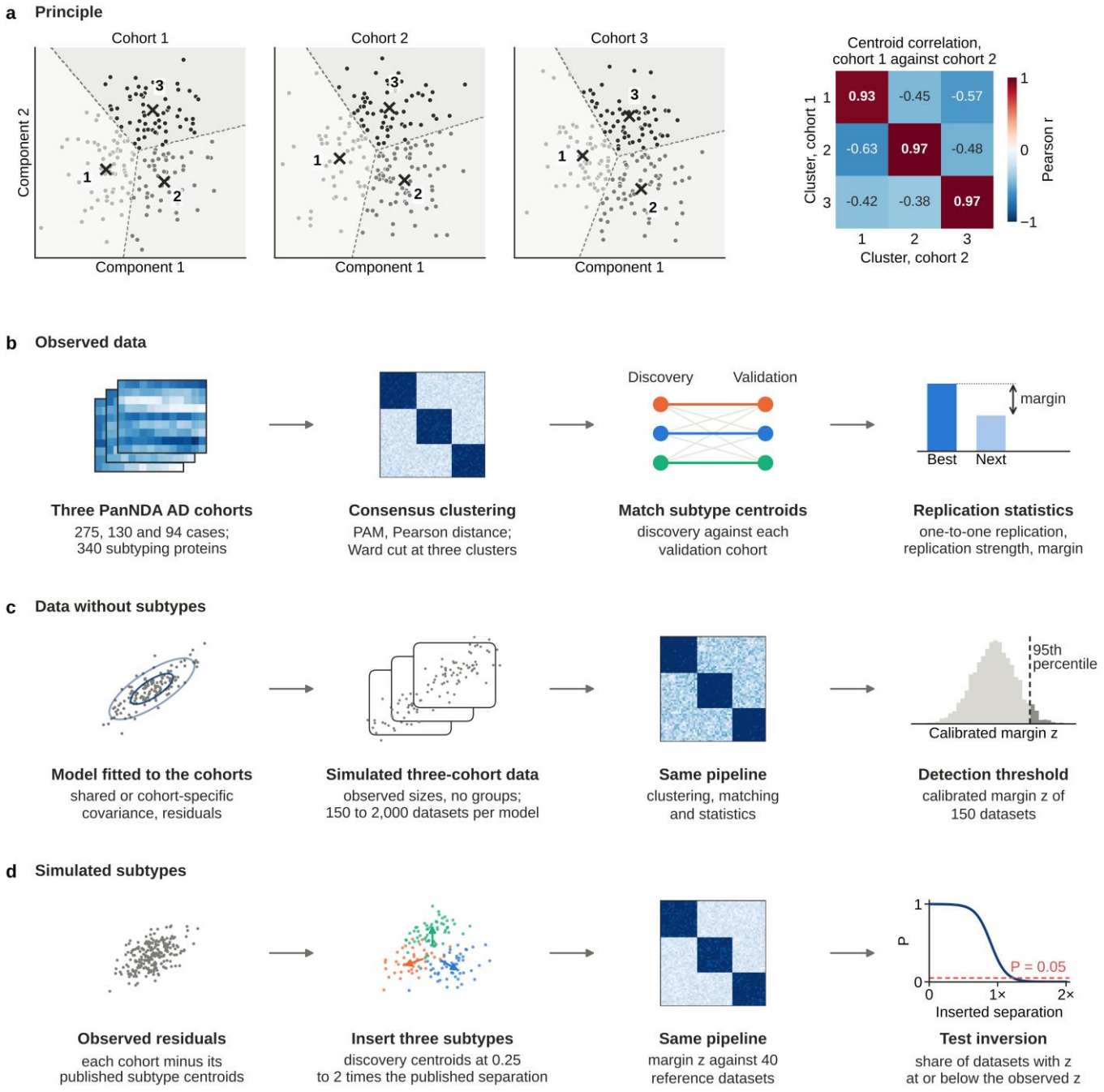


Fig. 2 | Principle of the test and study design. a, Three cohorts of 180 samples simulated from one continuous 40-dimensional Gaussian distribution without groups, divided by k-means into three clusters and projected onto the two leading principal components; crosses, cluster centroids; dashed lines, nearest-centroid boundaries. Clusters of cohorts 2 and 3 are numbered by their most correlated cluster in cohort 1. Right, Pearson correlation between the cluster centroids of cohorts 1 and 2: each cluster matches one counterpart, although no groups exist. **b**, Observed analysis: consensus clustering of the three PanNDA AD cohorts (499 cases) with partitioning around medoids (PAM) on Pearson distance, 100 resamplings of 90% of samples and a Ward cut at three clusters, and matching of discovery and validation subtype

centroids by Pearson correlation. **c**, Data without subtypes: models fitted to the observed cohorts (Supplementary Table 3) generated three-cohort datasets of the observed sizes, analysed in the same way; 150 datasets of the shared continuum gave the detection threshold (95th percentile) for the margin z-score. **d**, Simulated subtypes: three subtypes with the discovery centroids and published proportions inserted into the residuals of each observed cohort at 0.25 to 2 times the published separation; for each separation, the proportion of datasets with a z-score at or below the observed value gives the test-inversion P value.

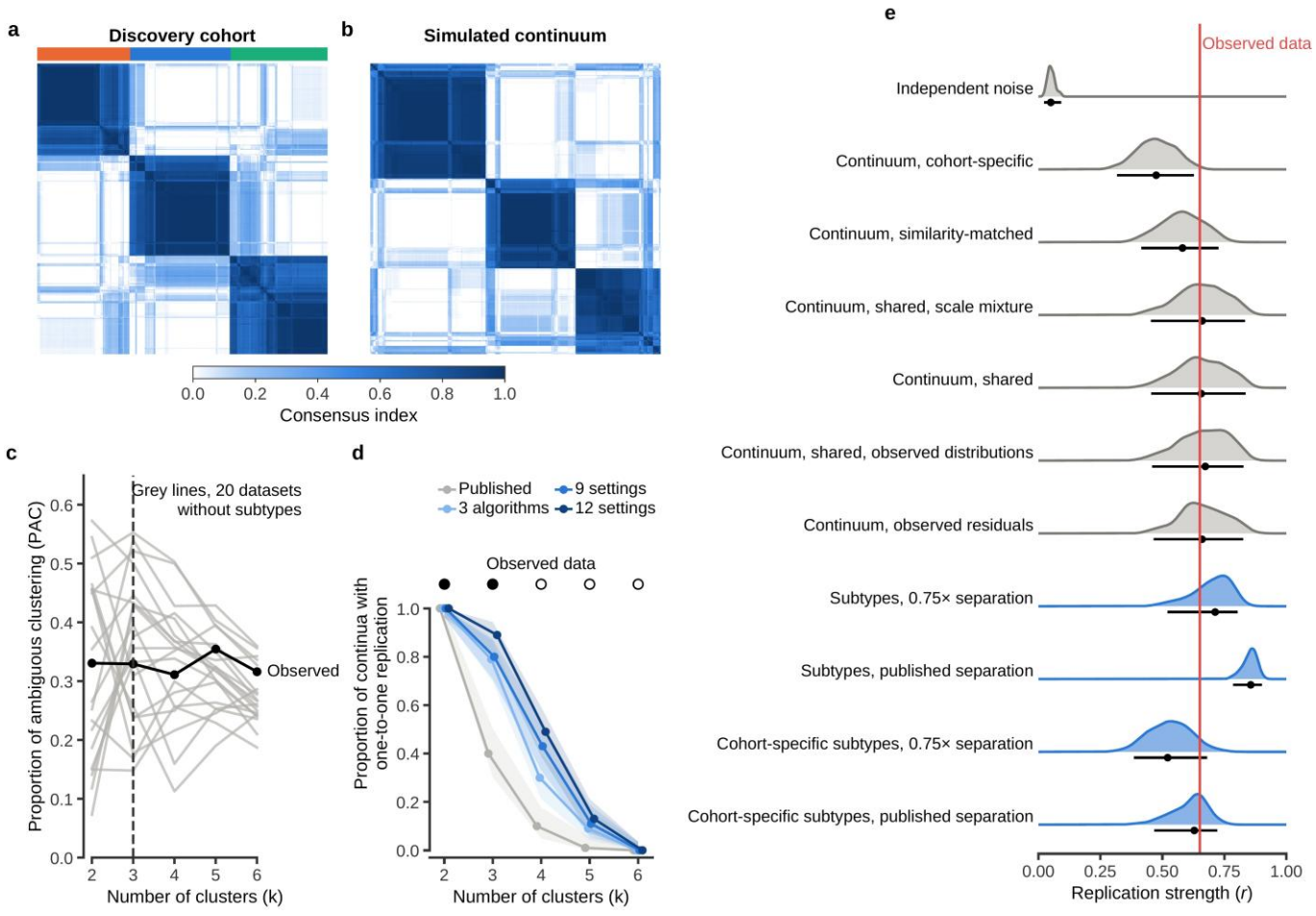


Fig. 3 | Consensus clustering and cross-cohort replication in observed data and in data without subtypes. a, Consensus matrix of the discovery cohort at three clusters; ribbon, published subtype (ST1 orange, ST2 blue, ST3 green). **b**, Consensus matrix of one discovery-sized dataset of the shared continuum. Matrices in a and b used 1,000 resamplings for display. **c**, Proportion of ambiguous clustering (PAC) against the number of clusters (k) in the discovery cohort (black) and in 20 datasets of the shared continuum (grey). **d**, Proportion of 100 shared-continuum datasets in which at least one setting of a family gives one-to-one replication in both validation cohorts: the published setting; the three algorithms tested by PanNDA (PAM, k-means and hierarchical clustering with Ward linkage); the nine settings of the PanNDA script; and all twelve settings (Methods). Bands, 95% Clopper-Pearson intervals. Circles above the panel,

observed data: filled, at least one of the twelve settings gives one-to-one replication; open, no setting does. **e**, Replication strength (mean correlation of matched centroids) in data without subtypes (grey; 150 to 2,000 datasets per model) and with simulated subtypes (blue; 50 datasets each). Dot, median; bar, central 95% range; red line, observed data.

