Title: Large-scale cerebrospinal fluid proteomic analysis in Alzheimer's disease patients reveals five molecular subtypes with distinct genetic risk profiles.

Authors: Betty M Tijms[1,2], Ellen M Vromen[1,2], Olav Mjaavatten[3], Henne Holstege[1,2,8], Lianne M Reus[1,2,5], Sven van der Lee[1,2,6], Kirsten EJ Wesenhagen[1,2], Luigi Lorenzini[7,8], Lisa Vermunt[2,9], Vikram Venkatraghavan[1,2], Niccoló Tesi[6,10], Jori Tomassen[1,2], Anouk den Braber[1,2], Julie Goossens[11], Eugeen Vanmechelen[11], Frederik Barkhof[5,6,12], Yolande AL Pijnenburg[1,2], Wiesje M van der Flier[1,2], Charlotte E Teunissen[2,9], Frode Berven[3], Pieter Jelle Visser[1,2,13,14].

Affiliations:
1 Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, the Netherlands.
2 Amsterdam Neuroscience, Neurodegeneration, Amsterdam, the Netherlands.
3 PROBE, department of biomedicine, University of Bergen, Bergen, Norway.
4 Department of Clinical Genetics, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, the Netherlands.
5 Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, USA
6 Genomics of Neurodegenerative Diseases and Aging, Human Genetics, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, the Netherlands.
7 Department of Radiology and Nuclear Medicine, Amsterdam UMC location VUmc, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.
8 Amsterdam Neuroscience, Neuroimaging, Amsterdam, the Netherlands.
9 Neurochemistry Laboratory, Department of Clinical Chemistry, Amsterdam UMC location VUmc, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.
10 Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands
11 ADx NeuroSciences, Ghent, Belgium.
12 Institutes of Neurology and Healthcare Engineering, University College London, London, UK
13 Alzheimer Center Limburg, School for Mental Health and Neuroscience, Maastricht University, Maastricht, the Netherlands.
14 Department of Neurobiology, Care Sciences and Society, Division of Neurogeriatrics, Karolinska Institutet, Stockholm, Sweden.

Keywords: Alzheimer's disease, cerebrospinal fluid, proteomics, disease heterogeneity

Abstract

Alzheimer's disease (AD) is heterogenous on the molecular level. Understanding this heterogeneity is critical for AD drug development. We aimed to define AD molecular subtypes by mass spectrometry proteomics in cerebrospinal fluid (CSF). Of the 3863 proteins detected in CSF, 1058 proteins had different levels in individuals with AD (n=419) compared with controls (n=187). Cluster analyses of AD individuals on these 1058 proteins revealed five subtypes: subtype 1 was characterized by neuronal hyperplasticity; subtype 2 by innate immune activation; subtype 3 by RNA dysregulation; subtype 4 by choroid plexus dysfunction; and subtype 5 by blood-brain barrier dysfunction. Distinct genetic profiles were associated with subtypes, e.g., subtype 1 was enriched with *TREM2 R47H*. Subtypes also differed in brain atrophy and clinical outcomes. For example, survival was shorter in subtype 3 compared to subtype 1 (5.6 versus 8.9 years). These novel insights into AD molecular heterogeneity highlight the need for personalized medicine.

Alzheimer's disease (AD) is the leading cause of dementia, affecting about 44 million people worldwide[1]. AD is histopathologically defined by amyloid plaques and hyperphosphorylated tau tangles in the brain[2], but its underlying pathophysiology remains largely unclear. Genetic studies, as well as brain tissue gene expression and proteomic studies have indicated that many different pathophysiological processes are associated with amyloid and tau pathology, including but not limited to synaptic plasticity, the innate immune system, neuroinflammation, lipid metabolism, RNA metabolism, the matrisome, and vascular function[3–9]. This heterogeneity may explain why previous AD trials had no or limited clinical effects[10–12]. For example, we previously found that CSF BACE1 levels were abnormally increased in a specific AD subtype, suggesting that BACE inhibition may be effective in a subgroup only[9,13]. This highlights the need for personalized treatments and for in-vivo tools to define such molecular subtypes.

CSF is the most accessible biofluid to study molecular complexity of neurodegenerative diseases during life: CSF is in close contact with the brain, and protein concentrations in CSF reflect the brain's ongoing (patho)physiological processes. We previously discovered and replicated three distinct molecular AD subtypes through investigation of respectively 707 and 204 proteins in CSF[9]. The proteins involved in these subtypes represent distinct biological processes such as neuronal plasticity, innate immune activation and blood-brain barrier dysfunction[9]. Subtype specific molecular alterations were already present at a very early stage of AD, when cognition was still intact and neuronal damage still limited. Furthermore, these molecular processes were previously also identified in AD post-mortem tissue proteomic studies [3,4,6,8]. This supports the value of CSF proteomics to detect AD pathophysiological processes in living patients[4].

Proteomic techniques have greatly improved since then, and can now detect thousands of proteins in CSF, providing an unprecedented opportunity to dissect the molecular processes associated with AD in detail. In our current study, we took advantage of these novel techniques in a new cohort and detected more than 3000 proteins in CSF. We also increased the number of individuals to 609 individuals to replicate and refine existing subtypes, to test if the higher complexity allows us to uncover more AD subtypes, and to study underlying genetic factors of these subtypes.

In our previous studies, we compared CSF AD subtypes on *APOE* e4 carriership (the strongest genetic risk factor for sporadic AD)[9,14] and on AD polygenic risk scores. In the current study we further extent genetic analyses, and compared subtypes on AD risk variants from a recent GWAS[5]. Moreover, we enriched for the relatively rare *TREM2* R47H and R62H mutations, as these are associated with a resp. 2.3 and 1.4-fold increased risk of AD[5]. *TREM2* R47H and R62H are supposed to impair microglia activation in AD[15]. We therefore hypothesized that carriers of *TREM2* variants could group together in a subtype with impaired microglial activation. A small number of patients (n=6) were carrier of autosomal dominant mutation in PSEN1 or APP and we performed exploratory analysis to which subtypes these genetic variants were associated.

