Neurobiology of Aging 83 (2019) 42–53

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

Biomarker-guided clustering of Alzheimer's disease clinical syndromes

Nicola Toschi [a,b,c,1], Simone Lista [d,e,f,1], Filippo Baldacci [d,e,f,g], Enrica Cavedo [d,e,f,h], Henrik Zetterberg [i,j,k,l], Kaj Blennow [i,j], Ingo Kilimann [m], Stefan J. Teipel [m], Antonio Melo dos Santos [d,e,f], Stéphane Epelbaum [d,e,f], Foudil Lamari [n], Remy Genthon [d,f], Marie-Odile Habert [o,p,q], Bruno Dubois [d,e,f], Roberto Floris [a], Francesco Garaci [a,r], Andrea Vergallo [d,e,f], Harald Hampel [d,s,*], the INSIGHT-preAD study group for the Alzheimer Precision Medicine Initiative (APMI)

[a] *Department of Biomedicine and Prevention, University of Rome "Tor Vergata", Rome, Italy*
[b] *Department of Radiology, "Athinoula A. Martinos" Center for Biomedical Imaging, Boston, MA, USA*
[c] *Harvard Medical School, Boston, MA, USA*
[d] *Sorbonne University, GRC n° 21, Alzheimer Precision Medicine (APM), AP-HP, Pitié-Salpêtrière Hospital, Paris, France*
[e] *Brain & Spine Institute (ICM), INSERM U 1127, Paris, France*
[f] *Department of Neurology, Institute of Memory and Alzheimer's Disease (IM2A), Pitié-Salpêtrière Hospital, Paris, France*
[g] *Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy*
[h] *Qynapse, Paris, France*
[i] *Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, The Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden*
[j] *Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden*
[k] *Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK*
[l] *UK Dementia Research Institute, London, UK*
[m] *Department of Psychosomatic Medicine, University of Rostock & DZNE Rostock, Rostock, Germany*
[n] *AP-HP, UF Biochimie des Maladies Neuro-métaboliques, Service de Biochimie Métabolique, Groupe Hospitalier Pitié-Salpêtrière, Paris, France*
[o] *Sorbonne Université, CNRS, INSERM, Laboratoire d'Imagerie Biomédicale, Paris, France*
[p] *Centre pour l'Acquisition et le Traitement des Images, France*
[q] *Département de Médecine Nucléaire, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France*
[r] *Casa di Cura "San Raffaele Cassino", Cassino, Italy*
[s] *Eisai Inc., Neurology Business Group, Woodcliff Lake, NJ, USA*

ARTICLE INFO

Article history:
Received 24 March 2019
Received in revised form 30 August 2019
Accepted 31 August 2019
Available online 10 September 2019

Keywords:
Alzheimer's disease
Biomarker-guided categorization
Clustering
Pathophysiology
Precision medicine

ABSTRACT

Alzheimer's disease (AD) neuropathology is extremely heterogeneous, and the evolution from preclinical to mild cognitive impairment until dementia is driven by interacting genetic/biological mechanisms not fully captured by current clinical/research criteria. We characterized the heterogeneous "construct" of AD through a cerebrospinal fluid biomarker-guided stratification approach. We analyzed 5 validated pathophysiological cerebrospinal fluid biomarkers ($A\beta_{1\text{-}42}$, t-tau, p-tau_{181}, NFL, YKL-40) in 113 participants (healthy controls [N = 20], subjective memory complainers [N = 36], mild cognitive impairment [N = 20], and AD dementia [N = 37], age: 66.7 ± 10.4, 70.4 ± 7.7, 71.7 ± 8.4, 76.2 ± 3.5 years [mean ± SD], respectively) using Density-Based Spatial Clustering of Applications with Noise, which does not require a priori determination of the number of clusters. We found 5 distinct clusters (sizes: N = 38, 16, 24, 14, and 21) whose composition was independent of phenotypical groups. Two clusters showed biomarker profiles linked to neurodegenerative processes not associated with classical AD-related pathophysiology. One cluster was characterized by the neuroinflammation biomarker YKL-40. Combining nonlinear data aggregation with informative

* Corresponding author at: Director of the Sorbonne University Groupe de Recherche Clinique (GRC n°21), "Alzheimer Precision Medicine (APM)", Établissements Publics à caractère Scientifique et Technologique (E.P.S.T.), President of the Alzheimer Precision Medicine Initiative (APMI). Tel.: +33 1 57 27 46 74; fax: +33 1 42 16 75 16.
E-mail address: harald.hampel@med.uni-muenchen.de (H. Hampel).
[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.neurobiolaging.2019.08.032

biomarkers can generate novel patient strata which are representative of cellular/molecular pathophysiology and may aid in predicting disease evolution and mechanistic drug response.

1. Introduction

Over the past 3 decades, technological advances have transformed the conceptual framework of Alzheimer's disease (AD). Postmortem studies demonstrated a high degree of neuropathological heterogeneity in patients who received a clinical diagnosis of AD, emphasizing the need to develop reliable in vivo biomarkers for AD-related pathophysiology (Rabinovici et al., 2017). Accordingly, the current research diagnostic criteria (Albert et al., 2011; Dubois et al., 2010, 2007; McKhann et al., 2011; Sperling et al., 2011) recommend the biomarker-based in vivo demonstration of AD pathomechanistic alterations, that is, brain overaccumulation of both brain amyloid-β (Aβ) (as indicated by low Aβ peptide, forty-two amino acid–long Aβ peptide [$A\beta_{1\text{-}42}$] concentrations in the cerebrospinal fluid [CSF]) and neurofibrillary tangles (as indicated by elevated CSF concentrations of hyperphosphorylated tau [p-tau] and total tau [t-tau] proteins).

Furthermore, the growing need for early-stage clinical trials for disease-modifying therapies is fostering a progressive shift from a clinical toward a biological definition of AD, along its clinical continuum (Jack et al., 2018). For the sake of stratifying individuals by AD pathomechanistic alterations, the "A/T/N" scheme, an agnostic conceptual framework focused on a biological definition of the pathophysiological continuum of AD, has been proposed. The A/T/N categorizes individuals by the presence/absence of core AD-related pathophysiological hallmarks (Aβ and tau proteinopathies plus neurodegeneration) (Jack et al., 2016a, 2018).

In the next years, the implementation of the current A/T/N scheme with additional pathophysiological process, including neuroinflammation, axonal damage, and synaptic dysfunction is expected to occur (Hampel et al., 2018b; Molinuevo et al., 2018). The integration of the A/T/N system in clinical trials is expected to support time-sensitive pathway-based therapeutic approaches.

Ideally, all relevant pathophysiological mechanisms supporting AD should be integrated into a biomarker-guided stratification strategy to support the aggregation of patients from large-scale population studies into homogeneous "clusters" (i.e., strata of individuals with distinct biomarker profiles) independent of clinical phenotypes. This regrouping (i.e., stratification) strategy has the potential to establish subsets of individuals which share similar disease-related trajectories and drug responses. In this context, one study predicted patient evolution based on clusters generated from both CSF and magnetic resonance imaging data (Nettiksimmons et al., 2010). Previous work clustered patients with AD using, for example, the CLustering In QUEst (CLIQUE) strategy (Gamberger et al., 2016a,b) (not including dimensionality reduction and focused on patients with a very large number of attributes), hierarchical clustering (which provides a multiscale picture of clusters but no indication of which dimensionality to choose), or k-means clustering (which requires a priori determination of the number of clusters) for profiling patients with AD, older adults in general (Escudero et al., 2012; Zemedikun et al., 2018), or patients with other disorders, including aphasia (Hoffman et al., 2017).

In this exploratory analysis, we sought to investigate whether different AD clinical syndromes may drop into different biomarker-driven clusters as well as whether different clinical syndromes may drop into the same biomarker-driven cluster.

In view of this objective, we recruited the entire continuum of the AD "construct"—from the asymptomatic cognitively normal preclinical stage (Dubois et al., 2016), including subjective memory complainers (SMCs), through mild cognitive impairment (MCI), all the way to clinically overt dementia. We used a panel of validated and innovative CSF biomarkers, including (1) the traditional CSF core, feasible (Frank et al., 2003; Hampel et al., 2008) biomarkers to track AD pathophysiology—the $A\beta_{1\text{-}42}$, t-tau, and tau hyperphosphorylated at threonine 181 ($p\text{-}tau_{181}$) proteins (Dubois et al., 2010, 2014, 2007; Albert et al., 2011; Jack et al., 2016a; McKhann et al., 2011; Sperling et al., 2011)—and (2) 2 additional CSF candidate biomarkers: neurofilament light chain (NFL) (Lista et al., 2017)—a structural component of the neuroaxonal cytoskeleton—and YKL-40 (Baldacci et al., 2017b)—a specific macrophage differentiation glycoprotein highly expressed in astrocytes. These are biomarkers of large-caliber myelinated axon disintegration (Olsson et al., 2016) and astrocytic activation, which are core mechanisms of neurodegeneration and neuroinflammation (Baldacci et al., 2017a; Olsson et al., 2016), respectively.

We believe that asking this question may provide useful insights for next clinical trials investigating biomarker-guided molecular combination therapies for AD.

2. Material and methods

2.1. Study participants

We conducted a multicenter cross-sectional study in a convenience sample (N = 113) recruited in 3 independent academic memory clinics. Particularly, healthy controls (HCs) (N = 20), SMCs (N = 36), MCI (N = 20), and AD dementia (ADD) (N = 37) individuals were examined. Age, sex, and Mini–Mental State Examination (MMSE) are reported in Table 1. Specifically, 58 participants were recruited from the Institute for Memory and Alzheimer's Disease (*Institut de la Mémoire et de la Maladie d'Alzheimer, IM2A*)—a subcohort of the Alzheimer Precision Medicine Initiative Cohort Program; available at https://www.apmiscience.com/) (Hampel et al., 2017, 2018c; Hampel et al., 2019)—at the Pitié-Salpêtrière University Hospital (Paris, France), 42 from the German Center for Neurodegenerative Diseases (Rostock, Germany), and 13 from the Institute of Neuroscience and Physiology at Sahlgrenska University Hospital (Mölndal, Sweden). The recruitment center will be referred to as "site" in this article.

The study complied with the tenets of the Declaration of Helsinki of 1975 and was approved by the local Ethical Committees at each participating university center. All participants or their representatives gave written informed consent for the use of their clinical data for research purposes.

2.2. Clinical diagnosis

The clinical diagnosis of ADD was performed according to the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association consensus criteria (McKhann et al., 1984) and the clinical core of the current diagnostic criteria for the amnestic presentation of ADD (Dubois et al., 2014; McKhann et al., 2011). N = 37 patients with ADD were included. All participants had a diagnosis of typical ADD (i.e., hippocampal type).