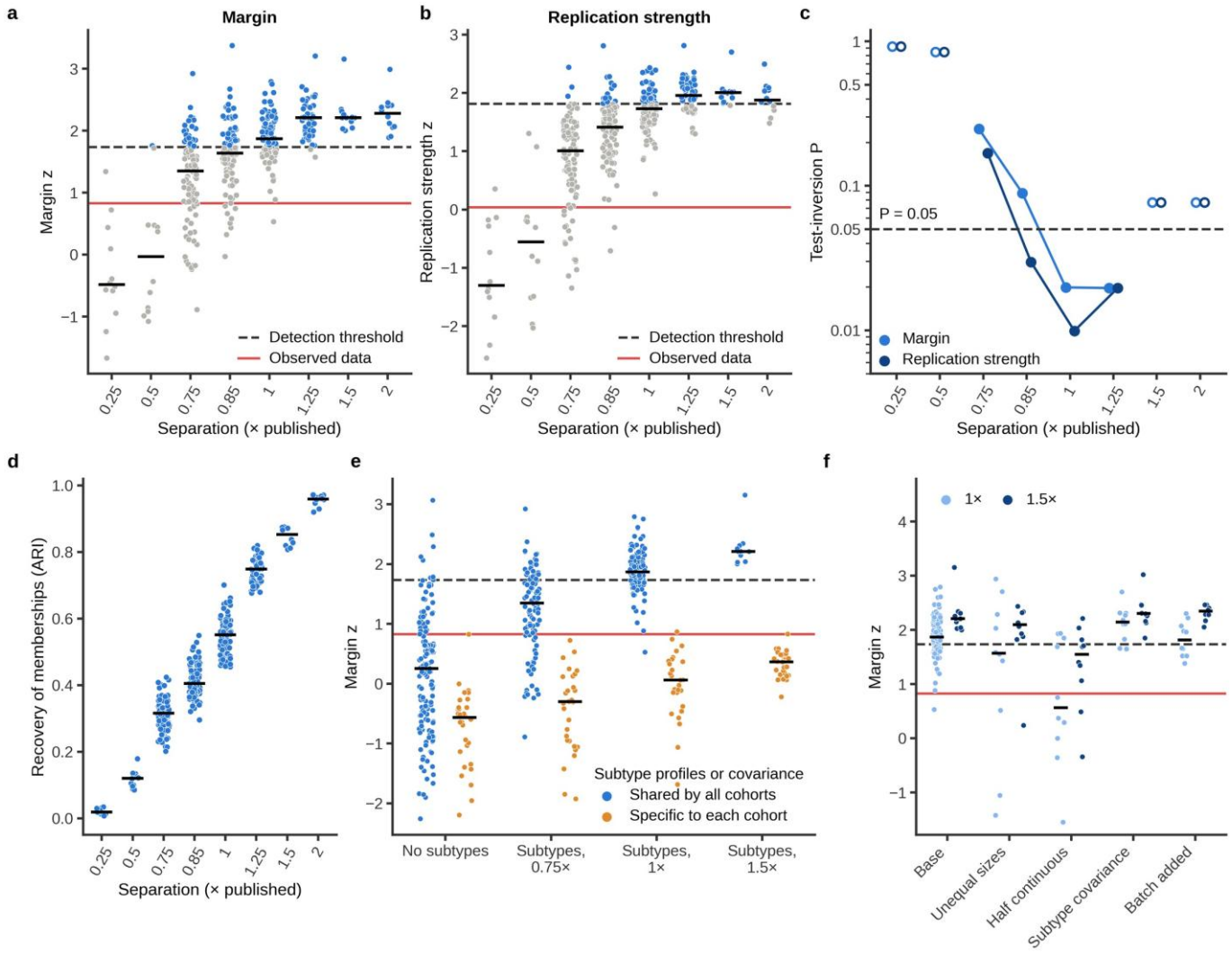
With the published setting, the shared continuum met the one-to-one criterion in 37% of 2,000 datasets (95% confidence interval 35% to 39%; Supplementary Fig. 2). A variant of the criterion that follows the PanNDA similarity dendrogram more closely was met in 63% of 200 datasets. PanNDA tested hierarchical clustering, k-means and partitioning around medoids (PAM) and selected the algorithm that gave the best clusters[19]. If the algorithm were chosen among these three for replication alone, at least one would meet the criterion in 79% of datasets (70% to 87%; Fig. 3d). The observed data met the criterion only with PAM on Pearson distance (3 of 12 settings) and in 23 of 30 random subsamples of 90% of the cases (Supplementary Table 3). Subtypes with shared profiles met the criterion in 50 of 50 datasets at the published separation and in 38 of 50 at 0.75 times. Meeting the criterion is therefore at most 2.7 times more likely under these subtypes than under the shared continuum (Methods). Other models without subtypes met it in 20% to 44% of datasets.

As the one-to-one criterion records only whether clusters match, we also measured how closely they matched. Replication strength is the mean correlation between the centroids of matched subtypes in the discovery and each validation cohort. The margin, our primary statistic, is the mean amount by which each subtype correlates more strongly with its counterpart than with the next-best subtype. We expressed the margin of each dataset as a z-score relative to 40 datasets simulated without subtypes from a model refitted to that dataset (Methods). The observed margin gave $z = 0.83$, below the detection threshold of 1.74, the 95th percentile of 150 datasets without subtypes (calibrated $P = 0.27$; Supplementary Fig. 1e; Supplementary Table 4). It was also typical of continuous data built on the observed residuals, the data minus the published subtype centroids ($P = 0.16$). Replication strength (0.651) was typical of the shared continuum and its variants (medians 0.657 to 0.673; Fig. 3e).

Weaker subtypes cannot be excluded

The observed replication was therefore consistent with continuous variation. This result is informative only if real subtypes would have been detected[31]. We therefore inserted three subtypes, with the published proportions and discovery centroids, into the residuals of each observed cohort at multiples of the published separation (Fig. 4). At the published separation, the inserted subtypes explained about 11% of the variance. We then asked how often simulated subtypes of a given separation gave a margin z-score at or below the observed value (test inversion; primary analysis). One of 100 datasets at the published separation gave a z-score this low (P = 0.020), compared with 8 of 100 at 0.85 times (P = 0.089) and 24 of 100 at 0.75 times (P = 0.25; Fig. 4a, c; Supplementary Table 5). By the margin, the data exclude the published separation but not 0.85 times it; by replication strength, they also exclude 0.85 times (P = 0.030; Fig. 4b, c). Clustering itself inflated separation: clusters cut from the shared continuum were 87% to 113% as far apart as the observed subtypes, and clusters as far apart as the observed ones arose from simulated subtypes at about 0.84 times the published separation (0.79 to 1.03 across cohorts). The margin exceeded the detection threshold in 27 of 100 datasets at 0.75 times and in 70 of 100 at the published separation (Fig. 4a). Clustering recovered the simulated memberships only partly (median adjusted Rand index (ARI) 0.55 at the published separation; Fig. 4d).

When each cohort carried its own subtype profiles, the margin exceeded the threshold in none of 30 datasets at any separation tested (Fig. 4e), although clustering recovered these subtypes within each cohort. The cross-cohort statistics therefore cannot distinguish a continuum from subtypes whose profiles differ between cohorts. Subtypes of strongly unequal size, or with half of the cases drawn from a continuous distribution, were also often missed (Fig. 4f). Subtype assignments were not associated with tandem mass tag (TMT) batch (P = 0.24 to 0.72), although nearest neighbours shared a batch more often than expected by chance, most strongly in validation 2 (Supplementary Fig. 1a, b; Supplementary Table 6).



193

194 **Fig. 4 | Detection of simulated subtypes across cohorts.** Subtypes were inserted as in Fig. 2d, and each dataset was
195 compared with 40 reference datasets from a model without subtypes refitted to it. **a**, Margin z-score of each simulated
196 dataset with the same subtype profiles in every cohort (blue, above the detection threshold; bar, median). Dashed line,
197 detection threshold (1.74); red line, observed data (z = 0.83). Datasets per separation: 12 at 0.25, 0.5, 1.5 and 2 times;
198 100 at 0.75, 0.85 and 1 times; 50 at 1.25 times. **b**, The same for replication strength (threshold 1.81). **c**, Test-inversion P
199 value for the margin and replication strength; dashed line, P = 0.05; open symbols, separations with 12 datasets. **d**,
200 Recovery of simulated memberships (adjusted Rand index (ARI), mean over cohorts). **e**, Margin z with subtype profiles or
201 covariance shared by all cohorts (blue) or specific to each cohort (orange), for continuous data and for subtypes at 0.75,
202 1 and 1.5 times the published separation. **f**, Margin z for alternative structures at 1 and 1.5 times the published
203 separation: unequal subtype sizes (50%, 35% and 15%), half of the cases continuous, subtype-specific covariance and
204 added TMT batch structure.