To further characterize the subtypes, we compared the groups on clinical measures, atrophy patterns and enrichment for biological processes, transcription factors, and cell type specificity.

This new large scale CSF proteomic study revealed five molecular AD subtypes. Three subtypes recapitulated our previously identified three subtypes (hyperplasticity, innate immune activation and blood-brain barrier dysfunction)[9]. We also identified two new AD subtypes: one with RNA dysregulation, and one with choroid plexus dysfunction. These subtypes were associated with distinct genetic risk profiles, further validating the biological underpinning of AD subtypes. The proteomic signatures associated with AD subtypes were present already in the preclinical stage and largely remained stable with increasing disease severity. Subtypes differed in the amount and pattern of cortical atrophy, cell type specific expression of proteins, vascular damage, progression rate from mild cognitive impairment to dementia and survival times in dementia. Our results highlight the importance of neuronal plasticity, microglial impairment, innate immune activation, RNA processing choroid plexus and blood-brain barrier dysfunction in AD pathogenesis, and provide a comprehensive resource that informs on which proteins and pathways are dysregulated in particular subtypes of AD patients.

Results

We analysed CSF samples from 609 individuals that were selected from Alzheimer Center Amsterdam related studies[16–19] (see online methods). Of this sample, 419 had AD as defined by abnormal CSF amyloid beta 1-42 (abeta42) levels across clinical stages (i.e., 107 with normal cognition, 103 with mild cognitive impairment (MCI), and 209 with dementia). The 187 controls were required to have normal cognition and normal CSF abeta42 and tau levels. CSF proteins from each sample were enzymatically digested and the peptides were labelled with tandem mass tags (TMT), fractionated, and analysed by LC-MS/MS (see online methods). A total of 3863 proteins was identified, of which 1309 were observed across all individuals. We then selected proteins that different between controls and AD (S-table 2). We repeated those analyses stratified on tau-levels or disease stage, because protein levels can change in a non-linear way with these variables[13,20]. This led to a total of 1058 AD-related proteins that were selected for cluster analyses (S-table 2). We then clustered AD individuals on AD-related proteins using the dual clustering approach 'non-negative matrix factorisation'[21] (figure 1). A particular strength of the algorithm is that individuals will per definition be allocated to one subtype, which is useful for diagnosis or patient stratification for trials.

Five AD subtypes with distinct differences in clinical, molecular and genetic characteristics

Patients' proteomic profiles clustered into 5 subtypes (figure 2a; S-table 3 for fit statistics): Subtypes 1, 2, and 5 recapitulated our previously detected subtypes with neuronal hyperplasticity (subtype 1), innate immune activation (subtype 2) and blood-brain barrier dysfunction (subtype 5). Additionally, two new subtypes emerged: one with RNA dysregulation (subtype 3) and one with choroid plexus dysfunction (subtype 4). The next sections discuss each subtype in detail on molecular, and genetic and clinical characteristics, here we briefly summarise these characteristics. Compared to controls, subtypes 1, 2 and 3 had increased CSF t- and p-tau levels, while subtype 4 and 5 had mostly normal tau levels (table 1). Subtypes differed in clinical stage, sex, and age, and so all subsequent analyses took these characteristics into account. Compared to controls, subtypes differed in rates of progression from MCI to dementia, with subtypes 2 and 5 having the highest risk, and subtype 4 the lowest (figure 3d), albeit not significantly different between subtypes (S-table 10a). Subtype 3 individuals with dementia had the most aggressive disease course of 5.6 years, which was shorter than subtype 1 with the longest survival time of 8.9 years (p=0.04; S-table 10b). These results suggests that different underlying molecular

aspects may explain variability in decline. Analysis of MRI scans in individuals with dementia (n=159) indicated that subtypes differed in the degree and anatomical location of cortical atrophy (figure 2c; S-table 9). All subtypes had a higher prevalence of the APOE e4 genotype than controls, and a higher AD polygenic risk score, supporting their underlying AD genetic risk architecture. Subtypes had, however, different AD genetic risk profiles.

We next examined the molecular processes associated with AD subtypes. For each subtype we compared the levels of 2878 proteins against the control group (figure 2b; S-table 5). Proteins with different levels between a subtype and the control group were included in enrichment analyses to study associated biological processes and transcription factors. In order to aid comparability with e.g., gene expression literature we report gene names for proteins (UniProt codes are listed in S-table5). Stratification by disease stage resulted in similar differences to controls (correlations between 0.85-0.98; S-figure 1; S-table5 columns CY-HU), providing further support that AD subtypes reflect specific disease traits[9,14].

Below, we discuss the most distinct biological processes types, cell types AD risk variants and atrophy patterns that were associated with each subtype. Detailed results are reported in the supplemental material.

Subtype 1: Hyperplasticity

Subtype 1 individuals (n=137, 32.7%) had compared to controls 827 proteins with increased CSF levels and 408 proteins with decreased levels. Of all subtypes, subtype 1 had the highest proportion of proteins specific for neurons, astrocytes, oligodendrocytes and oligodendrocyte precursor cells (figure 2c). Proteins with increased levels were associated with neuronal plasticity processes, including synapse assembly, axon guidance, neurogenesis, and gliogenesis (figure 2e, S-table 6). In addition, this neuronal hyperplasticity subtype had high BACE1 and amyloid beta 1-40 (abeta40) CSF levels, as well as high tau levels (table 1), like in our previous study[9]. While high tau levels were previously thought to reflect neuronal loss due to tangle formation, now more studies indicate that this may also reflect other processes[22,23]. For example, hyperactive neurons and astrocytes have been reported to surround plaques[24,25]. Neurons with increased activity secrete more amyloid as well as tau[26–30], and fragments of those proteins may in turn drive hyperplasticity through enhanced gene transcription[31]. Indeed, proteins increased in subtype 1 included were enriched for the transcription factors REST ($p_{adjusted}$=.018x10^{-13}, figure 1d, S-table7) and SUZ12 ($p_{adjusted}$ =.016x10^{-12}), which regulate plasticity related processes through repression of neuronal differentiation genes[32,33]. Previous studies pointed towards REST de-repression and increases of tau and plasticity related processes in AD brain tissue[8,34], iPSC neurons[35,36] and tau tangle carrying neurons[37]. Comparing subtype 1 increased proteins with those studies, we found an overlap 5 of 6 from[8], 65 of 173 from[35] and 46 of 127 from[37] (S-table 5). Moreover, with the higher number of proteins that we measured compared to our previous study, we found other mechanisms that may contribute to the plasticity response observed in this subtype. For example, the lysosomal protein PLD3 was highest in subtype 1. High PLD3 levels have been reported in dystrophic neurites associated with 'amyloid axonal spheroids' [38]. Such spheroids have previously been found to trigger axonal remodelling and local hyperactivity[38]. Furthermore, dystrophic neurites accumulate BACE1, which has been found to increase processing of APP[39], and may explain the elevated BACE1 and abeta levels in this subtype.