The clinical diagnosis of MCI was based on the MCI core clinical criteria (Albert et al., 2011). The group clinically defined as MCI included 20 participants. These individuals, predominantly

Table 1
Demographic and clinical data of the individuals stratified by clinical diagnosis

	HC (N = 20)	SMC (N = 36)	MCI (N = 20)	ADD (N = 37)
Sex (M/F)	7/13	11/25	13/7	18/19
Age at the time of CSF collection (yrs)	60.7 ± 10.3	76.2 ± 3.5	71.7 ± 8.4	70.4 ± 7.7
MMSE	29.4 ± 0.8	28.6 ± 1.1	25.5 ± 2.3	21.7 ± 5.0
CSF biomarkers				
p-tau$_{181}$ (pg/mL)	43.5 ± 8.5	54.0 ± 20.1	82.3 ± 39.4	102.0 ± 51.0
t-tau (pg/mL)	198.6 ± 78.6	431.6 ± 202.2	380.0 ± 190.3	449.5 ± 167.5
A$\beta_{1\text{-}42}$ (pg/mL)	912.0 ± 146.4	439.6 ± 173.1	620.8 ± 375.8	553.9 ± 299.1
NFL (pg/mL)	684.7 ± 296.1	975.6 ± 353.1	1245.7 ± 583.0	1622.1 ± 583.4
YKL-40 (ng/mL)	102.910 ± 40.895	224.557 ± 72.723	145.371 ± 57.041	153.358 ± 54.308

Numbers denote frequency for sex, mean ± standard deviation for age, MMSE, and CSF biomarkers.
Key: ADD, Alzheimer's disease dementia; A$\beta_{1\text{-}42}$, forty-two amino acid–long amyloid-β peptide; CSF, cerebrospinal fluid; F, female; HCs, healthy controls; M, male; MCI, mild cognitive impairment; MMSE, Mini–Mental State Examination; NFL, neurofilament light-chain protein; p-tau$_{181}$, tau hyperphosphorylated at threonine 181; SMC, subjective memory complainers; t-tau, total tau.

amnestic type MCI, had a one-year clinical follow-up. Three of them converted to ADD within one year.

Thirty-six participants with SMC, available at the time of study execution, were recruited from the "INveStIGation of AlzHeimer's PredicTors in Subjective Memory Complainers" (INSIGHT-preAD) study, a French standardized large-scale, observational, monocentric, academic, university-based cohort which is part of the Alzheimer Precision Medicine Initiative Cohort Program (https://www.apmiscience.com/) (Hampel et al., 2017, 2018c; Dubois et al., 2018; Hampel et al., 2019) at the time of the preparation of the article. The status of SMC was confirmed as follows: (1) participants gave an affirmative answer ("YES") to both questions: "Are you complaining about your memory?" and "Is it a regular complaint that has lasted now more than 6 months?"; (2) participants presented intact cognitive functions based on MMSE score ≥27, Clinical Dementia Rating scale = 0, and Free and Cued Selective Rating Test (total recall score ≥41) (Dubois et al., 2016). Amyloid-positron emission tomography (PET) imaging investigation was performed at baseline visit, as mandatory study inclusion criterion. Ten SMC participants were amyloid-PET positive and 26 participants were amyloid-PET negative. Amyloid-PET imaging processing has been previously described (Dubois et al., 2018; Habert et al., 2018).

Cognitively HCs (N = 20) were individuals who (1) volunteered for lumbar puncture, (2) were free of neurological or psychiatric diseases, and (3) had an MMSE score between 27 and 30.

2.3. CSF sampling and immunoassays for CSF core biomarkers, NFL, and YKL-40

2.3.1. CSF withdrawn and preanalytical procedures

CSF was taken in the morning through a standard lumbar puncture. All CSF samples used in the present study and deriving from the 3 different cohorts were collected in polypropylene tubes and processed as follows: centrifugation at 1000 g for 10 minutes at the temperature of +4 °C (IM2A, Pitié-Salpêtrière University Hospital, in Paris), 1500 g for 10 minutes at the temperature of +4 °C (samples collected at the Department of Psychosomatic Medicine inside the University of Rostock), 1800 g for 10 minutes at the temperature of +4 °C (samples collected at the Clinical Neurochemistry Laboratory in Mölndal). The obtained supernatant was collected, homogenized, and aliquoted into multiple 0.5 mL cryovial-sterilized tubes and finally stored at −80 °C within 1 hour from collection and until biochemical assessment.

2.3.2. Immunoassays for the assessment of CSF core biomarkers of AD pathophysiology

CSF measurement of AD core biomarkers—that is, A$\beta_{1\text{-}42}$, t-tau, p-tau$_{181}$—were performed, for the Paris cohort, at the Laboratory of Biochemistry, Unit of Biochemistry of Neurometabolic diseases, Pitié-Salpêtrière University Hospital in Paris; for the Rostock cohort in 2 different units: the Laboratory of Neurochemistry, Department of Neurology, Göttingen University Medical Center, before June 2012, and the Institute of Clinical Chemistry and Laboratory Medicine, Rostock University Medical Centre, as of June 2012; for the Gothenburg cohort, at the Clinical Neurochemistry Laboratory at the Sahlgrenska University Hospital in Mölndal. The concentrations of the 3 AD core biomarkers were measured using established sandwich ELISA methods, INNOTEST β-AMYLOID (1–42) (Vanderstichele et al., 2000), INNOTEST hTAU-Ag (Blennow et al., 1995), and INNOTEST Phospho-Tau [181P] (Vanmechelen et al., 2000) (Fujirebio Europe NV, Ghent, Belgium), respectively. All analyses were carried out by experienced laboratory technicians who were blinded to clinical data. All laboratories involved in the present study participate in the Alzheimer's Association Quality Control Program for CSF biomarkers (Mattsson et al., 2011) and the Global Biomarker Standardization Consortium (Carrillo et al., 2013). The threshold cutoff values for identifying pathologic concentrations of the core biomarkers were different across laboratories, that is, at the IM2A in Paris, A$\beta_{1\text{-}42}$ < 500 pg/mL, t-tau > 450 pg/mL, p-tau$_{181}$ > 60 pg/mL; at the German Center for Neurodegenerative Diseases in Rostock, A$\beta_{1\text{-}42}$ < 567 pg/mL, t-tau > 512 pg/mL, p-tau$_{181}$ > 66 pg/mL (for the CSF samples assessed before June 2012) and A$\beta_{1\text{-}42}$ < 450 pg/mL, t-tau > 450 pg/mL, p-tau$_{181}$ > 62 pg/mL (for the CSF samples assessed after June 2012); at Clinical Neurochemistry Laboratory in Mölndal, A$\hat{a}_{1\text{-}42}$ < 550 pg/mL, t-tau > 400 pg/mL, p-tau$_{181}$ > 80 pg/mL.

2.3.3. Immunoassays for the assessment of CSF biomarkers NFL and YKL-40

CSF NFL and YKL-40 were analyzed at the Clinical Neurochemistry Laboratory at the Sahlgrenska University Hospital in Mölndal, Sweden. In particular, CSF NFL protein concentrations were quantified using a sensitive sandwich ELISA method (NF-light ELISA kit; UmanDiagnostics AB, Umeå, Sweden), according to the recommendations provided by the manufacturer. The lower limit of quantification (LLOQ) for this assay was 50 ng/L. CSF YKL-40 protein concentrations were quantified using a commercially available ELISA kit (R&D Systems, Minneapolis, MN, USA), according to manufacturer instructions. The LLOQ for this assay was 60 pg/mL. All patient samples showed values greater than the in-house LLOQ. Intra-assay coefficients of variation were less than 10%. The measurements of each biomarker were performed in one round of experiments, using the same batch of reagents, by board-certified laboratory technicians who were blinded to the clinical data.

2.4. Statistical analysis

Our data allowed us to represent every participant as a "point" in a five-dimensional space, where the 5 coordinates are the 5 CSF

biomarkers under investigation ($A\beta_{1\text{-}42}$, t-tau, p-tau_{181}, NFL, and YKL-40). In accordance with our hypothesis that the embedding of individuals in this biomarker-based space could be able to reveal novel, unknown associations as well as possibly categories of individuals—clusters—we proceeded as follows.

As a first step, to eliminate age- and sex-related confounds, both variables were simultaneously regressed out of all biomarker values using a quadratic regression model which also included age-sex interaction. Then, this five-dimensional space of adjusted (for age, sex, and their interaction) biomarker values was fed into t-Distributed Stochastic Neighbor Embedding (t-SNE) to reduce dimensionality and, successively, into density-based clustering to formally identify clusters of arbitrary shape/different densities without imposing prior constraints on the number of clusters. t-SNE is a nonlinear dimensionality reduction technique well-suited for embedding high-dimensional data for visualization in a low-dimensional space of 2 or 3 dimensions. Specifically, in t-SNE, a probability distribution over pairs of high-dimensional objects is constructed so that similar objects have a high probability of being picked while minimizing the probability of picking dissimilar points. Successively, another probability distribution is designed over the points in a lower-dimensional map, and finally the Kullback-Leibler divergence between the 2 distributions is minimized with respect to where the points are located in the map. This allows an efficient and high-performing dimensionality reduction problem which optimally preserves high-dimensional relationship between data points (van der Maaten, 2008, 2009, 2014; van der Maaten and Hinton, 2012). Then, density-based clustering (Density-Based Spatial Clustering of Applications with Noise [DBSCAN]) (Patwary et al., 2012; Tran et al., 2013) can be applied to the low-dimensional representation to automatically identify clusters. It groups closely packed points (i.e., points with a high number of close neighbors) and marks points in low-density regions as outliers (i.e., noise). Briefly, when a point is found to be a dense part of a cluster, its ε-neighborhood is also considered to belong to the same cluster and all corresponding points are added, as well as their own ε-neighborhood when they are also dense, continuing until the density-connected cluster is complete. Successively, a point which has not previously been visited is retrieved and processed the same way—this can lead to the discovery of either noise or of a further cluster. Both t-SNE and DBSCAN can be used with any distance function and do not require initial, arbitrary determination of the number of clusters. In terms of validation, although some methods have been proposed to test the generalizability of unsupervised clustering methods (Tibshirani and Walther, 2005), it should be noted that the DBSCAN approach seeks to explore the data for partitions with high-density areas of points (clusters, i.e., clusters which, importantly, are not necessarily globular) separated by low-density areas, possibly containing noise objects. Dealing with noise objects correctly in a validation approach is a challenge that makes the definition of external or even cross-validated methods arduous, and no method has yet been proposed for this particular case. In addition, standard internal validity metrics commonly applied in unsupervised clustering applications (such as, e.g., silhouette analysis) fail for arbitrarily shaped, nonconvex clusters and, again, are not defined for noise objects. For the purpose of this revision, we have implemented a recently presented method for density-based, arbitrarily shaped clusters (Moulavi et al., 2014), which assesses clustering quality based on the relative density connection between pairs of objects and is based on a new kernel density function. After cluster identification, to explore cluster profiles and how different we were from each other, we used several strategies: first, the average minimum Manhattan Distance between every pair of clusters was calculated. In addition, biomarker values were compared across clusters using Kruskal-Wallis tests, and when a significant ($p < 0.05$) effect of group (i.e., cluster) was found, post hoc tests were conducted to identify the main drivers of this effect (2-sided tests adjusted for multiple comparisons across pairs of clusters). In addition, for each biomarker, the overall data set was divided in quintiles (see Table 2 for quintile boundaries) to observe to which quintile the median value of each biomarker in each cluster belonged. Using quintiles has the advantage of defining robust ranking intervals, centered on median of each distribution, which can provide a sense of where the median of each biomarker in each cluster falls with respect to the whole population median. Quintiles are intervals built using percentiles (i.e., bottom quintile = 20th percentile, second quintile = 21st–40th percentile, center quintile = 41st–60th percentile, fourth quintile = 61st–80th percentile, top quintile = 81st–100th percentile). The overall workflow of the clustering strategy is exemplified in Fig. 1.