Subtypes at the published separation replicated more strongly (median replication strength 0.861) than the observed data (0.651) and were more tightly grouped (in-group proportion, the proportion of samples whose nearest neighbour belongs to the same cluster; median 0.791, compared with 0.642). The observed in-group proportion was lower than in all 150 datasets without subtypes and in 565 of 568 datasets with subtypes (calibrated z-scores; Supplementary Fig. 1d). No simulation model therefore reproduced the local neighbourhood structure of the observed data, to which batch structure contributes (median 0.692 with batch added; Supplementary Table 3), and our inferences on separation are conditional on these models. The gap statistic[32] and Hartigan's dip test[33] did not detect subtypes in the observed data, but they also missed 37% and all, respectively, of the simulated datasets with subtypes at the published separation (Supplementary Table 4).

Transfer of subtype assignments between cohorts resembles that of clusters of continuous data

We next asked whether subtype assignments transfer between cohorts, as required for assigning new patients. The published subtyping proteins were selected with case-control statistics from all three cohorts. We therefore selected 340 proteins with discovery statistics alone, re-clustered the discovery cases and derived both axes in discovery (Supplementary Table 7). The projected axes matched principal components of the validation cohorts (|r| 0.73 to 0.97; Fig. 5a, b). By contrast, assigning each validation case to the discovery cluster with the most similar centroid agreed poorly with clusters obtained from each validation cohort on its own (ARI[34] 0.25 and 0.37; Fig. 5c). A logistic regression classifier, the approach used by PanNDA[19], agreed similarly poorly (0.22 and 0.38). With the published proteins, the nearest-centroid agreement equalled the median of data without subtypes and was lower than in all 100 datasets with subtypes at the published separation. Axis weights also agreed between cohorts in data with a shared covariance (mean |r| across cohort pairs, median 0.89 and 0.82 for the two axes, compared with 0.71 and 0.57 in the observed data). Replication of the axes, like that of subtypes, therefore reflects a shared correlation structure (Supplementary Table 3).

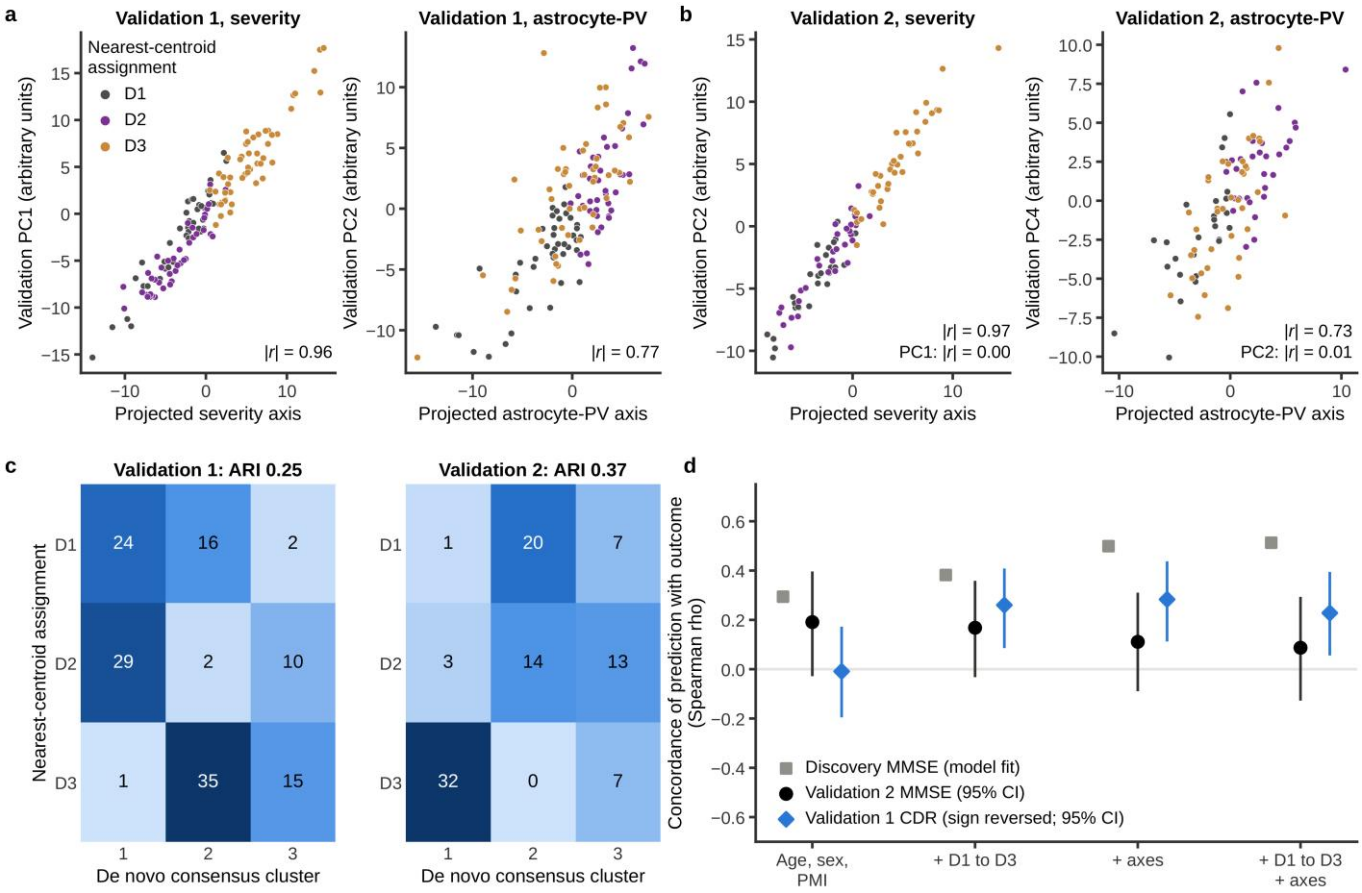


Fig. 5 | Transfer of discovery-derived subtypes and axes to the validation cohorts, and prediction of cognition. Proteins were selected and subtypes and axes derived in discovery cases only (Methods). Each validation case was assigned to the discovery cluster with the most correlated centroid (nearest-centroid assignment) and projected onto the discovery axis weights. **a**, **b**, Projected severity and astrocyte-parvalbumin (astrocyte-PV) axis scores against the best-matching of principal components (PCs) 1 to 4 in validation 1 (*n* = 134 cases) and validation 2 (*n* = 97). Points, cases coloured by assignment to discovery clusters D1 to D3; component scores in arbitrary units; numbers, absolute correlation (|r|) with the plotted component and, in validation 2, with the component of the same rank. **c**, Nearest-centroid assignments against the de novo consensus clusters of each cohort; numbers, cases; ARI, adjusted Rand index. **d**, Spearman correlation between predicted and observed outcome for four models trained on discovery Mini-Mental State Examination (MMSE) scores: age, sex and post-mortem interval (PMI), with D2 and D3 indicators, with the two axis scores, or with both. Grey squares, discovery fit; black circles, validation 2 MMSE; blue diamonds, validation 1 Clinical Dementia Rating (CDR, sign reversed); bars, 95% bootstrap intervals.

We then tested whether subtypes or axes improve the prediction of cognition. Adding subtypes or axes to a model of age, sex and post-mortem interval (PMI) did not clearly improve the prediction of Mini-Mental State Examination (MMSE) scores in validation 2 (Fig. 5d). In validation 1, predictions from

subtypes and from axes were concordant with the Clinical Dementia Rating (CDR; rho −0.26 and −0.28; negative values indicate concordance for the CDR), and both were more concordant than the baseline model (paired bootstrap differences −0.27, 95% interval −0.38 to −0.16, and −0.29, −0.44 to −0.15; Supplementary Table 7).

One axis follows disease severity

If the subtypes divide continuous variation, the axes describe the patients more directly than the labels, and we therefore examined both axes. The severity axis increased with tangle burden in all three cohorts (Spearman rho 0.25 to 0.59) and with plaque burden and cognitive impairment in discovery and validation 1 (Supplementary Fig. 3; Supplementary Table 8). A severity axis is expected, as the subtyping proteins were selected as differences between AD and control brains. Its relation to tangles and Braak stage was stronger than that of at least 96% of signatures built from random sets of 340 differentially abundant proteins, but in validation 2 it was not stronger than that of the leading principal components of these random sets. The axis is therefore close to the general direction of difference between AD and control brains. Cases lay 1.03 to 1.96 control standard deviations (SD) above controls, with MCI samples in between (Supplementary Table 2), and the protein weights of both axes were similar between cohorts (r = 0.48 to 0.81; Supplementary Fig. 4a, b). To test whether subtype assignment depends on severity, we removed the tangle score from each protein and re-clustered. Removing even a permuted, uninformative tangle score changed up to 30% of assignments; removing the observed tangle score changed more assignments than this in validation 1 (P = 0.01), and less clearly in discovery (P = 0.09) and validation 2 (P = 0.06; Supplementary Fig. 3b). Subtype assignment therefore depends partly on severity and is sensitive to small changes in the data.