Next, we tested which AD risk variants[5] were more common in this subtype compared to controls. We found enrichment in subtype 1 for variants in *TREM2 R47H*, *LILRB2*, *RHOH*, *NCK2* and in *APP* (figure 2f; S-tables 8a, b). This subtype also included 3 of the 4 *PSEN1* carriers (S-table 8b). TREM2 is a transmembrane protein, of which the extracellular part binds ligands [15], including amyloid fibrils, that can activate microglia. The genetic variant *TREM2 R47H* has been associated with dampened microglia activation due to decreased ligand binding to TREM2's extracellular part[40,41]. *LILRB2* has been associated with a similar dampened immune activation as *TREM2*[42,43]. Furthermore, *RHOH* and *NCK2* are signalling molecules downstream from TREM2 that influence cytoskeleton rearrangement of microglia, necessary to enable migration towards pathogens and amyloid plaques[44]. Normally, activated microglia form a tight barrier around plaques that decreases plaque surface, thereby minimizing plaque contact with neurites[15,40,45]. When microglial activation is dampened, such as observed in carriers of *TREM2* variants, amyloid plaques are less compact, with toxic oligomers sticking out that could damage nearby neurites[46–48] and may lead to axonal dystrophy[47], which may trigger a plasticity response. *TREM2* has also been implicated in impaired microglial synaptic pruning, which could further contribute to the hyperplasticity signature observed in this subtype[49–53]. Such an excess of synapses was previously associated with milder atrophy on MRI in *TREM2* mouse models[49]. After analysing MRI in our data, we found that this subtype had the lowest degree of atrophy compared to the other subtypes (figure 3c; S-table9), and was restricted to the temporal and parietal lobes.

Together, our results provide further support for a hyperplasticity subtype in AD, which we now observe could be related to a dampened microglial response. Currently, therapies that boost TREM2 activation are in development[54]. We argue that individuals with this subtype may also respond to such treatments, even without carrying the *TREM2 R47H variant*.

Subtype 2: Innate immune activation

Subtype 2 individuals (n=124, 29.6%) had compared to controls 986 proteins with increased CSF levels and 506 with decreased levels. A high proportion of proteins increased in subtype 2 was specific to microglia. Proteins with increased levels were associated with innate immune activation, including regulation of cytokine production. These included proteins from the complement complex (C1QA, C1QB, C1QC, C1S and C1R), as well as APOE and LPL. This subtype also had increased levels of microglial TAM receptors AXL and MERTK, and GAS6 (a MERTK ligand), which can detect and engulf plaques[55,56]. We further found, for the first time, increased PYCARD levels specifically in subtype 2. PYCARD is also known as Apoptosis-associated speck-like protein containing a CARD (ASC), and is released by microglia with NLRP3 inflammasome activation[57,58]. PYCARD can form ASC specks[59], which are fibrils that worsen amyloid aggregation[58] and induce tau phosphorylation[60], providing a potential mechanism through which microglial activation may aggravate AD pathology. Indeed, subtype 2 individuals had higher p-tau levels than seen in subtype 1 (table 1). Other subtype 2 increased proteins were related to neuron-microglia signalling, such as CSF1, CSF1R, and CX3CL1. Neuro-immune signalling occurs during normal neuronal development, during which microglia prune immature synapses[61–65]. However, activated microglia near diffuse and neuritic plaques may lead to excessive synaptic pruning[62]. This could lead to exacerbated atrophy on MRI as shown in mouse models[66,67]. In line with those models, subtype 2 was one of the two subtypes with the most severe and widespread cortical atrophy on MRI compared to subtype 1, 3 and 5

(figure 3c; S-figure 2). Still, despite this severe atrophy, the levels of proteins related to neuroplasticity were increased in this subtype and these proteins overlapped with subtype 1, and were also enriched for the transcription factors REST and SUZ12. Possibly, the increase of plasticity related proteins may reflect an attempt to repair synaptic contacts, which succumbs in the presence of activated microglia. Alternatively, increased protein levels may reflect neuronal loss as a result of atrophy.

AD genetic variants associated with this subtype were *IDUA*, *CLNK*, and *SCIMP*, which are all implied in immune processes[5,68,69].

Together, these results give detailed insight into the innate immune activation AD subtype, and suggest that an overactive innate immune system worsens the disease.