3. Results

Demographic and clinical data of the individuals stratified by clinical diagnosis are reported in Table 1.

We found 5 distinct clusters of individuals (sizes: N = 38, N = 16, N = 24, N = 14, and N = 21). The low-dimensional embedding, which provides an easily readable 2D representation of the clusters we found, is shown in Fig. 2. We obtained a density-based cluster validation index of 0.67, indicating good-quality clustering.

Table 2
Quintile boundaries; cluster distances; cluster profiles according to the quintile (computed on the whole population) in which the median biomarker value of each cluster falls

Quintile boundaries	Bottom quintile-second quintile	Second quintile-center quintile	Center quintile-fourth quintile	Fourth quintile-top quintile
CSF NFL (pg/mL)	677.0	968.6	1174.4	1513.0
CSF YKL-40 (pg/mL)	106,022.0	139,240.0	170,989.0	220,130.0
CSF $A\beta_{1-42}$ (pg/mL)	346.0	424.0	507.9	805.0
CSF t-tau (pg/mL)	202.3	334.0	433.0	598.0
CSF p-tau_{181} (pg/mL)	42.0	53.0	68.0	98.0

Cluster distances

	C1	C2	C3	C4	C5
C1		2.79	2.17	0.94	1.53
C2	2.79		3.74	0.88	0.72
C3	2.17	3.74		2.49	2.92
C4	0.94	0.88	2.49		0.33
C5	1.53	0.72	2.92	0.33	

Cluster profiles	CSF NFL	CSF YKL-40	CSF $A\beta_{1\text{-}42}$	CSF t-tau	CSF p-tau_{181}
C1	–	–	–	↑	↑
C2	↓	↓↓	↑	↓	↓
C3	–	↑↑	–	↑	↑
C4	↓	–	–	–	–
C5	–	↓	↑	↓	–

↑↑ = top quintile; ↑ = fourth quintile; – = center quintile; ↓ = second quintile; ↓↓ = bottom quintile.
Key: $A\beta_{1-42}$, forty-two amino acid–long amyloid-β peptide; CSF, cerebrospinal fluid; NFL, neurofilament light-chain protein, p-tau_{181}, tau hyperphosphorylated at threonine 181; t-tau, total tau protein, C1, cluster 1; C2, cluster 2, C3, cluster 3; C4, cluster 4, C5, cluster 5.

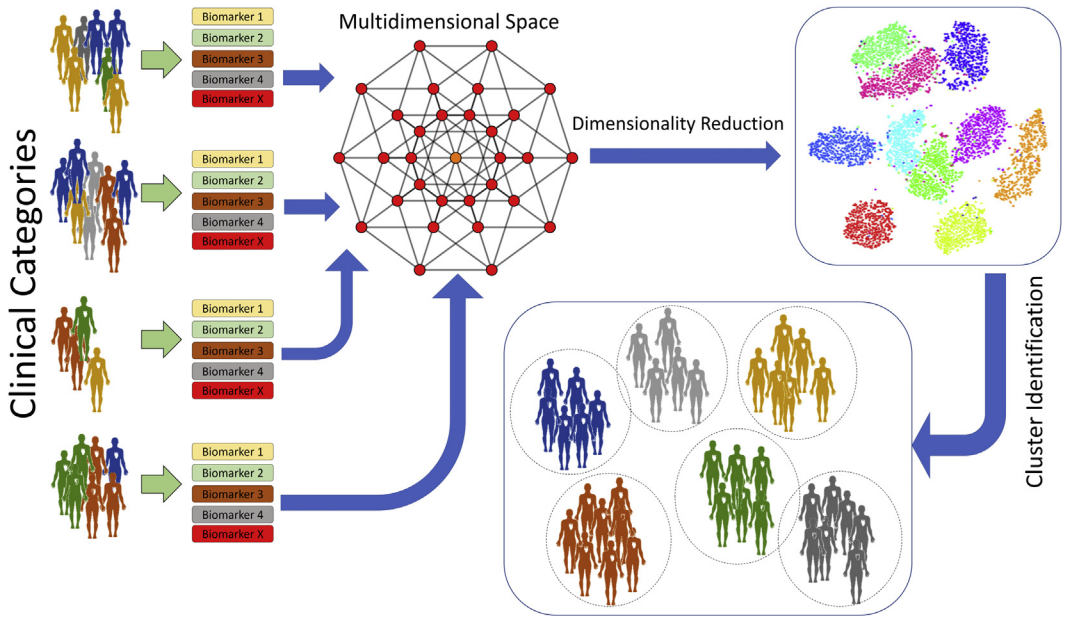

Fig. 1. Overall workflow for clustering procedure. For every clinical category (which may be composed of individuals with heterogeneous biomarker profiles), all N biomarkers under study are sampled and the values (after adjustment for confounds like sex and age through quadratic regression) are arranged into a multidimensional space where each individual is represented by a dot with N coordinates. Successively, a dimensionality reduction algorithm is applied, followed by a cluster identification procedure which allows redefining subject groupings based on biomarker profiles exclusively. All computational procedures are purely data driven.

Cross-tabulation based on cluster assignment and clinical categorization—that is, HC, SMC, MCI, and ADD—is shown in Table 3 and the distribution of each biomarker across clusters is shown in Fig. 3. Particularly, Table 3 shows how all individuals and their respective clinical categories are redistributed across clusters (both in absolute number and in percentage), as well as how the individuals belonging to each cluster are distributed across clinical categories. As an example, when looking at the first row from left to

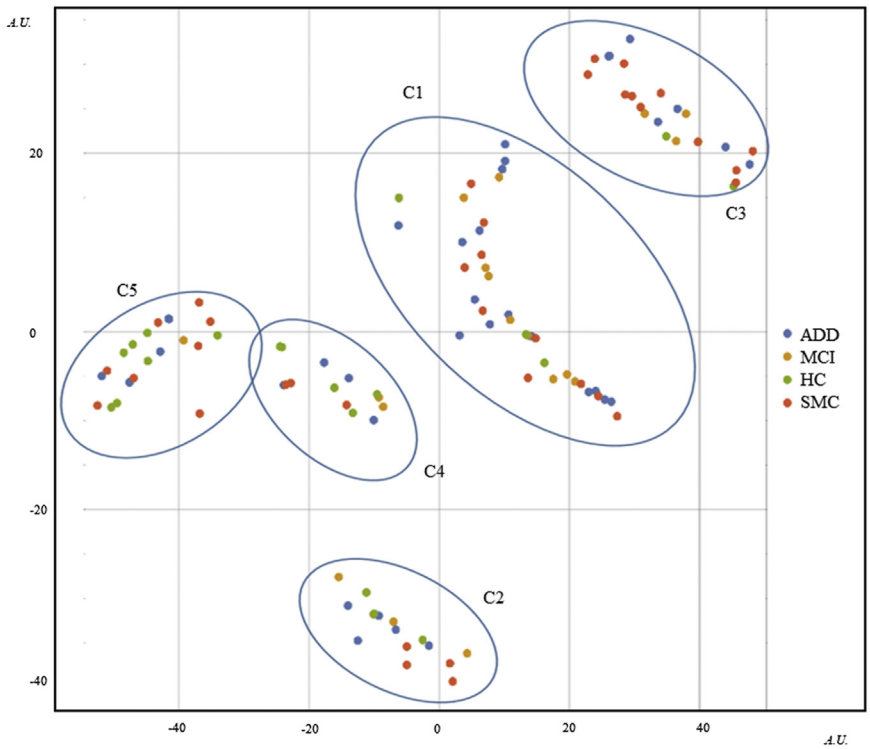

Fig. 2. t-SNE low-dimensional embedding of identified clusters of individuals. The colors represent the clinical diagnosis (blue = ADD; orange = MCI; green = HC; red = SMC). The blue ellipses delimit the clusters identified by the DBSCAN procedure. Abbreviations: ADD, Alzheimer's disease dementia; C1, cluster 1; C2, cluster 2; C3, cluster 3; C4, cluster 4; C5, cluster 5; DBSCAN, Density-Based Spatial Clustering of Applications with Noise; HC, healthy controls; MCI, mild cognitive impairment; SMC, subjective memory complainers. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Cross-tabulation based on clinical categorization: HC, SMC, MCI, and ADD

			Clinical category				Total
			HC	SMC	MCI	ADD	
CLUSTER	C1	Count	3	10	9	16	38
		% Within cluster	7.9%	26.3%	23.7%	42.1%	100.0%
		% Within clinical category	15.0%	27.8%	45.0%	43.2%	33.6%
		% Of total	2.7%	8.8%	8.0%	14.2%	33.6%
	C2	Count	3	4	4	5	16
		% Within cluster	18.8%	25.0%	25.0%	31.3%	100.0%
		% Within category	15.0%	11.1%	20.0%	13.5%	14.2%
		% Of total	2.7%	3.5%	3.5%	4.4%	14.2%
	C3	Count	2	11	4	7	24
		% Within cluster	8.3%	45.8%	16.7%	29.2%	100.0%
		% Within clinical category	10.0%	30.6%	20.0%	18.9%	21.2%
		% Of total	1.8%	9.7%	3.5%	6.2%	21.2%
	C4	Count	5	3	2	4	14
		% Within cluster	35.7%	21.4%	14.3%	28.6%	100.0%
		% Within clinical category	25.0%	8.3%	10.0%	10.8%	12.4%
		% Of total	4.4%	2.7%	1.8%	3.5%	12.4%
	C5	Count	7	8	1	5	21
		% Within cluster	33.3%	38.1%	4.8%	23.8%	100.0%
		% Within clinical category	35.0%	22.2%	5.0%	13.5%	18.6%
		% Of total	6.2%	7.1%	0.9%	4.4%	18.6%
TOTAL		Count	20	36	20	37	113
		% Within cluster	17.7%	31.9%	17.7%	32.7%	100.0%
		% Within clinical Category	100.0%	100.0%	100.0%	100.0%	100.0%
		% Of total	17.7%	31.9%	17.7%	32.7%	100.0%

Key: ADD, Alzheimer's disease dementia; C1, cluster 1; C2, cluster 2; C3, cluster 3; C4, cluster 4; C5, cluster 5; CSF, cerebrospinal fluid; HCs, healthy controls; MCI, mild cognitive impairment; SMC, subjective memory complainers.

right, one can see that cluster 1 is composed of 7.9% HC (3 individuals), 26.3% SMC (10 individuals), 23.7% MCI (9 individuals), and 42.3% ADD (16 individuals), which in total form 33.6% (38 individuals) of all participants (113 individuals). Similarly, looking at the first column from the top to bottom, we can see that HC individuals were distributed as follows: 15%, 15%, 10%, 25%, and 35% in clusters 1, 2, 3, 4, and 5, respectively, for a total of 20 individuals, which represent 17.7% of the total (113 individuals). This reasoning can be applied to every row and column to better understand the relationship between the clinical categorization and the data-driven clusters we derived. Importantly, chi-square tests "Clinical category × Cluster" and "Site × Cluster" yielded p values of 0.16 and 0.14 (respectively), indicating that cluster formation was not significantly influenced by site-related effects and that the 2 categorizations—clinical and biomarker-guided—are not statistically related to each other.