The astrocyte-PV axis contrasts astrocyte and parvalbumin-interneuron proteins

The astrocyte-PV axis is less familiar. At one end, samples had higher levels of glial fibrillary acidic protein (GFAP), the hyaluronan receptor CD44 and vimentin. These proteins increase in reactive astrocytes[35,36], and in the normal human cortex CD44 marks subpial interlaminar, white-matter fibrous and large-vessel perivascular astrocytes[36]. The same samples had lower levels of homeostatic astrocyte

proteins (SLC1A2, SLC1A3, glutamine synthetase and ATP1A2; Fig. 6a; Supplementary Table 9). At the other end, samples had higher levels of channel subunits of fast-spiking interneurons (KCNC1, KCNC2, SCN4B and SCN1A), parvalbumin, neurofilaments, aggrecan and the perineuronal net (PNN) link proteins HAPLN1 and HAPLN4[37,38]; in the neocortex, PNNs surround mainly PV interneurons. A homeostatic astrocyte module (these four proteins and ALDOC) correlated positively with the PV proteins (rho 0.31 to 0.69). The name of the axis therefore describes proteins, not established changes in cell number or state.

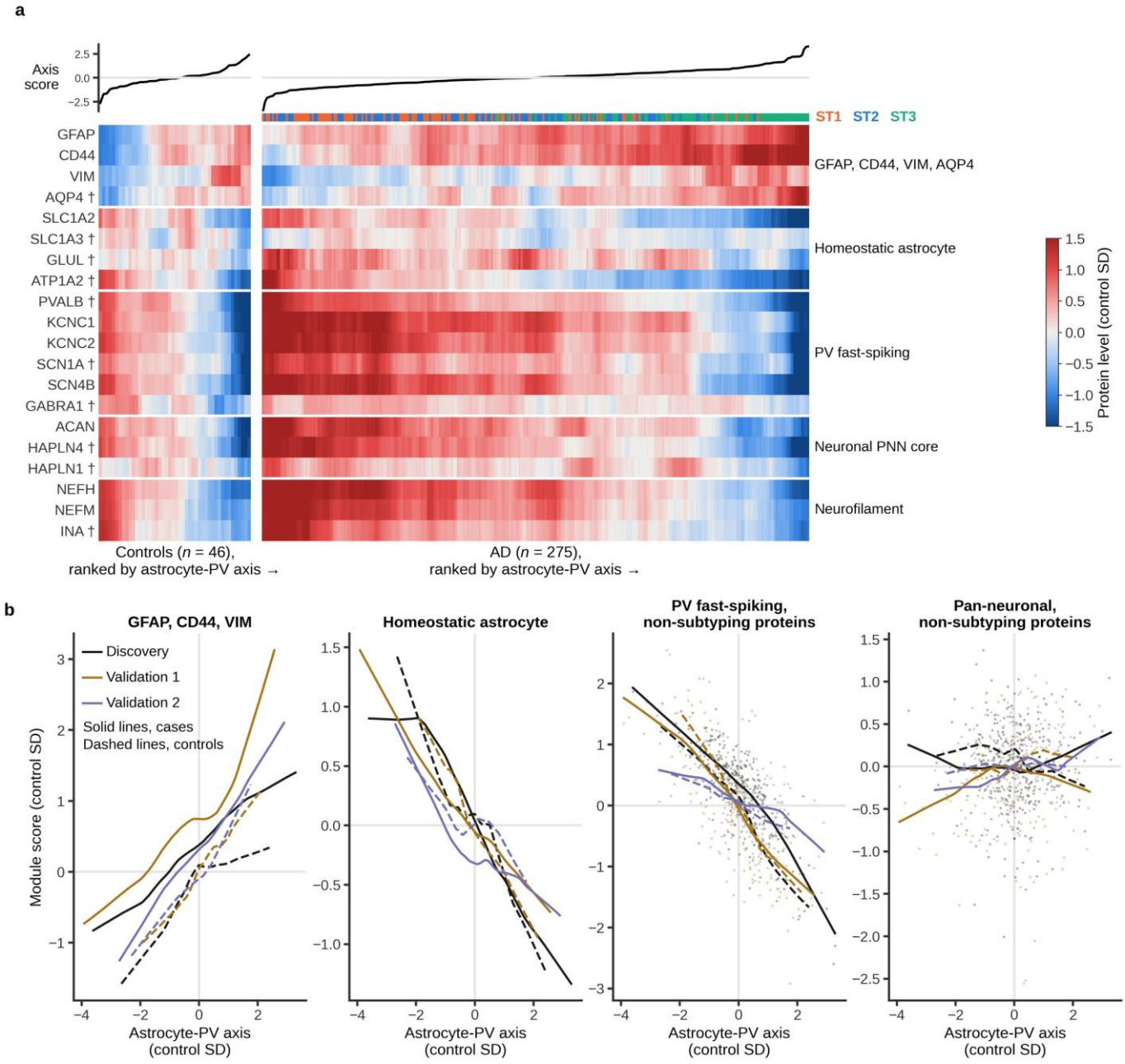


Fig. 6 | Proteins ordered along the astrocyte-parvalbumin axis in control brains and in Alzheimer's disease. a, Discovery controls (*n* = 46, left) and AD cases (*n* = 275, right), each ordered by axis score (top trace). Rows, marker proteins in five blocks: GFAP, CD44, vimentin (VIM) and aquaporin 4 (AQP4); homeostatic astrocyte; PV fast-spiking; neuronal perineuronal-net core; and neurofilament. †, not among the 340 subtyping proteins. Values are moving averages over 9 (controls) or 15 (cases) neighbouring samples, in control SD. Ribbon, published subtype of each case. **b**, Locally weighted regression (LOWESS) fits of four module scores against the axis score in discovery (black), validation 1 (brown) and validation 2 (purple) cases (solid) and controls (dashed); in the two right-hand panels, points show individual samples. The PV fast-spiking (PVALB, SCN1A, GABRA1) and pan-neuronal (SNAP25, SYT1, STMN2, RBFOX3, SYP, ENO2, GAP43) modules use only proteins outside the 340 subtyping proteins.

We reasoned that an axis caused by AD would be weak or absent in control brains. In controls alone, however, the leading principal component matched the case weights (r = 0.61 to 0.77; 95th percentile of random weights 0.15), more closely than any of case components 3 to 10 in discovery and validation 1 (0.39 and 0.49) but not in validation 2 (0.71). In controls, the axis was not related to plaque or tangle score, Braak stage or amyloid-β abundance. The component was also present in controls at Braak stages 0 to II in validations 1 and 2 (r = 0.79 and 0.60); in discovery, which had 12 such controls, it appeared as component 3 (r = 0.42; Supplementary Tables 2 and 10). The axis is therefore present in control brains, including those at Braak stages 0 to II, although many controls had some AD pathology (Braak stage III or IV in 34 of 46 discovery, 16 of 63 validation 1 and 32 of 76 validation 2 controls with a recorded stage). The relation of the axis to cognition was inconsistent across cohorts (Supplementary Fig. 4c).

After adjustment for a pan-neuronal score, a PV score built from proteins outside the subtyping proteins (PVALB, SCN1A and GABRA1) was lower at the astrocyte end in controls and cases of every cohort (partial rho −0.34 to −0.88; Fig. 6b). PVALB alone, however, was not related to the axis in validation 2 (−0.19 in controls, −0.16 in cases), and markers of other interneuron classes were higher at the astrocyte end. Several PV-end proteins are also expressed by large pyramidal neurons, and PanNDA had corrected the data for the proportions of neurons and four other cell classes[19]; a contribution of neuronal composition is therefore possible.

Cases and controls had a similar spread along the axis. Cases were shifted towards the astrocyte end only in validation 2 (0.33 control SD; P = 0.042; Benjamini-Hochberg false discovery rate q = 0.14; 0.39 control SD after adjustment for age and sex, P = 0.021, q = 0.12; Supplementary Table 10). A simpler contrast score, GFAP, CD44 and vimentin minus PV and PNN proteins, was higher in cases in validation 1 and validation 2 (0.37 and 0.38 control SD; q = 0.019 for both) but not in discovery (0.02 control SD; Supplementary Table 2). In discovery, PV and PNN proteins (0.54 control SD) and neurofilaments (0.51) were higher in cases than in controls (Supplementary Table 11), possibly a result of the PanNDA cell-class correction. GFAP and CD44 were higher in cases in all three cohorts, consistent with increased GFAP immunoreactivity in AD[39]. After adjustment for severity, ST3 had higher GFAP and CD44 and lower homeostatic astrocyte, PV and PNN proteins than other cases (Supplementary Fig. 5); these differences are partly circular, as the subtype was defined from these proteins[11]. A mitochondrial network module[23] was lower at the astrocyte end; after adjustment for PV and pan-neuronal scores, this association largely disappeared in controls and in the validation cases but remained in discovery cases (partial rho −0.42; Supplementary Table 12).