Subtype 3: RNA dysregulation

Subtype 3 (n=24, 5.7%) emerged as a new subtype. Compared to controls this subtype had increased CSF levels for 516 proteins and decreased levels for 757 proteins. Proteins with increased levels were associated with cytoskeleton organisation, axonal transport, proteasome and protein folding (figure 2e; S-tables 5, 6). This subtype had the highest t-tau and NEFL CSF levels. BACE1 levels were higher than controls (table 1), but unlike subtypes 1 and 2, abeta40 levels were similar to controls, suggesting a different mechanism associated with higher BACE1 levels for this subtype. Proteins specifically increased in subtype 3 included heterogenous nuclear ribonucleoproteins (hnRNPs)[70] and other RNA binding proteins, which may point to RNA dysregulation. HNRNPs, are involved in maturation of pre-mRNAs, mRNA stabilisation during transport and local mRNA translation for many RNAs, including those important for cytoskeleton organisation [71,72]. Disruptions in HnRNPs and mRNA have been associated with tau tangles in previous proteomic studies[73]. Mislocalised hnRNPs could lead to mis- and/or cryptic splicing, resulting in dysfunctional proteins[70,74]. For example, cryptic splicing of STMN2 is a hallmark of TDP43 mislocalisation[75], resulting in shorter proteins and decreased STMN2 levels in tissue[76]. In our data STMN2 was detected in a subset (n=84), and subtype 3 specifically had decreased levels of STMN2 compared to controls. Transcription factors associated with subtype 3 increased proteins were KLF4 ($p_{adjusted}=0.02x10^{-15}$), which is associated with axon regeneration ability[77], as well as on TAF1 ($p_{adjusted}=0.008x10^{-13}$) and MYC ($p_{adjusted}=0.02x10^{-10}$), which are interacting factors in cell differentiation processes[78,79]. A previous study based on brain tissue gene expression found a similar AD subtype with increased TAF1 and MYC signalling and decreased synapse organisation[8].

When testing AD risk factors, we found that subtype 3 was enriched for *BIN1*, which is known as 'myc-box dependent interaction protein'. One of BIN1's functions is to physically inhibit MYC[80]. BIN1 mainly localises in axons and has many isoforms arising from splicing[80]. BIN1 mis-splicing has been associated with de-inhibition of MYC and cytoskeleton disruptions[81]. *TREM2* R62H was also associated with this subtype. On MRI this subtype showed relatively mild atrophy like subtype 1, but also encompassed atrophy in the superior frontal gyrus. Other genetic risk variants associated with subtype 3 included *SPDYE3*, involved in the cell-cycle, *SNX1*, important for endosome sorting, and *KAT8* a lysine acetyltransferase[5,69].

While RNA dysfunction has been mainly observed in frontotemporal dementia[82], we now find that these disruptions are associated with a specific AD subtype as well.

Subtype 4: Choroid plexus dysfunction

Subtype 4 (n=78, 18.6%) was another new subtype. Compared to controls, this subtype had increased CSF levels of 467 proteins, and decreased levels of 626 proteins. A high proportion of proteins increased in subtype 4 were specific to microglia and other immune cells. Moreover, a large subset of proteins with increased levels (45%) was associated with high expression in the lateral ventricle choroid plexus (S-table 5), including TTR, SPARC, and extracellular matrix proteins such as DCN, LUM and COLA12. Biological processes associated with subtype 4 included cell adhesion, BMP and SMAD pathways, which are involved in choroid plexus development[83] (figure 2e). The choroid plexus is located along the ventricles, where it produces CSF and is responsible for nutrient, lipid, and protein transfer across the blood-CSF interface[83]. It consists of a highly developed extracellular matrix that connects a dense vasculature to its epithelial cells[84]. On MRI, subtype 4 had the largest choroid plexus volume (figure 3b). Increased choroid plexus volume has been associated with inflammation and structural alterations by previous studies in AD[85–87]. Although this subtype most often had normal t-tau and p-tau levels (table 1), it had more severe atrophy in comparison to subtype 1, 3 and 5, and had specific atrophy in anterior cingulate areas (figure 3; S-figure 2). Proteins increased in subtype 4 were also enriched for fibroblasts (S-table 5). Fibroblasts produce extracellular matrix proteins, providing structural support to the choroid plexus[84]. Other proteins increased in subtype 4 included cytokines, such as CCL2, CCL21 and CCL15, which can attract monocytes and T lymphocytes [88]. Of note, proteins with *decreased* levels in subtype 4 were related to axonal outgrowth and synaptic plasticity (e.g., BDNF), in part overlapping with proteins increased in subtype 1 and 2 and also enriched for REST and SUZ12. This suggests that 2 subtype is also characterised by neuronal hypo-plasticity.

When testing AD genetic risk variants, we found enrichment in subtype 4 of *ABCA7*, *PICALM*, and *IL34*, and also with *CLNK*. While *ABCA7* and *IL34* are expressed in the choroid plexus [89,90], *PICALM* is expressed in the blood-brain barrier[91]. Both *ABCA7* and *PICALM* play a role in lipid metabolism[92,93]. Both have been associated with amyloid clearance in combination with LRP1 across the blood-brain barrier (*PICALM*)[94,95] or choroid plexus (*ABCA7*), or via lysosomal degradation[92,96–100]. *IL34* has been associated with impaired macrophage function[101], which could interfere with macrophage uptake of amyloid fibrils[102]. Of note, this subtype had lower levels of BACE1 and abeta40 than controls (table 1), suggesting *decreased* amyloid metabolism. This suggests that impaired clearance mechanisms underlie subtype 4, rather than amyloid overproduction which was observed in subtype 1 and 2.

Taken together, these results provide further support for choroid plexus dysfunction as another contributor to AD pathogenesis, for a specific subgroup of patients.

Subtype 5: Blood-brain barrier dysfunction

Subtype 5 (n=56, 13.4%) was highly similar to our previously identified BBB dysfunction subtype with increased levels of 640 proteins that included blood proteins such as albumin, fibrinogens, plasminogen, prothrombin, and many immunoglobulins such as IGG-1, which are all proteins that leak into the brain when the BBB is compromised[103,104]. Pathways associated