The clusterwise biomarker concentrations are reported in Table 4. The Manhattan distance between clusters is shown in Table 2, demonstrating how, for example, the pair of clusters which were furthest apart was clusters 2 and cluster 3, whereas the clusters which were closest to each other were cluster 4 and cluster 5.

In terms of statistical comparison between clusters, the Kruskal-Wallis tests for comparing multiple medians yielded the following group effects: YKL-40 ($p < 0.001$), t-tau ($p < 0.001$), NFL ($p = 0.019$), $A\beta_{1\text{-}42}$ ($p = 0.532$), and p-tau_{181} ($p < 0.001$). In addition, the results of post hoc testing are shown in Table 5 (in terms of resulting p value).

Finally, computing biomarker quintiles across the whole population and allocating cluster medians for each biomarker to each quintile resulted in the cluster profiles shown in Table 2.

It should be noted that, although Table 5 is based on simply comparing medians (while the multivariate cluster formation process is based on matching multivariate probability distributions), along with Table 2, it can provide an idea of the strongest drivers of the cluster formation process. One could, for example, examine in how many biomarkers a given pairwise comparison is statistically significant or, conversely, in how many pairwise comparisons a given biomarker is significantly different between pairs of clusters (see Section 4). For example, the separations of clusters 2 versus 3 (which corresponds to the highest distance between clusters; Table 2) and clusters 3 versus 5 (which corresponds to the second-highest distance between clusters; Table 2) appear to be a result of several biomarkers in conjunction—possibly mainly driven by YKL-40 and t-tau in the first case and by YKL-40 in the second case. In addition, median biomarker values are not statistically different between clusters 2 versus 5 and clusters 4 versus 5 (which corresponds to the lowest distance between clusters; Table 2). As mentioned previously, this is not a discrepancy because the cluster forming process is based on the whole biomarker distribution as opposed to median values only. In addition, a weak effect is noted in YKL-40 and NFL in comparing cluster 3 versus cluster 5. In terms of cluster profiles within the overall population, from Table 2, one can infer that, for example, cluster 3 is characterized by a much higher (with respect to the whole population median) YKL-40 concentration, whereas cluster 2 is characterized by lower (with respect to the whole population median) NFL, t-tau, and p-tau_{181} concentrations, higher $A\beta_{1\text{-}42}$ concentration, and a much lower (with respect to the whole population median) YKL-40 concentration. By the same token and under the same approximations, one could infer that, overall, YKL-40 is the main driver of division, followed by t-tau, p-tau_{181}, and NFL that contribute very little to cluster separation. $A\beta_{1\text{-}42}$ does not significantly contribute to cluster separation.

3.1. Description of clusters

Cluster 1 was the largest (N = 38) and consisted of MCI (23.7%), SMC (26.3%), ADD (42.7%) participants, as well as of a small portion of HC (only 7.9%) participants. This cluster included a single MCI subject that converted to dementia at follow-up. CSF t-tau and p-tau_{181} median concentrations were in the 4th quintile (Table 2). These individuals showed significantly higher CSF concentrations of YKL-40 compared with those of clusters 2 and 5. CSF t-tau and p-

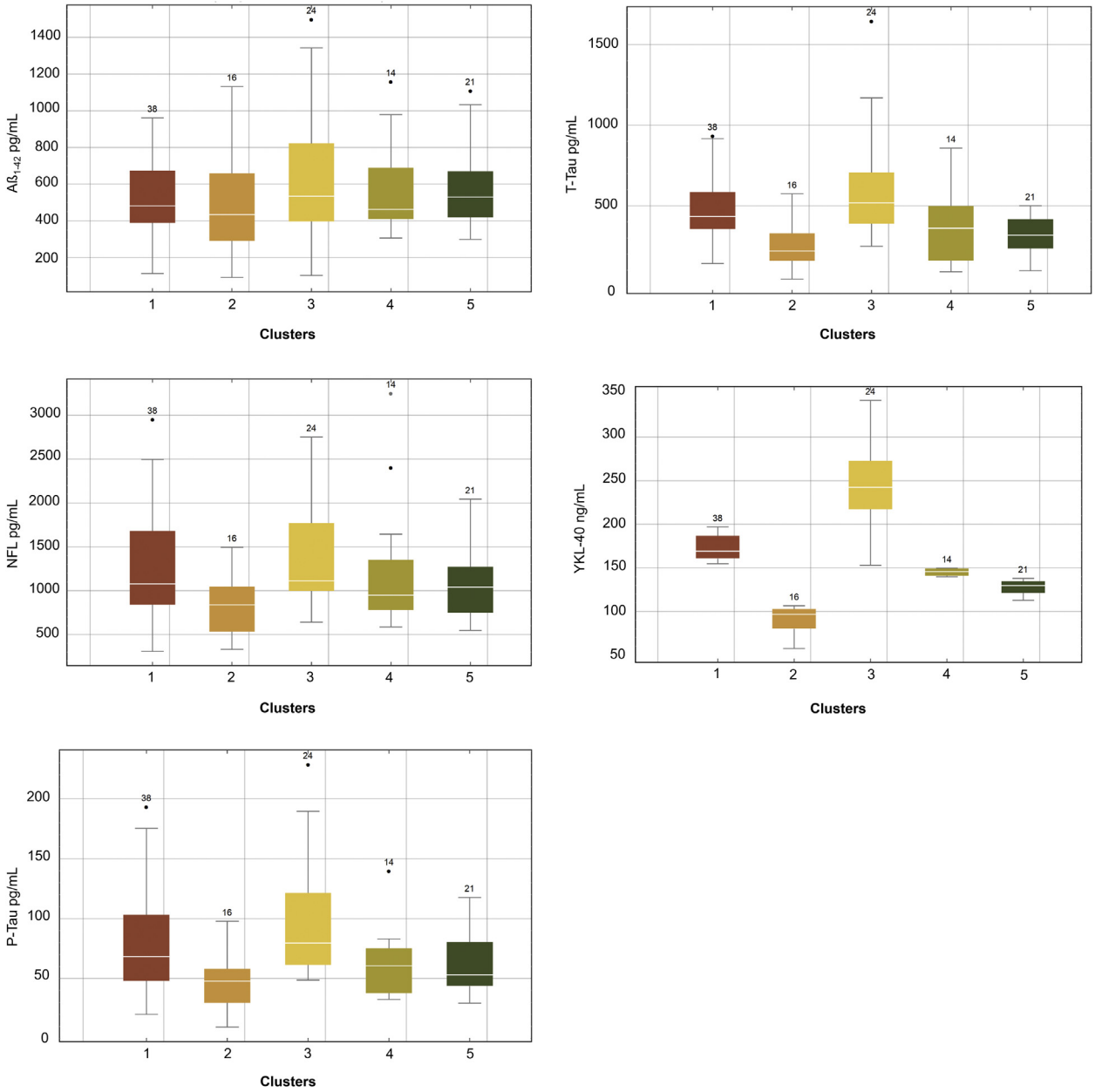

Fig. 3. Box-whisker plots depicting the distributions of biomarkers across the clusters. The X axis indicates reports the number of clusters; the Y axis indicates the biomarker concentration. Middle line: median. Boxes: interquartile rage. Whiskers: Extremes. Points: Outliers. Numbers: Group numerosities. Abbreviations: $A\beta_{1\text{-}42}$, forty-two amino acid–long amyloid-β peptide; NFL, neurofilament light-chain protein; p-tau_{181}, tau hyperphosphorylated at threonine 181; t-tau, total tau protein.

Table 4
Biomarker concentrations of the individuals stratified by biomarker-guided clusters

Cluster	CSF biomarkers				
	p-tau_{181} (pg/mL)	t-tau (pg/mL)	$A\beta_{1\text{-}42}$ (pg/mL)	NFL (pg/mL)	YKL-40 (pg/mL)
Cluster 1	80.2 ± 44.0 (68.1)	457.6 ± 209.5 (434.6)	522.6 ± 202.8 (481.7)	1253.0 ± 558.6 (1076.7)	172,711.0 ± 13,574.5 (169,037.0)
Cluster 2	47.0 ± 23.5 (47.8)	252.1 ± 140.2 (219.56)	504.5 ± 294.8 (434.7)	836.8 ± 354.2 (838.2)	89,272.4 ± 17,302.4 (96,569.3)
Cluster 3	95.7 ± 46.7 (79.5)	598.1 ± 319.1 (518.8)	621.9 ± 331.8 (534.3)	1384.4 ± 579.9 (1111.4)	250,776.0 ± 46,763.7 (242,451.0)
Cluster 4	62.1 ± 28.5 (60.5)	369.1 ± 238.9 (360.9)	572.0 ± 254.8 (462.3)	1236.5 ± 743.5 (950.1)	144,999.0 ± 4040.5 (145,820.0)
Cluster 5	60.0 ± 23.4 (53.0)	317.1 ± 114.2 (317.5)	583.1 ± 206.2 (529.9)	1057.2 ± 368.9 (1039.6)	127,808.0 ± 7567.9 (129,711.0)

Numbers denote mean ± standard deviation for CSF biomarkers.
Key: $A\beta_{1\text{-}42}$, forty-two amino acid–long amyloid-β peptide; CSF, cerebrospinal fluid; NFL, neurofilament light-chain protein, p-tau_{181}, tau hyperphosphorylated at threonine 181; t-tau, total tau protein.

Table 5
Results of post hoc testing after Kruskal-Wallis tests

Biomarker	C1 vs. C2	C1 vs. C3	C1 vs. C4	C1 vs. C5	C2 vs. C3	C2 vs. C4	C2 vs. C5	C3 vs. C4	C3 vs. C5	C4 vs. C5
YKL-40	**<0.001**	**0.009**	0.008	**<0.001**	**<0.001**	**0.027**	ns.	**<0.001**	**<0.001**	ns.
t-tau	**0.006**	ns.	ns.	0.015	**<0.001**	ns.	ns.	0.007	**0.003**	ns.
p-tau$_{181}$	**0.033**	ns.	ns.	ns.	**0.001**	ns.	ns.	0.013	**0.027**	ns.
NFL	0.010	ns.	ns.	ns.	**0.013**	ns.	ns.	ns.	ns.	ns.
Aβ$_{1-42}$	ns.	ns.	ns.	ns.	ns.	ns.	ns.	ns.	ns.	ns.

Only significant (<0.05) *p* values are shown. *p* values which are also significant after correction for multiple comparisons are shown in bold.
Key: Aβ$_{1-42}$, forty-two amino acid–long amyloid-β peptide; C1, cluster 1; C2, cluster 2; C3, cluster 3; C4, cluster 4; C5, cluster 5; NFL, neurofilament light-chain protein; ns, nonsignificant; p-tau$_{181}$, tau hyperphosphorylated at threonine 181; t-tau, total tau protein.

tau$_{181}$ concentrations were significantly higher than those of cluster 2 (Fig. 3 and Table 5).

Cluster 2 (N = 16) included ADD (31.3%), SMC (25.0%), MCI (25.0%), and HC (18.8%) participants. Only one subject, within the MCI group, converted to dementia at follow-up. CSF YKL-40 median concentrations were in the bottom quintile and CSF NFL, t-tau, and p-tau$_{181}$ were in the 2nd quintile. On the contrary, the median concentrations of CSF Aβ$_{1-42}$ were in the 4th quintile (Table 2). CSF YKL-40 concentrations were significantly lower than those of clusters 1, 3, and 4. CSF t-tau and p-tau$_{181}$ concentrations were significantly lower than those of clusters 1 and 3. CSF NFL concentrations were significantly lower than of those of cluster 3 (Fig. 3 and Table 5).