Covariates and batch do not explain the axis, but tissue composition contributes

An axis present in control brains could reflect properties of the tissue samples rather than disease, and we tested this possibility. PanNDA had regressed age, sex and PMI out of the data, and further adjustment for these variables and for brain pH did not alter the comparison of cases with controls (Supplementary Fig. 6a, b). The axis was weakly related to PMI in validation 2 cases (rho −0.21, q = 0.86) and to a protein proxy of pH in discovery (rho −0.47 in controls and −0.23 in cases), with the opposite sign in validation 1 cases, in which the proxy was derived (rho 0.23), but not to measured pH or a hypoxia-response score (Supplementary Table 10). Apolipoprotein E (APOE) ε4 alleles and recorded Lewy body and infarct co-pathology were not related to the axis after correction for multiple testing. The abundance of TAR DNA-binding protein 43 (TDP-43) was related to the axis only in validation 1 (rho −0.56 in controls, −0.38 in cases). Discovery cases without a recorded brain region lay further towards the astrocyte end (P = 0.011), but excluding them left both axes unchanged. ST3 remained separated

after adjustment for dissection, cell-class and network-module scores and for TMT batch (Supplementary Fig. 6c to e; Supplementary Table 6).

A tissue block can contain more or less of each cortical layer. In an exploratory analysis, we scored each sample for layer and white-matter marker genes outside the subtyping proteins and for meningeal marker genes (Methods). In discovery and validation 1, samples at the astrocyte end had more layer 1, 2 and 6 signal and less layer 4 signal; in validation 2, only the layer 4 association was clear (Supplementary Table 13). With interneuron-enriched genes removed, layer scores explained 59%, 55% and 21% of the variance of the axis in cases, compared with medians of 27%, 29% and 13% for random gene sets (P = 0.003, 0.002 and 0.10). An upper-layer minus deep-layer score was not clearly related to the axis (rho 0.12, 0.15 and −0.15; Supplementary Figs. 7 and 8; Supplementary Table 14). When the PV and layer 4 scores were adjusted for each other, the PV score remained related to the axis (partial rho −0.70, −0.76 and −0.27), whereas the layer 4 score was related only in discovery (−0.24, compared with −0.05 and −0.08 in the validation cohorts). The layer 4 term was also negative in hippocampal samples of the independent cohort described below, although the hippocampus lacks a layer 4 (Supplementary Fig. 8b). The layer 1 set consists largely of vascular and meningeal genes, and meningeal proteins were higher at the astrocyte end in discovery and validation 1 (Supplementary Fig. 9), suggesting that the proportion of pial surface in a tissue block contributes. Tissue composition and a PV-associated signal therefore both contribute to the axis, in proportions that these data cannot separate.

Independent data do not resolve the origin of the astrocyte-PV axis

We next asked whether independent data could separate these contributions. We projected both axes onto an independent proteomic cohort[30] with three brain regions from largely the same donors (Supplementary Figs. 10 to 12; Supplementary Table 15). The severity axis separated sporadic AD from controls in all three regions (standardised mean difference, Hedges' g, 1.66 to 2.67), whereas donors with high AD pathology but no dementia did not differ clearly from controls (g −0.10 to 0.70). Sporadic AD was also shifted towards the astrocyte end of the astrocyte-PV axis, but this difference was largely explained by severity and by the amyloid precursor protein (APP), and in one region by brain bank

(Supplementary Fig. 12). These data had not been corrected for cell-class proportions, and the axis was not clearly recovered in them (Supplementary Fig. 13; Supplementary Table 16); a similar correction made a matching component more prominent (|r| 0.29 to 0.45 before correction; 0.55 and 0.48 in two regions after) without creating the case-control difference. In cortex, dementia cases that most resembled ST3 were mostly autosomal dominant AD cases (Supplementary Fig. 14). In single-nucleus RNA sequencing data from the middle temporal gyrus of 69 donors[28], astrocyte and PV interneuron proportions were not inversely related along disease progression (rho 0.09, 95% interval −0.15 to 0.32; Supplementary Fig. 15; Supplementary Table 17), although the 11 most severely affected donors are absent from these data. Neither dataset therefore resolves whether the axis reflects cell numbers, tissue composition or the PanNDA cell-class correction.

Discussion

Replication of subtypes across cohorts is informative only in comparison with the replication expected from continuous data. For the PanNDA subtypes, data without subtypes met the one-to-one criterion in 37% of datasets with the published method, and the observed subtypes matched and transferred across cohorts to a similar extent as clusters of continuous data. Subtypes common to all cohorts and as separated as the published ones were unlikely, but clustering inflates separation, and subtypes at 0.85 times the published separation could not be excluded by the primary statistic. Subtypes whose profiles differ between cohorts would not be detected at all. The patients were ordered largely by disease severity and by an astrocyte-PV axis that is also present in control brains.

Replication of clusters across cohorts is widely used as evidence for disease subtypes, but on its own it does not distinguish groups from continuous variation. Cohorts that share a correlation structure yield matching clusters whether or not groups exist (Fig. 2a), and consensus matrices and separation in principal components can likewise arise from continuous variation. Such evidence becomes informative when replication statistics are compared with data simulated without subtypes, fitted to the same cohorts and analysed with the same pipeline, when the separations that these statistics can detect are estimated by inserting subtypes into the observed residuals, and when the clustering algorithm and number of

subtypes are fixed before replication is assessed. These steps require only the data and pipeline already used for subtyping, and the open release of the PanNDA data and clustering pipeline made them possible here.

These results do not imply that the subtypes are artefacts or that AD lacks molecular heterogeneity. The subtypes summarise differences between patients that recur across cohorts, and part of their biology was consistent: ST1 was the least severe subtype and ST3 was positioned at the astrocyte end in every cohort. PanNDA reported that the subtypes differed from established classifications[19]; in the present analysis, a large part of what separated them was disease severity, consistent with imaging subtypes that reflect disease stage[12] and with the single inferred ordering of AD progression (pseudo-time) in PanNDA[19]. For studies that use the PanNDA labels, this implies adjusting for disease stage, reporting the position of each patient on the two axes alongside the label, and treating individual assignments as uncertain. Simulated subtypes at the published separation were recovered with a median ARI of 0.55, and small changes to the data reassigned up to 30% of cases.

The astrocyte-PV axis combines GFAP, CD44 and vimentin with low homeostatic astrocyte proteins at one end, and PV-associated, PNN and neurofilament proteins at the other. Like proteomic subtype signatures in the cerebrospinal fluid of cognitively unimpaired people[40], it is present in control brains, and astrocyte states change with age[41]. Tissue composition is one likely contributor: CD44 marks subpial, fibrous and perivascular astrocytes[36], vimentin and CD44 are also expressed by vascular, meningeal and microglial cells, and pial and meningeal markers increased towards the astrocyte end. A PV-associated signal also contributes but does not indicate loss of PV interneurons, as the axis was not shifted in AD in two cohorts. Astrocyte states that differ between brain regions[42], in other neurodegenerative disorders[43] and between astrocyte subtypes in AD[44] are further possible contributors, as is the sensitivity of energy-demanding fast-spiking interneurons[45] to terminal hypoxia[46,47]; measured pH and a hypoxia-response score, however, were unrelated to the axis. The PanNDA cell-class correction may also have shaped the contrast, and PanNDA itself noted that bulk proteomics may obscure cell-type-specific signatures[19].

The analyses used the published proteins and pipeline, and the results therefore apply to the reported subtypes. Our simulation models were Gaussian, heavy-tailed (a Gaussian scale mixture) or built from the observed residuals; none reproduced the low in-group proportion of the observed data, and inferences on separation are conditional on these models. A covariance shared by all cohorts makes simulated cohorts more alike than the observed ones, which tends to raise replication rates. The axes, like the subtypes, recurred in data without subtypes and did not improve the prediction of cognition in validation 2.

Methods

Data

We analysed only published, de-identified data; ethical approval and consent are described in the original publications[19,28,30]. We obtained the PanNDA protein abundance tables, sample annotations and differential abundance statistics for the AD discovery, validation 1 and validation 2 cohorts (PanNDA AD cohorts 1 to 3)[19]. The PanNDA authors had normalised these values, regressed out TMT batch, age, sex and PMI, and corrected for the proportions of five cell classes with a constrained regression model[48]; these are the data used for subtyping[19]. The 340 subtyping proteins met the PanNDA selection rule: a combined Benjamini-Hochberg-adjusted P below 0.05 across the three cohorts and an absolute mean log2 fold-change z-score above 2.5. Our case-control comparisons used the 46 normal discovery controls, not the 77 early-stage or MCI samples that PanNDA grouped with controls.

The PanNDA authors provided their clustering script on request. The published PanNDA Methods describe hierarchical clustering, k-means and PAM with Ward linkage (the three-algorithm family), and the script tests PAM and k-means with Pearson or Euclidean distance and Ward, average or complete linkage (the nine-setting family). The PanNDA authors had excluded four, four and three cases as outliers by average-linkage hierarchical clustering and a static tree cut (WGCNA[49]); these cases have no published subtype and were used only in the transfer analysis. Cut heights of 35, 30.5 to 31.0 and 28.5 to 32.0 reproduced the published exclusions (Supplementary Table 1). We used the released subtype assignments (88, 95 and 92 discovery cases in ST1, ST2 and ST3) throughout.

Discovery samples came from the Banner Sun Health Research Institute (frontal gyrus), validation 1 from the Mount Sinai Brain Bank (perirhinal cortex, Brodmann area 36) and validation 2 from the Religious Orders Study and Rush Memory and Aging Project (ROSMAP; frontal gyrus). Brain region was not recorded for 20 discovery cases. Part of the PanNDA data were re-analysed from earlier studies[19,23,50]; validation 2 overlaps with the ROSMAP data from which one of the network module sets[23] was partly derived, and module analyses in validation 2 are therefore not fully independent of the module definitions.