with increased proteins included blood coagulation, B cell mediated immunity, and acute inflammatory response. No transcription factor enrichment was observed for proteins with increased CSF levels. On MRI this subtype had more microbleeds than controls (p=.03), unlike the other subtypes (table 1). The majority of proteins associated with subtype 5 (1013, 61%) had, however, *decreased* CSF levels compared to controls, and these were associated with neuroplasticity and converged on transcription factors SUZ12 and REST. This suggests that the BBB subtype, like the choroid plexus subtype, has hypo-plasticity. Neuronal plasticity processes can be impaired by leakage of blood proteins, including fibrin, which were specifically increased in this subtype[104]. Furthermore, in this dataset, we could now detect changes in protein levels that were associated with pericytes, cells that normally cover capillaries. We also found altered levels of proteins associated with particular vascular cell types, such as lower levels of PDGFRB, CDH2 (N-cadherin), MFGE8 (medin), HTRA1, LAMB1 (laminin), EDN1, LRP1, and JAM3, as well as increased levels of CDH5 (VE-cadherin), ANXA3, ICAM1, AMBP, VWF and PTPRB (see S-table 5 for detailed vascular cell annotation). These have been all been associated with deposition of blood proteins in the parenchyma in previous studies[91,105–109]. The low PDGFRb levels we observed may reflect loss of pericytes, which is in line with brain tissue measures of PDGFRb in rodent models and post-mortem AD[105,110–112]. Alternatively, the low concentrations we observed in the BBB subtype could reflect loss of other vascular cells such as arterial smooth muscle cells, which normally express PDGFRb[108,113]. Furthermore, the BBB subtype had, unlike the choroid plexus subtype, *decreased* levels of LRP1, which could indicate reduced pericyte numbers, further impeding amyloid clearance across the BBB[114].

In terms of genetic risk, this subtype had the highest proportion of *APOE* e4 carriers, albeit not significantly from the other subtypes (table 1). This subtype was further enriched for *IL-34*, *ECHDC3* and *APP*. *IL-34* risk factor was also associated with the choroid plexus subtype, suggesting it contributes to AD pathogenesis through the vasculature[115]. *ECHDC3* has been associated with lipid metabolism[69]. Some variants in *APP* have been associated with vascular disruption and/or increased occurrence of cerebral amyloid angiopathy (CAA), by producing amyloid fragments that are more difficult to clear[116–118]. The notion that this subtype has BBB dysfunction, suggests that this *APP* variant may contribute to AD risk through vascular integrity.

Together, these data provide new insights into the underlying pathophysiological processes associated with the BBB dysfunction AD subtype.

Conclusion

In this study we dissected AD disease heterogeneity at a patient level using CSF proteomics, at a level of detail level that approaches the level of complexity achieved in tissue proteomics[3,6,8]. Our analyses led to a more in-depth characterisation of three previously identified CSF subtypes (i.e., hyperplasticity, innate immune activation and blood-brain barrier dysfunction), and we identified two new subtypes, one with RNA dysregulation, which showed the most aggressive disease course, and one with choroid plexus dysfunction. Notably, we found that each subtype was associated with distinct AD genetic risk factors, further validating that each CSF AD subtype reflects specific underlying molecular mechanisms. The subtypes also showed profound differences in cortical atrophy patterns, and survival times, underscoring their clinical relevance.

Given the distinct patterns of molecular processes and AD genetic risk profiles, it is likely that AD subtypes will require specific treatments. For example, subtype 1 individuals may benefit from TREM2 activating treatments, subtype 2 from innate immune inhibitors, subtype 3 from antisense oligonucleotides that restore RNA processing, subtype 4 from inhibition of monocyte infiltration and subtype 5 from cerebrovascular treatments. At the same time, potential harmful side effects arising from certain treatments may also depend on subtype. For example, while antibodies may more easily cross the blood-brain barrier in subtype 5, these individuals may be at increased risk for cerebral bleeding that can occur with some antibody treatments. Future studies should aim to (re)analyse proteomics in clinical trials to test whether particular treatments may have effects for specific subtypes only. To conclude, CSF based subtyping will be highly useful for the selection of individuals likely to benefit most from a specific therapeutic, either for a priori subject stratification for clinical trials, or for responder and side effect analysis.

Online Methods

Participants

This study included individuals from Alzheimer center Amsterdam related studies (i.e., Amsterdam Dementia Cohort (ADC)[16], EMIF-AD preclin AD[17] and 90+ studies[18]) and participants who co-enrolled in the ADC biobank and the EPAD study[19]. Participants were selected when they had CSF available and either normal cognition with normal CSF abeta42 as controls (n=187) or if they had abnormal CSF abeta42 (n=419). Among the group with abnormal amyloid levels 107 individuals had intact cognition (i.e., preclinical AD)[119], 103 had mild cognitive impairment (i.e., prodromal AD)[120] and 209 had dementia according to international consensus criteria[121–123]. In case more individuals met criteria for inclusion, preference was given to individuals with a known *TREM2* R47H (n=8) or R62H mutation (n=28; see details in *Genetic data* below), and to individuals without dementia who had clinical follow up (n=216). Information on mortality was available from the Dutch Municipal Register for ADC and EMIF-AD participants. One person for whom proteomics was measured was excluded from further analyses in this study, because this person had normal AD markers in CSF, but a clinical diagnosis of primary progressive aphasia. All studies were approved by local Medical Ethical Committees.

CSF collection

CSF was collected by lumber puncture between the L3/L4, L4/L5 or L5/S1 intervertebral space with a 25-gauge needle and syringe and collected in polypropylene tubes [16]. CSF sample processing and biobank storage at the Alzheimer center biobank at the department of Clinical Chemistry was performed according to international guidelines[124].

CSF sample preparation for tandem mass tag proteomics

Six hundred and ten samples were stored at -80 °C in seven 96 well plates, 100 µl CSF in each well. Each sample plate was thawed on ice, and 30µg protein (separate 96 well plates containing 40 µl CSF were used to do BCA protein assay on 2x10 µl CSF for protein concentration measurements) from each well was transferred to 1.5 ml protein Low-Bind tubes, and immediately frozen on dry ice. Quality control samples and TMT reference samples were collected from the first plate only since it contained an even mixture of all sample groups. From the first plate, 38 µg of CSF from 93 wells were mixed in a 4 ml glass vial on ice, and 30 µg protein were transferred to 110 1.5ml Protein Low-Bind tubes, 22 QC-samples and 88 TMT

reference samples, and immediately frozen on dry ice. All samples were lyophilized using a freeze dryer and kept at -80 °C prior to digestion.