Cluster 3 (N = 24) was represented by SMC (45.8%), ADD (29.2%), MCI (16.7%), and HC (8.3%) participants. These individuals showed the highest CSF YKL-40 concentrations compared with those of all other clusters (Fig. 3 and Table 5) with a median concentration falling in the top quintile. CSF t-tau and p-tau$_{181}$ were in the 4th quintile (Table 2). CSF concentrations of t-tau and p-tau$_{181}$ were significantly higher than those of clusters 2 and 5. CSF NFL concentrations were significantly higher compared with those of cluster 2 (Fig. 3 and Table 5).

Cluster 4 was the smallest one (N = 14) and consisted of HC (35.7%), ADD (28.6%), MCI (14.3%), and SMC (21.4%) participants. One MCI subject converted to dementia at follow-up. CSF NFL median concentrations were in the 2nd quintile (Table 2). CSF YKL-40 concentrations were significantly higher than those of cluster 2 and significantly lower than cluster 3. CSF t-tau and p-tau$_{181}$ concentrations were significantly lower than those of cluster 3 (Fig. 3 and Table 5).

Cluster 5 (N = 21) included SMC (38.1%), HC (33.3%), ADD (23.8%), and MCI (4.8%) participants. CSF YKL-40 and t-tau median concentrations were in the 2nd quintile, whereas the median concentrations of CSF Aβ$_{1-42}$ were in the 4th quintile (Table 2). CSF YKL-40 concentrations were significantly lower compared with those of clusters 1 and 3. CSF t-tau and p-tau$_{181}$ concentrations were significantly lower than those of cluster 3 (Fig. 3 and Table 5).

4. Discussion

In this clustering investigation, we asked the question of how many pathophysiological profiles may underlie the clinical progression of the typical (hippocampal type) AD phenotype (that encompasses syndromic diagnoses ranging from SMC to MCI to overt dementia) and asymptomatic individuals. To this end, we compared participant groups, generated via biostatistical clustering analysis, to groups obtained using a traditional diagnostic categorization based on currently available clinical diagnostic criteria. Importantly, in our approach, the number of clusters is not imposed a priori but, rather, determined automatically by optimality criteria. We explored the possibility of deriving a meaningful stratification of individuals distributed along the broad biological and clinical *spectrum* of AD (Dubois et al., 2016; Aisen et al., 2017) using exclusively core biological fluid markers which reflect distinct pathomechanistic alterations associated with the disease (i.e., brain Aβ accumulation and neurofibrillary pathology, neuroinflammation, axonal damage, and neurodegeneration). In our population, we found no significant relationship between the 2 categorizations (i.e., clinical diagnosis and clusters), emphasizing how data-driven similarity criteria can uncover novel between-individual associations in biomarker space, which reflect different pathophysiological mechanisms not necessarily mirrored in clinically descriptive categories. In this context, it should however be noted that the chi-square test between clinical diagnosis and clusters returned a *p* value of 0.16. Although this is traditionally considered not statistically significant, it reflects the fact that, overall, clinical categories did somewhat polarize (albeit not to a statistically significant extent) the clustering results to a certain extent. This naturally reflects the effect of peculiar syndromic clinical diagnoses. Relatedly, we did not find a biomarker-based cluster that uniquely corresponded to the clinical AD diagnosis. This is not unexpected, as it would only occur if this clinical diagnosis was homogeneous in terms of biomarker profiles. However, it is interesting to note that cluster 1 contains 42.3% of all patients with AD and that patients with AD constitute 42.1% of the whole cluster count. This lends further support to the complex interplay and only partial overlap between clinical and biomarker-driven categories.

Cluster 3 was the furthest/most distant (Table 2) from the majority of other clusters (2, 4, and 5) but not from cluster 1. Cluster 1 and cluster 3 show a biomarker profile linked to neuronal dystrophy and loss, that is, neurodegeneration (t-tau) and tau-mediated brain proteinopathy (p-tau$_{181}$) (and additionally, neuroinflammation [YKL-40] in the case of cluster 3), which do not entirely correspond to the predefined AD-related pathophysiological core (Sperling et al., 2011; Dubois et al., 2014; Jack et al., 2016b, 2018). These 2 clusters were mostly represented by patients with ADD (cluster 1) and SMC individuals (cluster 3). Because CSF YKL-40 concentrations in cluster 3 are significantly higher than those of cluster 1, one could argue that an association between neurodegeneration, neurofibrillary deposition, and neuroinflammatory mechanisms may exist in cognitively normal subjects at risk for ADD (Baldacci et al., 2017a).

Clusters 2, 4, and 5 were close to each other and presented average or lower median concentrations of biomarkers charting axonal damage and neurodegeneration—namely NFL and t-tau, of tau-mediated brain proteinopathy (p-tau$_{181}$)—and neuroinflammation (YKL-40) compared with those found in clusters 1 and 3.

Moreover, cluster 2 was the most distant from clusters 1 and 3, which in turn showed the highest median concentrations of biomarkers of neurodegeneration and neuroinflammation. Therefore, in spite of including a large share of patients with ADD, cluster 1 and cluster 2 present different biomarker profiles, particularly with reference to the level of cerebral amyloidosis. In addition, cluster 2 showed, along with cluster 5, higher median concentrations of

$A\beta_{1-42}$ compared with all other clusters. The same clusters were characterized by a high degree of clinical heterogeneity, especially cluster 2 which, however, is the second of 5 for prevalence of ADD.

Moreover, cluster 5 was mostly represented by SMC and HC, whereas cluster 4 primarily included HC and ADD. At a speculative interpretation level, the different biomarker profiles in clusters 4 and 5 may suggest that neuroinflammation, tau-mediated toxicity, and cerebral amyloidosis could be the pathophysiological mechanisms driving cognitive decline. Interestingly, cluster 5, mainly composed by SMC and HC, is the only cluster where median t-tau concentration does not lie in the same quintile as median p-tau_{181} concentration. Moreover, in cluster 5, YKL-40 median concentration lies in the 2nd quintile, whereas $A\beta_{1-42}$ lies in the central quintile. This profile suggests that post-translational modifications of tau-protein alone may not be sufficient to determine cognitive decline. This possibility is in line with the comparison between clusters 1 and 3 (see aforementioned). Moreover, besides the core pathophysiological biomarkers of AD, innovative but robust biomarkers tracking distinctive pathophysiological mechanisms, such as neuroinflammation, may account (even if not alone) for cognitive decline along the clinical AD continuum. Detecting novel biologically determined (e.g., inflammation based) disease categories (for instance, cluster 3 which presents the highest CSF median YKL-40 concentration) is expected to substantially improve disease prediction and provide key tools for accurate pathway-based therapies, such as drugs targeting neuroinflammation.

It is important to note that the high-dimensional feature reduction and aggregation (i.e., clustering) approach we used in this article is able to deal with partial redundancies while still extracting additional information from multiple variables which exhibit collinearities. In line with the aforementioned, it is also important to note that partially collinear biomarkers do not necessarily imply a representation of the same pathomechanistic alteration. For example, in spite of p-tau and t-tau stemming from a common precursor, p-tau has been demonstrated to reflect neurofibrillary pathology while t-tau is an established marker of axonal damage and neuronal injury (Jack et al., 2018). Indeed, an agnostic hypothesis-independent biomarker-driven classification system (the A/T/N scheme) has been proposed to stratify individuals according to core AD-related pathological and pathophysiological hallmarks (brain overaccumulation of both Aβ and tau proteins aggregates and neurodegeneration) (Jack et al., 2018). In the currently available research diagnostic criteria and in the A/T/N scheme, CSF t-tau and p-tau_{181} play a different role. Despite the fact that NFL represents a marker of axonal damage, similarly to t-tau, the former is more tightly related to the damage of large-caliber fiber axons (Jack et al., 2018; Shahim et al., 2018, 2016).

In summary, we found that our biomarker-guided clustering approach generates a set of specific clusters not significantly bound to original distinct clinically phenotyped diagnostic groups. Specifically, none of the clusters appears homogeneous enough to be translated into predefined clinical categories that can be considered specific of AD pathophysiology. Instead, each cluster includes all phenotypical groups. Interestingly, CSF $A\beta_{1-42}$ concentrations are less likely to have contributed to the process of cluster segregation. A potential explanation for this finding could be that $A\beta_{1-42}$ peptide modifications appear several years before symptom onset and are perhaps not linearly associated with the progressions of neurodegeneration and cognitive decline (Jack et al., 2013; Sperling et al., 2011). In contrast, all other biomarkers significantly contribute to separating clusters. In this regard, previous studies reported that all these biomarkers are positively correlated and associated with worsening of cognitive performance (Hampel et al., 2018d; Mattsson et al., 2016). This finding allows us to hypothesize that the individuals included in cluster 3—those with the highest median YKL-40 concentration in CSF—may benefit from participating in targeted clinical trials using compounds acting against neuroinflammation. Moreover, our clusters do not show a significant "Site × Cluster" interaction and are generated after correction of biomarker values for age, sex, and their interaction, thus ensuring that our model is biologically robust to the impact of aging and sexual dimorphism on AD-related pathophysiological mechanisms (Cavedo et al., 2018; Ferretti et al., 2018; Hampel et al., 2018f).

Unsupervised clustering strategies are ideal for identifying multivariate, nonlinear associations between individuals which can be described or characterized by a large number of variables—for example, biomarker and clinical outcomes—by dealing with an n-dimensional space which cannot be treated appropriately through intuition or classical linear methods. Particularly, these algorithms search for clusters by simultaneously taking all variables into account, regardless of their nature. The similarity metric used can also be varied in case there is a strong hypothesis about the nature of the association between biomarkers. By clustering individuals according to a multidimensional profile which could span all data realms available (e.g., genetic risk factors, fluid biomarker concentrations, imaging modalities), it should be feasible to define groups of individuals who share main pathophysiological drivers and triggers and, possibly, similar longitudinal disease trajectories. Under the hypotheses that the input data contain all main information potentially driving the evolution of the disease, these findings might support the discovery and development of targeted and individualized therapies with proven disease-modifying effects, consistent with the standardized stepwise procedure of proof-of-pharmacology (Hampel et al., 2018a,g). Thus, it would be possible to treat individuals at asymptomatic preclinical stages of the disease, when the imbalance of homeostatic dynamics is still potentially restorable and the functionality of brain networks can be preserved. This innovation, applied in drug research and development programs as well as in clinical practice, is in line with the precision pharmacology (Hampel et al., 2018e) and precision medicine paradigms (Hampel et al., 2016, 2017, 2018c; Hampel et al., 2019). In this context, we anticipate that unsupervised methods—complemented by integrative disease modeling (Baldacci et al., 2018, 2016; Hampel et al., 2017; Younesi and Hofmann-Apitius, 2013)—will accelerate and optimize clinical trial development for individualized treatments through the implementation of biomarker matrices serving for different contexts-of-use, namely risk prediction, early stratification and early detection and diagnosis, treatment efficacy and safety monitoring, and prognostic evaluation.