The independent cohort was the data-independent acquisition (DIA) mass spectrometry dataset of the MacCoss laboratory (ProteomeXchange PXD034525)[30]. We used the superior and middle temporal gyrus (SMTG), inferior parietal lobule (IPL) and hippocampus, with donors grouped as controls, high pathology without dementia (HiPath), sporadic AD (sAD) and autosomal dominant AD (adAD), from the Adult Changes in Thought study (ACT), the University of Washington Alzheimer's Disease Research Center (UW ADRC) and the Dominantly Inherited Alzheimer Network (DIAN). We matched donors across regions by condition, age, sex, APOE genotype, PMI, Braak stage group, score of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) and study. Single-nucleus data were the middle temporal gyrus data of the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD)[28]. Cell-class markers came from the Human Protein Atlas[51,52], layer statistics from spatial transcriptomics of human dorsolateral prefrontal cortex (spatialLIBD)[53,54] and network modules from Johnson et al.[22,23].

Clustering and replication

We followed the PanNDA pipeline. We standardised proteins across the retained cases of each cohort and ran consensus clustering[7,8] with PAM on Pearson distance, 100 resamplings of 90% of samples and a final Ward (ward.D2)[55] cut at three clusters, using an implementation that reproduces ConsensusClusterPlus 1.74.0[8]. We used 100 rather than 1,000 resamplings to allow thousands of simulated datasets to be clustered; with 1,000 resamplings or with the authors' random seed, the observed results were unchanged, and with standardisation before outlier removal, all 499 published assignments were reproduced and the replication statistics changed by at most 0.02 (Supplementary Table 1).

For each pair of discovery and validation cohorts, we correlated the subtype centroids (Pearson) and matched them by the permutation that maximised the total correlation. Replication strength was the mean matched correlation. The margin was the difference between each matched correlation and the best unmatched correlation, averaged over subtypes and validation cohorts. One-to-one replication required every matched pair to be the best match in both directions in both validation cohorts. A dendrogram variant clustered the nine subtype signatures of the three cohorts by Ward linkage and required each resulting group to contain one subtype from each cohort. We assessed stability in 30 random 90% subsamples of the cases and analysed three alternative feature sets: one feature per gene, features corrected for TMT batch in discovery and validation 1, and 469 features selected with discovery statistics alone. In-group proportion was computed as described[56] with Pearson distance, and PAC from consensus values strictly between 0.1 and 0.9[9].

The setting search combined four base methods (PAM on Pearson or Euclidean distance, k-means, and hierarchical clustering on Pearson distance) with Ward, average or complete final linkage, for two to six clusters; a setting counted only if every cluster contained at least five samples and 5% of the cohort. We also applied the gap statistic[32] (k-means, k = 1 to 6, 100 reference datasets) and Hartigan's dip test[33] to the scores of components 1 and 2.

Data without subtypes and simulated subtypes

We generated three-cohort datasets with the observed sample sizes under six models without subtypes (Supplementary Table 3): independent Gaussian noise (300 datasets); the Ledoit-Wolf covariance[57] of each cohort (cohort-specific continuum; 1,000); one covariance shared by all cohorts (shared continuum; 2,000); the shared continuum with each protein's observed distribution (500); a scale mixture of the shared continuum that keeps the observed spread of case distances (1,000); and a mixture of the shared (weight 0.3) and cohort-specific models (150). We also inserted continuous offsets into resampled observed residuals (continuum on observed residuals; 150). Replication of the axes was the mean |r| between the protein weights of the same principal component in each pair of cohorts. Simulated subtypes used the discovery centroids in every cohort or each cohort's own centroids. For the one-to-one criterion, the likelihood ratio $P(\text{pass} \mid \text{subtypes})/P(\text{pass} \mid \text{continuum})$ of shared-profile subtypes at the

published separation against the shared continuum was 2.7 (50 of 50, compared with 745 of 2,000), the largest value that the criterion can give against this model.

For calibration, we compared each of 150 shared-continuum datasets with 40 reference datasets simulated from the same model refitted to it, as z = (value − reference mean)/reference standard deviation. The detection threshold was the 95th percentile of these z-scores, and the calibrated P of the observed data was (b + 1)/(B + 1), where b is the number of the B calibration z-scores at or above the observed z. We chose the margin as the primary statistic in a pilot power analysis (25 calibration and 20 simulated-subtype datasets, included in the totals). As sensitivity analyses, we used the median z over 1,000 random selections of 40 reference datasets (calibrated P of the observed margin, 0.23 to 0.45), an alternative threshold from the continuum on observed residuals (1.41) and a reference model with cohort-specific covariance, which proved anti-conservative (Supplementary Tables 3 to 5).

For the detectability analysis, we inserted subtypes at 0.25, 0.5, 0.75, 0.85, 1, 1.25, 1.5 and 2 times the published separation (12 datasets at 0.25, 0.5, 1.5 and 2 times; 100 at 0.75, 0.85 and 1 times; 50 at 1.25 times). Within each cohort, residuals were the observed data minus the published subtype centroids, with rows permuted; memberships were the published assignments permuted within the cohort; and offsets were the published discovery centroids multiplied by the separation factor. Each protein was re-standardised after insertion, and each dataset was scored against 40 reference datasets from a shared-covariance model without subtypes refitted to it. Test inversion compared the observed z with the simulated z-scores, with one-sided P = (k + 1)/(n + 1) for k of n datasets at or below the observed value. The refitted reference model absorbs part of the inserted structure, and z-scores therefore plateau at large separations; we used the threshold-free test inversion as the primary inference. We measured the separation of pipeline clusters as the mean Euclidean distance between centroids, relative to the observed cohort, in datasets with subtypes at 0 to 2 times the published separation and in the shared continuum (Supplementary Table 5).

We simulated alternative structures at 1 and 1.5 times the published separation (10 datasets each): unequal subtype sizes (50%, 35% and 15%), half of the cases drawn from a continuous distribution, subtype-specific residual covariance and added TMT batch profiles; and cohort-specific subtype

centroids at 0.75, 1 and 1.5 times (30 datasets each). Recovery of simulated subtypes was the ARI between simulated memberships and consensus clusters. We assessed batch structure in the observed data against 2,000 permutations (nearest-neighbour batch sharing, variance of each axis explained by batch, and association of subtype with batch). To classify pipeline clusters from principal components, we used multinomial logistic regression with five-fold cross-validation in the observed data and in ten simulated datasets per condition.

Axes of variation

Within each cohort, PCA of the standardised subtyping proteins in cases gave two axes, oriented to increase with tangle score and with GFAP, respectively. For comparison with controls, we centred each protein on its control mean, divided it by its case standard deviation, projected it on the case weights and expressed scores in control SD. We derived axes independently in controls and compared them with the case weights by Pearson correlation, against 1,000 random weight vectors and case components 3 to 10. We estimated case-control differences by linear regression and compared spread by the Brown-Forsythe test. We tested specificity by partial Spearman correlation adjusted for a pan-neuronal score, also with markers outside the 340 subtyping proteins. Gene-level values averaged the features of each gene (keratin entries combined with GFAP), and module scores were means of control-standardised gene-level values (Supplementary Table 18). To test the dependence of subtype assignment on severity, we regressed each protein on tangle score, re-clustered and compared the reassignment with 99 runs using permuted tangle scores. In controls, we related the axis to plaque score, tangle score, Braak stage and amyloid-β abundance by Spearman correlation and repeated the control PCA in controls with low pathology.

Alternative explanations

We related age, PMI and pH (validation 1 only) to the axis by Spearman correlation and sex by Mann-Whitney test, and included them as covariates in linear models comparing cases with controls. Dissection scores used spatialLIBD layer 1 and white-matter markers and meningeal and fibroblast genes; cell-class scores used 30 markers each for neurons, astrocytes, microglia, oligodendrocytes and

endothelial cells. We regressed the subtyping proteins on these scores and repeated the component analysis. To test whether a cell-class correction can create the axis, we applied the same correction to the independent cohort and compared the axis before and after correction. We tested enrichment of network modules[22,23] among the 100 proteins most correlated with each end of the axis by one-sided Fisher exact tests. Layer scores used up to 40 genes most specific to each spatialLIBD layer, excluding the subtyping proteins and marker sets; variants used excitatory-neuron-enriched genes or removed interneuron-enriched genes. We regressed the subtyping proteins and module scores on the layer, white-matter and meningeal scores, repeated the PCA and compared the variance explained with that of 1,000 random gene sets of the same sizes. The gene-wise layer analysis correlated, across genes, the correlation of each gene with the axis and its spatialLIBD layer t-statistic (Supplementary Figs. 7 and 8; Supplementary Table 14). A protein proxy of pH (the 50 proteins most positively minus the 50 most negatively correlated with pH in validation 1) and a hypoxia-response score (BNIP3, PDK1, SLC2A1, LDHA, PGK1, HK2, P4HA1, EGLN3, NDRG1, ERO1A, ANKRD37, ADM, CA9 and ENO1) were related to the axis in each cohort.