Urea protein digestion were performed as follows. Each day until there were no more samples to process, 28 samples together with one QC sample and two reference samples were added 20 µl 8M Urea/20mM Methylamine, vortexed for 5 min at 1000 rpm and sonicated for 30s in ice cold water. The Urea solution was diluted with 20 µl 50mM TrisHCl/1mM CaCl2 pH 7.6, followed by cysteine reduction (0.4 µmol Dithiotreitol, 1h incubation at RT) and alkylation (1 µmol Iodoacetamide, 1h incubation in the dark at RT). To avoid protease alkylation, excess iodoacetamide were allowed to react with Dithiotreitol by adding 0.08 µmol of the reagent before diluting the Urea to 1M with 50mM TrisHCl/1mM CaCl2 pH 7.6. Trypsin digestions were performed for 16h at 37 °C after adding 0.6mg of the protease (porcine trypsin from Promega, GmbH, Mannheim, Germany).

Sample clean-up was performed using a reverse-phase Oasis 96-well HLB µElution Plate 30 µm (2 mg HLB sorbent, Waters, Milford, MA). After lyophilization, QC samples were resuspended in 25 µl 2% acetonitrile/0.5% formic acid. All other samples were resuspended in 20 µl 50mM HEPES buffer pH 8.2 (4-(2-Hydroxyethyl) piperazin-1-ylethanesulfonic acid) prior to TMT labelling. All samples were vortexed for 30s at 1500 rpm and sonicated for 30s in an ultrasonic bath.

Each reporter in a 5mg TMTpro 16plex reagent set were dissolved in 1ml anhydrous acetonitrile. The 610 samples were labelled in 44 experiments, where each experiment contained 14 samples and 2 reference samples. For each sample/reference sample, 20 µl label was added (the two reference samples were labelled with 126 and 134N in each experiment). The labelling reaction was allowed for 75min before it was stopped by adding 5 µl 5% hydroxylamine. The 16 labelled samples for each experiment were combined and lyophilized (about 240 µg protein), and approximately 150 µg were desalted using a reverse-phase Oasis 96-well HLB Elution Plate 30 µm (10 mg HLB sorbent, Waters, Milford, MA). After lyophilization, the 44 samples were dissolved in 150µl 10mM Ammonium formate pH 7.9, and 65µl were fractionated using an off-line HPLC (Agilent 1260 infinity, Agilent Technologies, Santa Clara, California, USA) equipped with a reversed phase column (XSelect CSH C18, 130Å, 3.5 µm, 1 x 150 mm from Waters Corp, Milford, Massachusetts, USA). Using high pH reversed phase chromatography, peptides were separated during a biphasic Acetonitrile (ACN) gradient from two HPLC pumps (flow rate of 50 µl/min). Solvent A and B were 10mM Ammonium formate pH 7.9 in water and 90% ACN/10% water respectively. The gradient composition was 5%B during trapping (2min) followed by 5-12%B over 1 min, 12–44%B for the next 35min, 44-70%B over 10 min, and 70–95%B over 2min. Elution of very hydrophobic peptides and conditioning of the column were performed for 5 minutes isocratic elution with 95%B and 12 minutes isocratic elution with 5%B respectively. Peptide were collected in a 500 µl protein Low-Bind 96-well plate during peptide elution, 10 fractions were collected. The first fraction was collected in two wells from 5 to 16 min (5.5 min/well, merged into one fraction), the next 8 fractions (2.7 min/fraction) were collected between 16 and 37.6 min, and the last fraction collected in two wells between 37.6 and 53.6 min (8 min/well, merged into one fraction). Fractions were lyophilized and resuspended in 10 µl 2% ACN/0.5% formic acid (FA), and peptide concentrations were measured on NanoDrop UV-Vis spectrophotometer (Thermo Scientific, Waltham, MA, USA) prior to LC-MS/MS analysis.

Liquid Chromatography (LC) Tandem Mass Spectrometry (MS) Analysis
About 0.5ug tryptic TMT labelled peptides were injected into an Ultimate 3000 RSLC system (Thermo Fisher Scientific, Waltham, MA, USA) connected online to a Exploris 480 mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) equipped with EASY-spray nano-electrospray ion source. Peptides were desalted on a pre-column (Acclaim PepMap 100, 2cm x 75µm ID nanoViper column, packed with 3µm C18 beads) at a flow rate of 5µl/min for 5 min with 0.1% trifluoroacetic acid, before separation on a 50 cm analytical column (PepMap RSLC, 50cm x 75 µm ID EASY-spray column, packed with 2µm C18 beads). During a biphasic ACN gradient from two nanoflow UPLC pumps (solvent A and B were 0.1% FA (vol/vol) in water and 100% ACN respectively), peptides were separated through the reversed phase column at a flow rate of 200 nl/min. The gradient composition was 5%B during trapping (5min) followed by 5-8%B over 1 min, 8–30%B for the next 104min, 30-40%B over 15 min, and 40–80%B over 3min. Elution of very hydrophobic peptides and conditioning of the column was performed for 9 minutes isocratic elution with 80%B and 10 minutes isocratic elution with 5%B. The LC was controlled through Thermo Scientific SII for Xcalibur 1.6.

Peptides eluted from the column were detected in the Exploris 480 Mass Spectrometer (capillary temperature at 275 °C and Ion spray voltage at 2100V) with FAIMS (High field asymmetric waveform ion mobility spectrometry) enabled using two compensation voltages (CVs, -50V and -70V), and "Advanced Peak Determination" on. During each CV, the mass spectrometer was operated in the DDA-mode (data-dependent-acquisition) to automatically switch between one full scan MS and MS/MS acquisition. Instrument control was through Orbitrap Exploris 480 Tune 3.1 and Xcalibur 4.4. The cycle time was maintained at 1.5s/CV. The FAIMS filter performs gas-phase fractionation, enabling preferred accumulation of multiply charged ions to maximize the efficiency of DDA. FAIMS results in less precursor co-isolation, and cleaner MS2 spectra. MS spectra were acquired in the scan range 375-1500 m/z with resolution R = 60 000 at m/z 200, automatic gain control (AGC) target of 3e6 and a maximum injection time (IT) at auto (depending on transient length in the orbitrap). The most intense eluting peptides with charge states 2 to 6 and above an intensity threshold of 2e4 were sequentially isolated to standard target value of 2e5, or a maximum IT of 120 ms in the C-trap, and isolation width maintained at 0.7 m/z (quadrupole isolation), before fragmentation in the HCD (Higher-Energy Collision Dissociation). Fragmentation was performed with a normalized collision energy (NCE) of 32 %, and fragments were detected in the Orbitrap at a resolution of 45 000 at m/z 200, with first mass fixed at m/z 110. One MS/MS spectrum of a precursor mass was allowed before dynamic exclusion for 45s with "exclude isotopes" on. Lock-mass internal calibration was not enabled.