In follow-up studies, we plan to further expand the panel of pathophysiological biomarkers to establish a stratification algorithm which could include biomarkers of synaptic damage, such as neurogranin (Lista and Hampel, 2016) and alpha-synuclein (Vergallo et al., 2018). Moreover, we also aim at including a core blood-based biomarker panel to take optimal advantage of the broad accessibility of blood plasma–based analyses to screen international large-scale cohorts of asymptomatic individuals at risk (Hampel et al., 2018b,g). This is of great importance for future translation of these classification and early-detection approaches into worldwide primary clinical practice.

Our explorative pilot study presents some caveats. First, our follow-up data are restricted to SMC and MCI participants, hence hampering the characterization of the clusterwise subject-trajectories through time. Second, the sample size is relatively limited, and most clustering algorithms perform better on a large number of individuals. Still, it should be noted that an over-/under-representation of any clinical category within any particular cluster does not, per se, indicate instability of the clustering procedure. In this context, it is important to note that one of the overall aims of

our article is to demonstrate how the same clinical syndrome may be allocated to different biomarker-driven clusters as well as that different clinical syndromes may drop in the same biomarker-driven cluster. In addition, given the partial collinearity of the biomarkers which is, however, possibly associated with different biological underpinnings, and the limited sample size, we hypothesized that a highly nonlinear and state-of-the-art dimension reduction algorithm such as t-Distributed Stochastic Neighbor Embedding could allow better data separation and hence cluster identification as compared to running a clustering algorithm on the 5-dimensional data directly.

Our multicenter cohort is clinically heterogeneous—that is, it includes individuals ranging from the status of SMC to overt ADD—and, although data were corrected for age using nonlinear regression, residual differences in terms of , for example, age may still affect the corrected concentrations of CSF biomarkers (Jack et al., 2016b). In addition, our study did not use magnetic resonance imaging or amyloid-PET imaging data, which, however, could be seamlessly included to enhance our or other clustering strategies, as previously reported (Ten Kate et al., 2018; Young et al., 2018).

In conclusion, we found (1) a set of biologically defined clusters not significantly linked to the clinical diagnosis, (2) that $A\beta_{1-42}$ is less likely to have contributed to the cluster segregation, and (3) that all other biomarkers, especially YKL-40, significantly contribute to separating clusters. We believe that this stratification approach, based on state-of-the-art biomarker-guided clustering algorithms, should not replace but, rather, complement, optimize, and enrich the traditional diagnostic approaches in neurology to design treatments tailored to the individual biological-clinical profile (Hampel et al., 2016, 2017, 2018e; Hampel et al., 2019). The integration of clinical data with biological comprehensive information may represent the crucial step forward to the development of an accurate biomarker-guided stratification framework, which has the potential to inform and innovate traditional diagnostic workups as well as treatment selection strategies for next-generation clinical trials.

Disclosure

SL received lecture honoraria from Roche and Servier. HZ and KB are co-founders of Brain Biomarker Solutions in Gothenburg AB, a GU Ventures–based platform company at the University of Gothenburg. HZ has served at scientific advisory boards for Roche Diagnostics, CogRx, Samumed, and Wave. KB has served as a consultant or at advisory boards for Alzheon, CogRx, Biogen, Lilly, Novartis, and Roche Diagnostics. SE received lecture honoraria from Roche and Astellas Pharma and participated on scientific advisory boards of GE Healthcare and Eli Lilly. RG is a former employee of Sanofi and holds stocks of Sanofi; he declares no conflict of interest on this work. MOH has received consultant's honoraria from GE Healthcare, AVID-LILLY, and PIRAMAL. BD reports personal fees from Eli Lilly and company. AV received lecture honoraria from Roche, MagQu LLC, and Servier. HH is an employee of Eisai Inc and serves as Senior Associate Editor for the Journal Alzheimer's & Dementia; he received lecture fees from Servier, Biogen and Roche, research grants from Pfizer, Avid, and MSD Avenir (paid to the institution), travel funding from Eisai, Functional Neuromodulation, Axovant, Eli Lilly and company, Takeda and Zinfandel, GE Healthcare, and Oryzon Genomics and consultancy fees from Qynapse, Jung Diagnostics, Cytox Ltd, Axovant, Anavex, Takeda and Zinfandel, GE Healthcare, Oryzon Genomics, and Functional Neuromodulation, and participated in scientific advisory boards of Functional Neuromodulation, Axovant, Eisai, Eli Lilly and company, Cytox Ltd, GE Healthcare, Takeda and Zinfandel, Oryzon Genomics, and Roche Diagnostics. He is co-inventor in the following patents as a scientific expert and has received no royalties:

- In Vitro Multiparameter Determination Method for The Diagnosis and Early Diagnosis of Neurodegenerative Disorders Patent Number: 8916388;
- In Vitro Procedure for Diagnosis and Early Diagnosis of Neurodegenerative Diseases Patent Number: 8298784;
- Neurodegenerative Markers for Psychiatric Conditions Publication Number: 20120196300;
- In Vitro Multiparameter Determination Method for The Diagnosis and Early Diagnosis of Neurodegenerative Disorders Publication Number: 20100062463;
- In Vitro Method for The Diagnosis and Early Diagnosis of Neurodegenerative Disorders Publication Number: 20100035286;
- In Vitro Procedure for Diagnosis and Early Diagnosis of Neurodegenerative Diseases Publication Number: 20090263822;
- In Vitro Method for The Diagnosis of Neurodegenerative Diseases Patent Number: 7547553;
- CSF Diagnostic in Vitro Method for Diagnosis of Dementias and Neuroinflammatory Diseases Publication Number: 20080206797;
- In Vitro Method for The Diagnosis of Neurodegenerative Diseases Publication Number: 20080199966;
- Neurodegenerative Markers for Psychiatric Conditions Publication Number: 20080131921.

NT, FB, EC, IK, SJT, AMdS, FL, RF, and FG declare that they have no conflict of interest.

Acknowledgements

This research benefited from the support of the Program "**PHOENIX**" led by the Sorbonne University Foundation and sponsored by la *Fondation pour la Recherche sur Alzheimer*.

The study was promoted by INSERM in collaboration with ICM, IHU-A-ICM, and Pfizer and has received support within the "Investissement d'Avenir" (ANR-10-AIHU-06) program. The study was promoted in collaboration with the "CHU de Bordeaux" (coordination CIC EC7), the promoter of Memento cohort, funded by the Foundation Plan-Alzheimer. The study was further supported by AVID/Lilly. CATI is a French neuroimaging platform funded by the French Plan Alzheimer (available at http://cati-neuroimaging.com). HZ is a Wallenberg Academy Fellow and holds grants from the Swedish and European Research Councils as well as the Medical Research Council (UK). KB holds the Torsten Söderberg Professorship of Medicine and is supported by the Swedish Research Council (#2017-00915), the Swedish Alzheimer Foundation (#AF-742881), Hjärnfonden, Sweden (#FO2017-0243), and a grant (#ALFGBG-715986) from the Swedish state under the agreement between the Swedish government and the County Councils, the ALF agreement. HH is an employee of Eisai Inc. During part of this work, he was supported by the AXA Research Fund, the "Fondation partenariale Sorbonne Université" and the "Fondation pour la Recherche sur Alzheimer", Paris, France.

Contributors to the Alzheimer Precision Medicine Initiative—Working Group (APMI–WG): Mohammad AFSHAR (Paris), Lisi Flores AGUILAR (Montréal), Leyla AKMAN-ANDERSON (Sacramento), Joaquín ARENAS (Madrid), Claudio BABILONI (Rome), Filippo BALDACCI (Pisa), Richard BATRLA (Rotkreuz), Norbert BENDA (Bonn), Keith L. BLACK (Los Angeles), Arun L.W. BOKDE (Dublin), Ubaldo BONUCCELLI (Pisa), Karl BROICH (Bonn), Francesco CACCIOLA (Siena), Filippo CARACI (Catania), Juan CASTRILLO† (Derio),

Enrica CAVEDO (Paris), Roberto CERAVOLO (Pisa), Patrizia A. CHIESA (Paris), Jean-Christophe CORVOL (Paris), Augusto Claudio CUELLO (Montréal), Jeffrey L. CUMMINGS (Las Vegas), Herman DEPYPERE (Ghent), Bruno DUBOIS (Paris), Andrea DUGGENTO (Rome), Enzo EMANUELE (Robbio), Valentina ESCOTT-PRICE (Cardiff), Howard FEDEROFF (Irvine), Maria Teresa FERRETTI (Zürich), Massimo FIANDACA (Irvine), Richard A. FRANK (Malvern), Francesco GARACI (Rome), Hugo GEERTS (Berwyn), Filippo S. GIORGI (Pisa), Edward J. GOETZL (San Francisco), Manuela GRAZIANI (Roma), Marion HABERKAMP (Bonn), Marie-Odile HABERT (Paris), Harald HAMPEL (Paris), Karl HERHOLZ (Manchester), Dimitrios KAPOGIANNIS (Baltimore), Eric KARRAN (Cambridge), Steven J. KIDDLE (Cambridge), Seung H. KIM (Seoul), Yosef KORONYO (Los Angeles), Maya KORONYO-HAMAOUI (Los Angeles), Todd LANGEVIN (Minneapolis-Saint Paul), Stéphane LEHÉRICY (Paris), Pablo LEMERCIER (Paris), Simone LISTA (Paris), Jean LORENCEAU (Paris), Alejandro LUCÍA (Madrid), Dalila MANGO (Rome), Mark MAPSTONE (Irvine), Christian NERI (Paris), Robert NISTICÒ (Rome), Sid E. O'BRYANT (Fort Worth), Giovanni PALERMO (Pisa), George PERRY (San Antonio), Craig RITCHIE (Edinburgh), Simone ROSSI (Siena), Amira SAIDI (Rome), Emiliano SANTARNECCHI (Siena), Lon S. SCHNEIDER (Los Angeles), Olaf SPORNS (Bloomington), Nicola TOSCHI (Rome), Steven R. VERDOONER (Sacramento), Andrea VERGALLO (Paris), Nicolas VILLAIN (Paris), Lindsay A. WELIKOVITCH (Montréal), Janet WOODCOCK (Silver Spring), Erfan YOUNESI (Esch-sur-Alzette).

References

Aisen, P.S., Cummings, J., Jack, C.R., Morris, J.C., Sperling, R., Frölich, L., Jones, R.W., Dowsett, S.A., Matthews, B.R., Raskin, J., Scheltens, P., Dubois, B., 2017. On the path to 2025: understanding the Alzheimer's disease continuum. Alzheimers Res. Ther. 9, 60.

Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., Jagust, W.J., Petersen, R.C., Snyder, P.J., Carrillo, M.C., Thies, B., Phelps, C.H., 2011. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 270–279.

Baldacci, F., Lista, S., Cavedo, E., Bonuccelli, U., Hampel, H., 2017a. Diagnostic function of the neuroinflammatory biomarker YKL-40 in Alzheimer's disease and other neurodegenerative diseases. Expert Rev. Proteomics 14, 285–299.

Baldacci, F., Lista, S., Garaci, F., Bonuccelli, U., Toschi, N., Hampel, H., 2016. Biomarker-guided classification scheme of neurodegenerative diseases. J. Sport Health Sci. 5, 383–387.