Generalisation to validation cohorts and outcome analyses

For the transfer analysis, we selected the 340 proteins with the largest absolute discovery log2 fold-change z-score among those with a discovery false discovery rate below 0.05 and quantified in at least half of the samples of every cohort. Missing values were imputed with the median of each protein within the cases of each cohort. We clustered cases with the same consensus procedure and ordered discovery clusters by mean severity. We assigned each validation case to the discovery cluster with the most correlated centroid and, as in PanNDA, also trained a multinomial logistic regression classifier on 80% of the discovery cases (20 random subsets). Axis scores were projections onto the discovery weights. We measured agreement with the de novo clusters of each cohort by the ARI and agreement of axes by the absolute Pearson correlation with the components 1 to 4 of each cohort.

Outcome models were linear models of MMSE in 220 discovery cases with age, sex and PMI (baseline), plus indicators for discovery clusters D2 and D3, plus the two axis scores, or plus both. We applied the models unchanged to validation 2 (MMSE) and validation 1 (CDR) and measured performance by the

Spearman correlation between predicted and observed scores, with 2,000 paired bootstrap resamples of test cases.

Projection onto the independent cohort

We mapped protein groups to genes, kept the most completely quantified group per gene and retained genes quantified in at least 80% of samples (5,142, 5,237 and 4,379 genes in SMTG, IPL and hippocampus). We applied the PanNDA discovery weights, averaged over the features of each gene, to gene-level values standardised against the controls of each region (196, 195 and 170 of the 270 weighted genes were quantified). As the APP value of this cohort may include amyloid-β peptides, we repeated analyses without APP. We compared groups within the bank that supplied all controls, adjusted the severity-axis comparison of sAD with HiPath donors for Braak stage group, CERAD score and age, and adjusted the astrocyte-PV comparison of sAD with controls for bank and age. We also fitted linear mixed models across regions with a random intercept for each donor. We assigned dementia cases to the nearest PanNDA subtype centroid after removal of named marker genes and tested contrasts against 1,000 permutations of centroid genes. We compared components of the independent data with the axis against 500 random directions, and compared protein changes with NeuroPro, a compilation of published protein changes in AD[58].

Single-nucleus analysis

We took nuclei counts per supertype, donor covariates and compositional results from the SEA-AD middle temporal gyrus data, combining the 170 sequencing libraries into 69 donors by their shared covariates; the 15 SEA-AD donors not included had high (14) or intermediate (1) AD neuropathological change. Astrocyte abundance was the proportion of astrocytes among non-neuronal nuclei, and PV abundance the proportion of Pvalb nuclei among GABAergic nuclei (Supplementary Table 17). We related these proportions to the continuous pseudo-progression score (CPS) and to each other by Spearman and partial Spearman correlation. Compositional effects along CPS were taken from the SEA-AD scCODA[59] results, and differential expression along CPS from the NEBULA[60] results for astrocyte,

Pvalb and chandelier supertypes; we compared the median of each gene programme with 10,000 random gene sets of equal size.

Statistics and reproducibility

Empirical P values were computed as (b + 1)/(B + 1). Proportions of simulated datasets are given with Clopper-Pearson 95% intervals[61]. Group differences in the independent cohort are Hedges' g[62] with bias-corrected and accelerated bootstrap 95% intervals[63] (4,000 resamples), tested by Mann-Whitney test. We used Spearman correlation for clinical variables and module scores, and Pearson correlation for weights and axis scores. The primary analyses were the replication rates of data without subtypes and the calibrated margin with its test inversion. All analyses of the axes were exploratory; Supplementary Data 1 gives Benjamini-Hochberg false discovery rates[64] for them. Tests are two-sided, except comparisons with simulated or permuted data, which are one-sided. We ran the analyses in R 4.5.3 (cluster 2.1.8.2) and Python 3.14.3 on Windows 11; Supplementary Table 19 lists the package versions.

Data availability

All data supporting the findings are in the Supplementary Information and Supplementary Data 1. The analysed data are publicly available: PanNDA protein tables in the supplementary tables of ref. 19 (raw data MassIVE MSV000097496, ProteomeXchange PXD062529), the independent cohort at ProteomeXchange PXD034525, SEA-AD data in the public SEA-AD repository (s3://sea-ad-single-cell-profiling), the Human Protein Atlas single-nucleus data (https://www.proteinatlas.org/about/download), and the spatialLIBD layer statistics, consensus network modules and NeuroPro compilation in the supplementary tables of refs. 22, 23, 53 and 58.

Code availability

The analysis code and data used in this study will be available on Zenodo upon publication.

Acknowledgements

We thank the investigators of the Pan-Neurodegeneration Atlas (PanNDA) for making their proteomic data publicly available and for providing their clustering script and method note, and the MacCoss laboratory and the SEA-AD consortium for making their data publicly available. We thank the developers of R, Bioconductor and Python for maintaining the open-source resources used in this analysis. V.K. acknowledges Sharda University Agra for providing infrastructural facilities.

Funding

This study received no external funding.

Author contributions

Conceptualisation: B.Y. and V.K. Methodology: B.Y. and V.K. Formal analysis and software: B.Y. Visualisation: B.Y. and V.K. Writing (original draft): B.Y. Writing (review and editing): B.Y. and V.K. Supervision and project administration: V.K. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

1. Vogel, J. W. et al. Four distinct trajectories of tau deposition identified in Alzheimer's disease. Nat. Med. 27, 871–881 (2021). https://doi.org/10.1038/s41591-021-01309-6

2. Young, A. L. et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. Nat. Commun. 9, 4273 (2018). https://doi.org/10.1038/s41467-018-05892-0

3. Neff, R. A. et al. Molecular subtyping of Alzheimer's disease using RNA sequencing data reveals novel mechanisms and targets. Sci. Adv. 7, eabb5398 (2021). https://doi.org/10.1126/sciadv.abb5398

4. Higginbotham, L. et al. Unbiased classification of the elderly human brain proteome resolves distinct clinical and pathophysiological subtypes of cognitive impairment. Neurobiol. Dis. 186, 106286 (2023). https://doi.org/10.1016/j.nbd.2023.106286

5. Tijms, B. M. et al. Cerebrospinal fluid proteomics in patients with Alzheimer's disease reveals five molecular subtypes with distinct genetic risk profiles. Nat. Aging 4, 33–47 (2024). https://doi.org/10.1038/s43587-023-00550-7

6. Hou, X. H. et al. Proteomics reveals three molecular subtypes of Alzheimer's disease with distinct progression patterns. Alzheimers Dement. 22, e71106 (2026). https://doi.org/10.1002/alz.71106

7. Monti, S., Tamayo, P., Mesirov, J. & Golub, T. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Mach. Learn. 52, 91–118 (2003). https://doi.org/10.1023/A:1023949509487

8. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 26, 1572–1573 (2010). https://doi.org/10.1093/bioinformatics/btq170

9. Şenbabaoğlu, Y., Michailidis, G. & Li, J. Z. Critical limitations of consensus clustering in class discovery. Sci. Rep. 4, 6207 (2014). https://doi.org/10.1038/srep06207

10. Huang, H., Liu, Y., Yuan, M. & Marron, J. S. Statistical significance of clustering using soft thresholding. J. Comput. Graph. Stat. 24, 975–993 (2015). https://doi.org/10.1080/10618600.2014.948179

11. Gao, L. L., Bien, J. & Witten, D. Selective inference for hierarchical clustering. J. Am. Stat. Assoc. 119, 332–342 (2024). https://doi.org/10.1080/01621459.2022.2116331

12. Poulakis, K. et al. Multi-cohort and longitudinal Bayesian clustering study of stage and subtype in Alzheimer's disease. Nat. Commun. 13, 4566 (2022). https://doi.org/10.1038/s41467-022-32202-6

13. Dinga, R. et al. Evaluating the evidence for biotypes of depression: methodological replication and extension of Drysdale et al. (2017). Neuroimage Clin. 22, 101796 (2019). https://doi.org/10.1016/j.nicl.2019.101796

14. Mohanty, R. et al. Comparison of subtyping methods for neuroimaging studies in Alzheimer's disease: a call for harmonization. Brain Commun. 2, fcaa192 (2020). https://doi.org/10.1093/braincomms/fcaa192

15. Venkatraghavan, V. et al. A large-scale multi-centre study characterising atrophy heterogeneity in Alzheimer's disease. Neuroimage 318, 121381 (2025). https://doi.org/10.1016/j.neuroimage.2025.121381

16. Isella, C. et al. Stromal contribution to the colorectal cancer transcriptome. Nat. Genet. 47, 312–319 (2015). https://doi.org/10.1038/ng.3224

17. Calon, A. et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. Nat. Genet. 47, 320–329 (2015). https://doi.org/10.1038/ng.3225

18. Schwede, M. et al. The impact of stroma admixture on molecular subtypes and prognostic gene signatures in serous ovarian cancer. Cancer Epidemiol. Biomarkers Prev. 29, 509–519 (2020). https://doi.org/10.1158/1055-9965.EPI-18-1359

19. Shrestha, H. K. et al. Pan-neurodegeneration proteomics reveals disease subtypes and molecular signatures. Cell 189, 3124–3143.e15 (2026). https://doi.org/10.1016/j.cell.2026.02.026

20. Ullmann, T., Peschel, S., Finger, P., Müller, C. L. & Boulesteix, A. L. Over-optimism in unsupervised microbiome analysis: insights from network learning and clustering. PLoS Comput. Biol. 19, e1010820 (2023). https://doi.org/10.1371/journal.pcbi.1010820