The resulting .raw files were processed using Proteome Discoverer 2.5. The database file used for the search using Sequest HT was Swiss-Prot with 20395 entries (version 20210413.fasta). The following modifications were defined in the database search: Precursor Mass Tolerance: 10 ppm, fragment Mass Tolerance: 0.02 Da, Static Peptide N-Terminus: TMTpro / +304.207 Da (Any N-Terminus), static modification: TMTpro / +304.207 Da (K), static modification: Carbamidomethyl (C), and dynamic modification for Methionine oxidation. Maximum of missed Cleavage Sites was set to 2, with a minimum peptide length of 6. The validation settings were set to 0.01 for target FDR for PSMs and peptides (strict) and 0.05 for relaxed. Peptides used were set to unique + razor. Reporter abundance was based on intensity with a co-isolation threshold of 50

and average reporter S/N threshold of 10. The output results files where then gathered and subjected to further processing.

Technical deviations may influence protein abundances across TMT experiments[125,126]. We normalized protein abundances according to the Internal Reference Scaling normalization procedure[127] for TMT proteomic data that uses the common pool reference channels to normalize values between plex experiments, adapted to scale according to median instead of the total sum to reduce sensitivity of outliers. Briefly, this is a two-step approach, with the first step normalizing grand total intensities for each of the 14 channels within an experiment to match these to the two reference channels. In the second step a correction factor is calculated based common pooled internal standards to normalize reporter ion intensities of proteins between TMT experiments. Internal standards were unavailable for 113 proteins, which were excluded for subsequent analyses. S-figure 3 illustrates that the correction effectively removed batch effects. After batch correction, protein values were log2 transformed, and then scaled according the mean and standard deviation of the control group, such that positive and negative values indicate higher and lower than normal. In total 3863 proteins were identified that had at least 1 observation. For, all proteins we report GENE names to aid comparisons with other AD subtyping literature using either proteomics or RNAseq data.

CSF ELISA measures

Abeta42, t-tau and p-tau were previously determined using ELISA assays from Innotest (Fujirebio, formerly Innogenetics) in the ADC[128,129], or with the Roche Elecsys System (n=15 from ADC and in EPAD[130]. In EMIF-AD preclinAD study amyloid status was determined based on the abeta42 and abeta40, t-tau and p-tau181 were measured with ELISAs from ADx Neurosciences/EUROIMMUN[131]. In EMIF-AD 90+ amyloid status was determined with visual read of [^{18}F] flutemetamol positron emission tomography[132]. For these individuals (n=22) tau levels were computed from the TMT MAPT measures, which correlated strongly (r=0.81, p<.001) with Innotest t-tau levels in the ADC cohort (formula: Inno t-tau = -309.16+0.01*MAPT). For tau categorization, we used t-tau values as these were available in all individuals, and correlated strongly with p-tau levels (r=0.94, p<.001). We used published cutoffs to label individuals as having a normal AD CSF profile or abnormal amyloid based on CSF[128–131,133]. Three individuals with normal cognition had normal amyloid and abnormal CSF t-tau levels, which were excluded from the present analyses, resulting in a final sample size of 187 controls and 419 individuals with abnormal amyloid. We standardized continuous amyloid 1-42, t-tau and p-tau181 values within specific assays according to the mean and standard deviation of controls to compare these values between subtypes. Finally, we measured BACE1, abeta40, and neurogranin with EUROIMMUN ELISA assays (Germany), NEFL with ADx NeuroSciences (Belgium) ELISA assay, and VAMP2 with a prototype assay developed by ADx Neurosciences (Belgium) with single-molecule array (Simoa) technology (Quanterix Corp, Billerica, USA) as previously described[134]. These measures were not included in clustering, but used as independent markers to compare between AD subtypes.

Genetic data

APOE genotyping was performed in blood as previously described[16,17,135,136]. A subset of 560 individuals we had genotyping data available (Illumina Global Screening Array, GSA). Details on quality control procedures were previously described[137], and for EPAD available on GitHub

(https://github.com/marioni-group/epad-gwas). Genotype vcf files were imputed using the TopMed reference panel[138]. Eighty-three genetic risk loci for AD were selected based on their previous genome-wide association with AD[5]. These SNPs were extracted from the genetic data based on rsID/and or base pair location in the genome. Protective SNPs (i.e., odds ratios below 1) were inverted, such that higher values indicate more AD risk for all SNPs.

MRI data

A subset 503 individuals had structural T1 weighted MRI available. To test if subtypes were characterized by different atrophy patterns, we restricted analyses to subtypes in the dementia stage (n=159 and 160 controls), because in that stage atrophy is most pronounced. Acquisition details were previously described[17,132,139,140]. Cortical thickness, hippocampal volume, choroid plexus volume and total intracranial volumes were estimated with FreeSurfer version 7.1.1 (http://surfer.nmr.mgh.harvard.edu/)[141]. Cortical thickness and volumetric estimates were summarised in anatomical regions as defined by the FreeSurfer implementation of the Desikan-Killiany atlas. Choroid plexus and hippocampal volumes were adjusted for total intracranial volume to adjust interindividual differences in head size. Furthermore, microbleeds were counted on T2* sequences by an experienced neuroradiologist and defined as small round hypointense foci up to 10mm in the brain parenchyma[17,132,139,140].