Baldacci, F., Lista, S., O'Bryant, S.E., Ceravolo, R., Toschi, N., Hampel, H., Alzheimer Precision Medicine Initiative (APMI), 2018. Blood-based biomarker screening with agnostic biological definitions for an accurate diagnosis within the dimensional spectrum of neurodegenerative diseases. Methods Mol. Biol. 1750, 139–155.

Baldacci, F., Toschi, N., Lista, S., Zetterberg, H., Blennow, K., Kilimann, I., Teipel, S., Cavedo, E., Dos Santos, A.M., Epelbaum, S., Lamari, F., Dubois, B., Floris, R., Garaci, F., Bonuccelli, U., Hampel, H., 2017b. Two-level diagnostic classification using cerebrospinal fluid YKL-40 in Alzheimer's disease. Alzheimers Dement. 13, 993–1003.

Blennow, K., Wallin, A., Agren, H., Spenger, C., Siegfried, J., Vanmechelen, E., 1995. Tau protein in cerebrospinal fluid: a biochemical marker for axonal degeneration in Alzheimer disease? Mol. Chem. Neuropathol. 1995, 231–245.

Cavedo, E., Chiesa, P.A., Houot, M., Ferretti, M.T., Grothe, M.J., Teipel, S.J., Lista, S., Habert, M.-O., Potier, M.-C., Dubois, B., Hampel, H., INSIGHT-preAD Study Group, Alzheimer Precision Medicine Initiative (APMI), 2018. Sex differences in functional and molecular neuroimaging biomarkers of Alzheimer's disease in cognitively normal older adults with subjective memory complaints. Alzheimers Dement. 14, 1204–1215.

Carrillo, M.C., Blennow, K., Soares, H., Lewczuk, P., Mattsson, N., Oberoi, P., Umek, R., Vandijck, M., Salamone, S., Bittner, T., Shaw, L.M., Stephenson, D., Bain, L., Zetterberg, H., 2013. Global standardization measurement of cerebral spinal fluid for Alzheimer's disease: an update from the Alzheimer's Association Global Biomarkers Consortium. Alzheimers Dement. 9, 137–140.

Dubois, B., Epelbaum, S., Nyasse, F., Bakardjian, H., Gagliardi, G., Uspenskaya, O., Houot, M., Lista, S., Cacciamani, F., Potier, M.-C., Bertrand, A., Lamari, F., Benali, H., Mangin, J.-F., Colliot, O., Genthon, R., Habert, M.-O., Hampel, H., INSIGHT-preAD Study Group, 2018. Cognitive and neuroimaging features and brain β-amyloidosis in individuals at risk of Alzheimer's disease (INSIGHT-preAD): a longitudinal observational study. Lancet Neurol. 17, 335–346.

Dubois, B., Feldman, H.H., Jacova, C., Cummings, J.L., Dekosky, S.T., Barberger-Gateau, P., Delacourte, A., Frisoni, G., Fox, N.C., Galasko, D., Gauthier, S., Hampel, H., Jicha, G.A., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Sarazin, M., de Souza, L.C., Stern, Y., Visser, P.J., Scheltens, P., 2010. Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol. 9, 1118–1127.

Dubois, B., Feldman, H.H., Jacova, C., Dekosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P.J., Scheltens, P., 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol. 6, 734–746.

Dubois, B., Feldman, H.H., Jacova, C., Hampel, H., Molinuevo, J.L., Blennow, K., DeKosky, S.T., Gauthier, S., Selkoe, D., Bateman, R., Cappa, S., Crutch, S., Engelborghs, S., Frisoni, G.B., Fox, N.C., Galasko, D., Habert, M.-O., Jicha, G.A., Nordberg, A., Pasquier, F., Rabinovici, G., Robert, P., Rowe, C., Salloway, S., Sarazin, M., Epelbaum, S., de Souza, L.C., Vellas, B., Visser, P.J., Schneider, L., Stern, Y., Scheltens, P., Cummings, J.L., 2014. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol. 13, 614–629.

Dubois, B., Hampel, H., Feldman, H.H., Scheltens, P., Aisen, P., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., Broich, K., Cavedo, E., Crutch, S., Dartigues, J.-F., Duyckaerts, C., Epelbaum, S., Frisoni, G.B., Gauthier, S., Genthon, R., Gouw, A.A., Habert, M.-O., Holtzman, D.M., Kivipelto, M., Lista, S., Molinuevo, J.-L., O'Bryant, S.E., Rabinovici, G.D., Rowe, C., Salloway, S., Schneider, L.S., Sperling, R., Teichmann, M., Carrillo, M.C., Cummings, J., Jack, C.R., Proceedings of the Meeting of the International Working Group (IWG) and the American Alzheimer's Association on "The Preclinical State of AD"; July 23, 2015; Washington DC, USA, 2016. Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. Alzheimers Dement. 12, 292–323.

Escudero, J., Ifeachor, E., Zajicek, J.P., Alzheimer's Disease Neuroimaging Initiative, 2012. Bioprofile analysis: a new approach for the analysis of biomedical data in Alzheimer's disease. J. Alzheimers Dis. 32, 997–1010.

Ferretti, M.T., Iulita, M.F., Cavedo, E., Chiesa, P.A., Schumacher Dimech, A., Santuccione Chadha, A., Baracchi, F., Girouard, H., Misoch, S., Giacobini, E., Depypere, H., Hampel, H., Women's Brain Project and the Alzheimer Precision Medicine Initiative, 2018. Sex differences in Alzheimer disease - the gateway to precision medicine. Nat. Rev. Neurol. 14, 457–469.

Frank, R.A., Galasko, D., Hampel, H., Hardy, J., de Leon, M.J., Mehta, P.D., Rogers, J., Siemers, E., Trojanowski, J.Q., National Institute on Aging Biological Markers Working Group, 2003. Biological markers for therapeutic trials in Alzheimer's disease. Proceedings of the biological markers working group; NIA initiative on neuroimaging in Alzheimer's disease. Neurobiol. Aging 24, 521–536.

Gamberger, D., Ženko, B., Mitelpunkt, A., Lavrač, N., 2016a. Homogeneous clusters of Alzheimer's disease patient population. Biomed. Eng. Online 15, 78.

Gamberger, D., Ženko, B., Mitelpunkt, A., Shachar, N., Lavrač, N., 2016b. Clusters of male and female Alzheimer's disease patients in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Brain Inform. 3, 169–179.

Habert, M.-O., Bertin, H., Labit, M., Diallo, M., Marie, S., Martineau, K., Kas, A., Causse-Lemercier, V., Bakardjian, H., Epelbaum, S., Chételat, G., Houot, M., Hampel, H., Dubois, B., Mangin, J.-F., INSIGHT-AD Study Group, 2018. Evaluation of amyloid status in a cohort of elderly individuals with memory complaints: validation of the method of quantification and determination of positivity thresholds. Ann. Nucl. Med. 32, 75–86.

Hampel, H., Bürger, K., Teipel, S.J., Bokde, A.L.W., Zetterberg, H., Blennow, K., 2008. Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. Alzheimers Dement. 4, 38–48.

Hampel, H., Mesulam, M.-M., Cuello, A.C., Farlow, M.R., Giacobini, E., Grossberg, G.T., Khachaturian, A.S., Vergallo, A., Cavedo, E., Snyder, P.J., Khachaturian, Z.S., 2018a. The cholinergic system in the pathophysiology and treatment of Alzheimer's disease. Brain J. Neurol. 141, 1917–1933.

Hampel, H., O'Bryant, S.E., Castrillo, J.I., Ritchie, C., Rojkova, K., Broich, K., Benda, N., Nisticò, R., Frank, R.A., Dubois, B., Escott-Price, V., Lista, S., 2016. Precision medicine - the golden gate for detection, treatment and prevention of Alzheimer's disease. J. Prev. Alzheimers Dis. 3, 243–259.

Hampel, H., O'Bryant, S.E., Durrleman, S., Younesi, E., Rojkova, K., Escott-Price, V., Corvol, J.-C., Broich, K., Dubois, B., Lista, S., Initiative, for the A.P.M., 2017. A Precision Medicine Initiative for Alzheimer's disease: the road ahead to biomarker-guided integrative disease modeling. Climacteric 20, 107–118.

Hampel, H., O'Bryant, S.E., Molinuevo, J.L., Zetterberg, H., Masters, C.L., Lista, S., Kiddle, S.J., Batrla, R., Blennow, K., 2018b. Blood-based biomarkers for Alzheimer disease: mapping the road to the clinic. Nat. Rev. Neurol. 14, 639–652.

Hampel, H., Toschi, N., Babiloni, C., Baldacci, F., Black, K.L., Bokde, A.L.W., Bun, R.S., Cacciola, F., Cavedo, E., Chiesa, P.A., Colliot, O., Coman, C.-M., Dubois, B., Duggento, A., Durrleman, S., Ferretti, M.-T., George, N., Genthon, R., Habert, M.-O., Herholz, K., Koronyo, Y., Koronyo-Hamaoui, M., Lamari, F., Langevin, T., Lehéricy, S., Lorenceau, J., Neri, C., Nisticò, R., Nyasse-Messene, F., Ritchie, C., Rossi, S., Santarnecchi, E., Sporns, O., Verdooner, S.R., Vergallo, A., Villain, N., Younesi, E., Garaci, F., Lista, S., Alzheimer Precision Medicine Initiative (APMI), 2018c. Revolution of Alzheimer precision neurology. Passageway of systems biology and neurophysiology. J. Alzheimers Dis. 64, S47–S105.

Hampel, H., Toschi, N., Baldacci, F., Zetterberg, H., Blennow, K., Kilimann, I., Teipel, S.J., Cavedo, E., Melo Dos Santos, A., Epelbaum, S., Lamari, F., Genthon, R., Dubois, B., Floris, R., Garaci, F., Lista, S., Alzheimer Precision Medicine Initiative (APMI), 2018d. Alzheimer's disease biomarker-guided diagnostic workflow using the added value of six combined cerebrospinal fluid candidates: aβ1-42, total-tau, phosphorylated-tau, NFL, neurogranin, and YKL-40. Alzheimers Dement. 14, 492–501.

Hampel, H., Vergallo, A., Aguilar, L.F., Benda, N., Broich, K., Cuello, A.C., Cummings, J., Dubois, B., Federoff, H.J., Fiandaca, M., Genthon, R., Haberkamp, M., Karran, E., Mapstone, M., Perry, G., Schneider, L.S., Welikovitch, L.A., Woodcock, J., Baldacci, F., Lista, S., Alzheimer Precision Medicine Initiative (APMI), 2018e. Precision pharmacology for Alzheimer's disease. Pharmacol. Res. 130, 331–365.

Hampel, H., Vergallo, A., Giorgi, F.S., Kim, S.H., Depypere, H., Graziani, M., Saidi, A., Nisticò, R., Lista, S., Alzheimer Precision Medicine Initiative (APMI), 2018f. Precision medicine and drug development in Alzheimer's disease: the importance of sexual dimorphism and patient stratification. Front. Neuroendocrinol. 50, 31–51.

Hampel, H., Vergallo, A., Bonuccelli, U., Lista, S., 2018g. Editorial: turning point towards blood biomarker-guided targeted therapy for precision medicine in Alzheimer's disease. J. Prev. Alzheimers Dis. 5, 160–164.

Hampel, H., Vergallo, A., Perry, G., Lista, S., Alzheimer Precision Medicine Initiative (APMI), 2019. The Alzheimer Precision Medicine Initiative. J. Alzheimers Dis. 68, 1–24.