21. Kelley, K. W., Nakao-Inoue, H., Molofsky, A. V. & Oldham, M. C. Variation among intact tissue samples reveals the core transcriptional features of human CNS cell classes. Nat. Neurosci. 21, 1171–1184 (2018). https://doi.org/10.1038/s41593-018-0216-z

22. Johnson, E. C. B. et al. Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. Nat. Med. 26, 769–780 (2020). https://doi.org/10.1038/s41591-020-0815-6

23. Johnson, E. C. B. et al. Large-scale deep multi-layer analysis of Alzheimer's disease brain reveals strong proteomic disease-related changes not observed at the RNA level. Nat. Neurosci. 25, 213–225 (2022). https://doi.org/10.1038/s41593-021-00999-y

24. Sosunov, A. A. et al. Phenotypic heterogeneity and plasticity of isocortical and hippocampal astrocytes in the human brain. J. Neurosci. 34, 2285–2298 (2014). https://doi.org/10.1523/JNEUROSCI.4037-13.2014

25. Simpson, J. E. et al. Astrocyte phenotype in relation to Alzheimer-type pathology in the ageing brain. Neurobiol. Aging 31, 578–590 (2010). https://doi.org/10.1016/j.neurobiolaging.2008.05.015

26. Hof, P. R. et al. Parvalbumin-immunoreactive neurons in the neocortex are resistant to degeneration in Alzheimer's disease. J. Neuropathol. Exp. Neurol. 50, 451–462 (1991). https://doi.org/10.1097/00005072-199107000-00006

27. Leng, K. et al. Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. Nat. Neurosci. 24, 276–287 (2021). https://doi.org/10.1038/s41593-020-00764-7

28. Gabitto, M. I. et al. Integrated multimodal cell atlas of Alzheimer's disease. Nat. Neurosci. 27, 2366–2383 (2024). https://doi.org/10.1038/s41593-024-01774-5

29. Verret, L. et al. Inhibitory interneuron deficit links altered network activity and cognitive dysfunction in Alzheimer model. Cell 149, 708–721 (2012). https://doi.org/10.1016/j.cell.2012.02.046

30. Merrihew, G. E. et al. A peptide-centric quantitative proteomics dataset for the phenotypic assessment of Alzheimer's disease. Sci. Data 10, 206 (2023). https://doi.org/10.1038/s41597-023-02057-7

31. Dalmaijer, E. S., Nord, C. L. & Astle, D. E. Statistical power for cluster analysis. BMC Bioinformatics 23, 205 (2022). https://doi.org/10.1186/s12859-022-04675-1

32. Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B Stat. Methodol. 63, 411–423 (2001). https://doi.org/10.1111/1467-9868.00293

33. Hartigan, J. A. & Hartigan, P. M. The dip test of unimodality. Ann. Stat. 13, 70–84 (1985). https://doi.org/10.1214/aos/1176346577

34. Hubert, L. & Arabie, P. Comparing partitions. J. Classif. 2, 193–218 (1985). https://doi.org/10.1007/BF01908075

35. Escartin, C. et al. Reactive astrocyte nomenclature, definitions, and future directions. Nat. Neurosci. 24, 312–325 (2021). https://doi.org/10.1038/s41593-020-00783-4

36. Al-Dalahmah, O. et al. The matrix receptor CD44 is present in astrocytes throughout the human central nervous system and accumulates in hypoxia and seizures. Cells 13, 129 (2024). https://doi.org/10.3390/cells13020129

720 37. Bekku, Y. et al. Molecular cloning of Bral2, a novel brain-specific link protein, and immunohistochemical
721 colocalization with brevican in perineuronal nets. Mol. Cell. Neurosci. 24, 148–159 (2003).
722 https://doi.org/10.1016/s1044-7431(03)00133-7

723 38. Carulli, D. et al. Animals lacking link protein have attenuated perineuronal nets and persistent plasticity. Brain
724 133, 2331–2347 (2010). https://doi.org/10.1093/brain/awq145

725 39. Serrano-Pozo, A. et al. Reactive glia not only associates with plaques but also parallels tangles in Alzheimer's
726 disease. Am. J. Pathol. 179, 1373–1384 (2011). https://doi.org/10.1016/j.ajpath.2011.05.047

727 40. Tijms, B. M. et al. CSF proteomic Alzheimer's disease-predictive subtypes in cognitively intact amyloid negative
728 individuals. Proteomes 9, 36 (2021). https://doi.org/10.3390/proteomes9030036

729 41. Habib, N. et al. Disease-associated astrocytes in Alzheimer's disease and aging. Nat. Neurosci. 23, 701–706
730 (2020). https://doi.org/10.1038/s41593-020-0624-8

731 42. Serrano-Pozo, A. et al. Astrocyte transcriptomic changes along the spatiotemporal progression of Alzheimer's
732 disease. Nat. Neurosci. 27, 2384–2400 (2024). https://doi.org/10.1038/s41593-024-01791-4

733 43. Al-Dalahmah, O. et al. Single-nucleus RNA-seq identifies Huntington disease astrocyte states. Acta
734 Neuropathol. Commun. 8, 19 (2020). https://doi.org/10.1186/s40478-020-0880-6

735 44. Sadick, J. S. et al. Astrocytes and oligodendrocytes undergo subtype-specific transcriptional changes in
736 Alzheimer's disease. Neuron 110, 1788–1805.e10 (2022). https://doi.org/10.1016/j.neuron.2022.03.008

737 45. Kann, O., Papageorgiou, I. E. & Draguhn, A. Highly energized inhibitory interneurons are a central element for
738 information processing in cortical networks. J. Cereb. Blood Flow Metab. 34, 1270–1282 (2014).
739 https://doi.org/10.1038/jcbfm.2014.104

740 46. Li, J. Z. et al. Systematic changes in gene expression in postmortem human brains associated with tissue pH
741 and terminal medical conditions. Hum. Mol. Genet. 13, 609–616 (2004). https://doi.org/10.1093/hmg/ddh065

742 47. Atz, M. et al. Methodological considerations for gene expression profiling of human brain. J. Neurosci. Methods
743 163, 295–309 (2007). https://doi.org/10.1016/j.jneumeth.2007.03.022

744 48. Bai, B. et al. Deep multilayer brain proteomics identifies molecular networks in Alzheimer's disease
745 progression. Neuron 105, 975–991.e7 (2020). https://doi.org/10.1016/j.neuron.2019.12.015

746 49. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC
747 Bioinformatics 9, 559 (2008). https://doi.org/10.1186/1471-2105-9-559

50. Coleman, C. et al. Multi-omic atlas of the parahippocampal gyrus in Alzheimer's disease. Sci. Data 10, 602 (2023). https://doi.org/10.1038/s41597-023-02507-2

51. Siletti, K. et al. Transcriptomic diversity of cell types across the adult human brain. Science 382, eadd7046 (2023). https://doi.org/10.1126/science.add7046

52. Sjöstedt, E. et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. Science 367, eaay5947 (2020). https://doi.org/10.1126/science.aay5947

53. Maynard, K. R. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat. Neurosci. 24, 425–436 (2021). https://doi.org/10.1038/s41593-020-00787-0

54. Pardo, B. et al. spatialLIBD: an R/Bioconductor package to visualize spatially-resolved transcriptomics data. BMC Genomics 23, 434 (2022). https://doi.org/10.1186/s12864-022-08601-w

55. Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? J. Classif. 31, 274–295 (2014). https://doi.org/10.1007/s00357-014-9161-z

56. Kapp, A. V. & Tibshirani, R. Are clusters found in one dataset present in another dataset? Biostatistics 8, 9–31 (2007). https://doi.org/10.1093/biostatistics/kxj029

57. Ledoit, O. & Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. J. Multivar. Anal. 88, 365–411 (2004). https://doi.org/10.1016/S0047-259X(03)00096-4

58. Askenazi, M. et al. Compilation of reported protein changes in the brain in Alzheimer's disease. Nat. Commun. 14, 4466 (2023). https://doi.org/10.1038/s41467-023-40208-x

59. Büttner, M., Ostner, J., Müller, C. L., Theis, F. J. & Schubert, B. scCODA is a Bayesian model for compositional single-cell data analysis. Nat. Commun. 12, 6876 (2021). https://doi.org/10.1038/s41467-021-27150-6

60. He, L. et al. NEBULA is a fast negative binomial mixed model for differential or co-expression analysis of large-scale multi-subject single-cell data. Commun. Biol. 4, 629 (2021). https://doi.org/10.1038/s42003-021-02146-6

61. Clopper, C. J. & Pearson, E. S. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26, 404–413 (1934). https://doi.org/10.1093/biomet/26.4.404

62. Hedges, L. V. Distribution theory for Glass's estimator of effect size and related estimators. J. Educ. Stat. 6, 107–128 (1981). https://doi.org/10.3102/10769986006002107

63. Efron, B. Better bootstrap confidence intervals. J. Am. Stat. Assoc. 82, 171–185 (1987). https://doi.org/10.1080/01621459.1987.10478410

64. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57, 289–300 (1995). https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Supplementary Files

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

- SupplementaryData1.xlsx
- YarlagaddaPanNDAsubtypesSupplementaryInformation.docx