Hoffman, P., Sajjadi, S.A., Patterson, K., Nestor, P.J., 2017. Data-driven classification of patients with primary progressive aphasia. Brain Lang. 174, 86–93.

Jack, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J.L., Montine, T., Phelps, C., Rankin, K.P., Rowe, C.C., Scheltens, P., Siemers, E., Snyder, H.M., Sperling, R., Contributors, 2018. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement. 14, 535–562.

Jack, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Feldman, H.H., Frisoni, G.B., Hampel, H., Jagust, W.J., Johnson, K.A., Knopman, D.S., Petersen, R.C., Scheltens, P., Sperling, R.A., Dubois, B., 2016a. A/T/N: an unbiased descriptive classification scheme for Alzheimer disease biomarkers. Neurology 87, 539–547.

Jack, C.R., Knopman, D.S., Chételat, G., Dickson, D., Fagan, A.M., Frisoni, G.B., Jagust, W., Mormino, E.C., Petersen, R.C., Sperling, R.A., van der Flier, W.M., Villemagne, V.L., Visser, P.J., Vos, S.J.B., 2016b. Suspected non-Alzheimer disease pathophysiology–concept and controversy. Nat. Rev. Neurol. 12, 117–124.

Jack, C.R., Knopman, D.S., Jagust, W.J., Petersen, R.C., Weiner, M.W., Aisen, P.S., Shaw, L.M., Vemuri, P., Wiste, H.J., Weigand, S.D., Lesnick, T.G., Pankratz, V.S., Donohue, M.C., Trojanowski, J.Q., 2013. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol. 12, 207–216.

Lista, S., Hampel, H., 2016. Synaptic degeneration and neurogranin in the pathophysiology of Alzheimer's disease. Expert Rev. Neurother. 17, 47–57.

Lista, S., Toschi, N., Baldacci, F., Zetterberg, H., Blennow, K., Kilimann, I., Teipel, S.J., Cavedo, E., Dos Santos, A.M., Epelbaum, S., Lamari, F., Dubois, B., Floris, R., Garaci, F., Hampel, H., Alzheimer Precision Medicine Initiative (APMI), 2017. Diagnostic accuracy of CSF neurofilament light chain protein in the biomarker-guided classification system for Alzheimer's disease. Neurochem. Int. 108, 355–360.

Mattsson, N., Insel, P.S., Palmqvist, S., Portelius, E., Zetterberg, H., Weiner, M., Blennow, K., Hansson, O., Alzheimer's Disease Neuroimaging Initiative, 2016. Cerebrospinal fluid tau, neurogranin, and neurofilament light in Alzheimer's disease. EMBO Mol. Med. 8, 1184–1196.

Mattsson, N., Andreasson, U., Persson, S., Arai, H., Batish, S.D., Bernardini, S., Bocchio-Chiavetto, L., Blankenstein, M.A., Carrillo, M.C., Chalbot, S., Coart, E., Chiasserini, D., Cutler, N., Dahlfors, G., Duller, S., Fagan, A.M., Forlenza, O., Frisoni, G.B., Galasko, D., Galimberti, D., Hampel, H., Handberg, A., Heneka, M.T., Herskovits, A.Z., Herukka, S.K., Holtzman, D.M., Humpel, C., Hyman, B.T., Iqbal, K., Jucker, M., Kaeser, S.A., Kaiser, E., Kapaki, E., Kidd, D., Klivenyi, P., Knudsen, C.S., Kummer, M.P., Lui, J., Lladó, A., Lewczuk, P., Li, Q.X., Martins, R., Masters, C., McAuliffe, J., Mercken, M., Moghekar, A., Molinuevo, J.L., Montine, T.J., Nowatzke, W., O'Brien, R., Otto, M., Paraskevas, G.P., Parnetti, L., Petersen, R.C., Prvulovic, D., de Reus, H.P., Rissman, R.A., Scarpini, E., Stefani, A., Soininen, H., Schröder, J., Shaw, L.M., Skinningsrud, A., Skrogstad, B., Spreer, A., Talib, L., Teunissen, C., Trojanowski, J.Q., Tumani, H., Umek, R.M., Van Broeck, B., Vanderstichele, H., Vecsei, L., Verbeek, M.M., Windisch, M., Zhang, J., Zetterberg, H., Blennow, K., 2011. The Alzheimer's Association external quality control program for cerebrospinal fluid biomarkers. Alzheimers Dement. 7, 386–395.e6.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. Neurology 34, 939–944.

McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., Mohs, R.C., Morris, J.C., Rossor, M.N., Scheltens, P., Carrillo, M.C., Thies, B., Weintraub, S., Phelps, C.H., 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 263–269.

Molinuevo, J.L., Ayton, S., Batrla, R., Bednar, M.M., Bittner, T., Cummings, J., Fagan, A.M., Hampel, H., Mielke, M.M., Mikulskis, A., O'Bryant, S., Scheltens, P., Sevigny, J., Shaw, L.M., Soares, H.D., Tong, G., Trojanowski, J.Q., Zetterberg, H., Blennow, K., 2018. Current state of Alzheimer's fluid biomarkers. Acta Neuropathol. 136, 821–853.

Moulavi, D., Jaskowiak, P., Campello, R., Zimek, A., Sander, J., April 24-26, 2014. Density-based clustering validation. In: Proceedings of the 2014 SIAM International Conference on Data Mining, Proceedings. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, pp. 839–847.

Nettiksimmons, J., Harvey, D., Brewer, J., Carmichael, O., DeCarli, C., Jack, C.R., Petersen, R., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Beckett, L., Alzheimer's Disease Neuroimaging Initiative, 2010. Subtypes based on cerebrospinal fluid and magnetic resonance imaging markers in normal elderly predict cognitive decline. Neurobiol. Aging 31, 1419–1428.

Olsson, B., Lautner, R., Andreasson, U., Öhrfelt, A., Portelius, E., Bjerke, M., Hölttä, M., Rosén, C., Olsson, C., Strobel, G., Wu, E., Dakin, K., Petzold, M., Blennow, K., Zetterberg, H., 2016. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. Lancet Neurol. 15, 673–684.

Patwary, M.A., Palsetia, D., Agrawal, A., Liao, W., Manne, F., Choudhary, A., 2012. A new scalable parallel DBSCAN algorithm using the disjoint-set data structure. In: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, SC '12. IEEE Computer Society Press, Los Alamitos, CA, USA, p. 62.

Rabinovici, G.D., Carrillo, M.C., Forman, M., DeSanti, S., Miller, D.S., Kozauer, N., Petersen, R.C., Randolph, C., Knopman, D.S., Smith, E.E., Isaac, M., Mattsson, N., Bain, L.J., Hendrix, J.A., Sims, J.R., 2017. Multiple comorbid neuropathologies in the setting of Alzheimer's disease neuropathology and implications for drug development. Alzheimers Dement. (N. Y.) 3, 83–91.

Shahim, P., Tegner, Y., Gustafsson, B., Gren, M., Ärlig, J., Olsson, M., Lehto, N., Engström, Å., Höglund, K., Portelius, E., Zetterberg, H., Blennow, K., 2016. Neurochemical Aftermath of repetitive mild traumatic brain injury. JAMA Neurol. 73, 1308–1315.

Shahim, P., Tegner, Y., Marklund, N., Blennow, K., Zetterberg, H., 2018. Neurofilament light and tau as blood biomarkers for sports-related concussion. Neurology 90, e1780–e1788.

Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, C.R., Kaye, J., Montine, T.J., Park, D.C., Reiman, E.M., Rowe, C.C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M.C., Thies, B., Morrison-Bogorad, M., Wagster, M.V., Phelps, C.H., 2011. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 280–292.

Ten Kate, M., Dicks, E., Visser, P.J., van der Flier, W.M., Teunissen, C.E., Barkhof, F., Scheltens, P., Tijms, B.M., Alzheimer's Disease Neuroimaging Initiative, 2018. Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline. Brain J. Neurol. 141, 3443–3456.

Tibshirani, R., Walther, G., 2005. Cluster validation by prediction strength. J. Comput. Graph. Stat. 14, 511–528.

Tran, T.N., Drab, K., Daszykowski, M., 2013. Revised DBSCAN algorithm to cluster data with dense adjacent clusters. Chemom. Intell. Lab. Syst. 120, 92–96.

van der Maaten, L., 2014. Accelerating t-SNE using tree-based algorithms. J. Mach. Learn. Res. 15, 3221–3245.

van der Maaten, L., Hinton, G., 2012. Visualizing non-metric similarities in multiple maps. Mach. Learn. 87, 33–55.

van der Maaten, L., 2009. Learning a parametric embedding by preserving local structure. In: Artificial Intelligence and Statistics, pp. 384–391.

van der Maaten, L., Hinton, G., 2008. Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Vanderstichele, H., Van Kerschaver, E., Hesse, C., Davidsson, P., Buyse, M.A., Andreasen, N., Minthon, L., Wallin, A., Blennow, K., Vanmechelen, E., 2000. Standardization of measurement of beta-amyloid(1-42) in cerebrospinal fluid and plasma. Amyloid 7, 245–258.

Vanmechelen, E., Vanderstichele, H., Davidsson, P., Van Kerschaver, E., Van Der Perre, B., Sjögren, M., Andreasen, N., Blennow, K., 2000. Quantification of tau phosphorylated at threonine 181 in human cerebrospinal fluid: a sandwich ELISA with a synthetic phosphopeptide for standardization. Neurosci. Lett. 285, 49–52.

Vergallo, A., Bun, R.-S., Toschi, N., Baldacci, F., Zetterberg, H., Blennow, K., Cavedo, E., Lamari, F., Habert, M.-O., Dubois, B., Floris, R., Garaci, F., Lista, S., Hampel, H., INSIGHT-preAD study group, Alzheimer Precision Medicine Initiative (APMI), 2018. Association of cerebrospinal fluid α-synuclein with total and phospho-tau181 protein concentrations and brain amyloid load in cognitively normal subjective memory complainers stratified by Alzheimer's disease biomarkers. Alzheimers Dement. 14, 1623–1631.

Younesi, E., Hofmann-Apitius, M., 2013. From integrative disease modeling to predictive, preventive, personalized and participatory (P4) medicine. EPMA J. 4, 23.

Young, A.L., Marinescu, R.V., Oxtoby, N.P., Bocchetta, M., Yong, K., Firth, N.C., Cash, D.M., Thomas, D.L., Dick, K.M., Cardoso, J., van Swieten, J., Borroni, B., Galimberti, D., Masellis, M., Tartaglia, M.C., Rowe, J.B., Graff, C., Tagliavini, F., Frisoni, G.B., Laforce, R., Finger, E., de Mendonça, A., Sorbi, S., Warren, J.D., Crutch, S., Fox, N.C., Ourselin, S., Schott, J.M., Rohrer, J.D., Alexander, D.C., Genetic FTD Initiative (GENFI), Alzheimer's Disease Neuroimaging Initiative (ADNI), 2018. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. Nat. Commun. 9, 4273.

Zemedikun, D.T., Gray, L.J., Khunti, K., Davies, M.J., Dhalwani, N.N., 2018. Patterns of multimorbidity in middle-Aged and older adults: an analysis of the UK biobank data. Mayo Clin. Proc. 93, 857–866.