AD subtype discovery

Our objective was to identify subtypes *within* AD, and so we first selected proteins that were related to AD. For this, we compared all AD individuals to controls on CSF levels of proteins that were observed in the complete sample with Kruskal-Wallis tests. Because previous studies have indicated that AD related alterations of CSF protein levels may depend on cognitive states and/or tau status in a non-linear way[9,13,142], we repeated these analyses stratified for these factors. This resulted in 1058 proteins that were selected for clustering with non-negative matrix factorization as implemented in the 'NMF' package[143] v0.25 in R version 4.2.2. "Bird Hippie". We followed the procedure as in our previous study[9]. Briefly, proteins were first scaled to have positive values between 1 and 2, keeping relative values intact. Next, we performed 30 different runs of NMF to determine the number of clusters (i.e., subtypes) that best described the data. We tested up to 10 clusters, and found that 5 clusters showed an optimal balance of a high co-phonetic coefficient, at least 2-fold improved fit over lower clustering solution as compared to improvement by random cluster solution, and a silhouette score of >0.5 (supplemental table 3). We then labelled each individual patient according to the subtype that best matched their proteomic profile. Patient level subtype clusters were visualized by projecting the NMF subtype scores to a UMAP embedding, via construction of a k-nearest neighbor graph using the 'uwot' R package (v 0.1.14). Patient level subtype labels provide the basis for all subsequent post-hoc comparisons, as described in the next sections.

Biological characterization of AD subtypes

We characterized the biological processes associated with AD subtypes, by comparing the subtypes on CSF protein levels of all available proteins, including in addition to the fully observed proteins also proteins with missing values when they had at least 5 observations available in each subtype group (2907 proteins in total). For this we used linear models with participant subtype status as predictors and protein levels as outcomes. We repeated analyses correcting for age and sex, and stratifying according to cognitive state to determine influence of

these factors on the results. All subtypes were compared to the control group, as well as to each other, and results from all comparisons are reported in supplementary table 5. In the main manuscript we report results of comparisons to controls only. To test the biological pathways associated with subtypes, we performed pathway enrichment analyses for biological processes from the Gene Ontology (GO) release 2022-01-13 as accessed by Panther[144] version 16.0, for the proteins that were associated with each subtype (i.e., differed from controls with $p<.05$), separately for increased and decreased alterations. Hypergeometric Fisher's exact test were used for pathway enrichment, and pathway p values were corrected for multiple testing with the false discovery rate procedure. We further tested if AD subtype related proteins were associated with potential up stream transcription factors from the CHEA and ENCODE databases through ENRICHR[145]. We further annotated proteins for cell type specificity according to the Human Protein Atlas (https://www.proteinatlas.org), and the RNAseq Barres database[146]; for specific vascular cell types with Garcia et al., (2022)[108] and Yang et al., (2022)[91]; for choroid plexus associations according to the Harmonizome database[147]; for REST signalling associations based on Meyer et al., (2019) [35] and Otero-Garcia et al., (2022)[148]; for blood-brain barrier dysfunction according to Dayon et al., (2019)[103]; and for CSF pathway panels informed by on tissue proteomics to Higginbotham et al, (2020)[4].

Post hoc comparisons between subtypes on clinical, MRI and genetic characteristics
We performed post-hoc tests to characterize AD subtypes clinically and biologically with Chi^2 tests for discrete variables (sex and APOE e4 genotype), and with linear regression models for continuous variables correcting for age and sex when applicable. Subtype differences in time to progress to dementia was tested with Cox proportional hazard models, and restricted to individuals the prodromal stage for reasons of statistical power. Subtype differences in survival times were also tested with Cox proportional hazard models, and restricted to individuals in the dementia stage for reasons of statistical power. Subtype differences with controls on genetic variants (i.e., SNPs, as continuous outcome) were tested with linear regression models, taking imputation uncertainty into account when possible, and repeated including age and sex as covariates. Subtype differences in hippocampal volume was tested in all stages, because this structure is altered in very early clinical stages.

References

1. Collaborators, G. 2019 D. F. *et al.* Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Heal* 7, e105–e125 (2022).

2. Montine, T. J. *et al.* National Institute on Aging–Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathologica* 123, 1–11 (2011).

3. 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* 1–13 (2022) doi:10.1038/s41593-021-00999-y.

4. Higginbotham, L. *et al.* Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic Alzheimer's disease. *Sci Adv* 6, eaaz9360 (2020).

5. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* 1–25 (2022) doi:10.1038/s41588-022-01024-z.

6. 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).

7. Hondius, D. C. *et al.* Profiling the human hippocampal proteome at all pathologic stages of Alzheimer's disease. *Alzheimer's & Dementia* 12, 654–668 (2016).

8. Caldwell, A. B. *et al.* Transcriptomic profiling of sporadic Alzheimer's disease patients. *Mol Brain* 15, 83 (2022).

9. Tijms, B. M. *et al.* Pathophysiological subtypes of Alzheimer's disease based on cerebrospinal fluid proteomics. *Brain* 143, 3776–3792 (2020).

10. Haeberlein, S. B. *et al.* Two Randomized Phase 3 Studies of Aducanumab in Early Alzheimer's Disease. *J Prev Alzheimer's Dis* 1–14 (2022) doi:10.14283/jpad.2022.30.

11. Dyck, C. H. van *et al.* Lecanemab in Early Alzheimer's Disease. *New Engl J Med* (2022) doi:10.1056/nejmoa2212948.

12. Egan, M. F. *et al.* Randomized Trial of Verubecestat for Prodromal Alzheimer's Disease. *New England Journal of Medicine* 380, 1408–1420 (2019).

13. Visser, P. J. *et al.* Cerebrospinal fluid tau levels are associated with abnormal neuronal plasticity markers in Alzheimer's disease. *Mol Neurodegener* 17, (2022).

14. Tijms, B. *et al.* CSF Proteomic Alzheimer's Disease-Predictive Subtypes in Cognitively Intact Amyloid Negative Individuals. *Proteomes* 9, 36 (2021).

15. Ulrich, J. D., Ulland, T. K., Colonna, M. & Holtzman, D. M. Elucidating the Role of TREM2 in Alzheimer's Disease. *Neuron* 94, 237–248 (2017).

16. Flier, W. M. van der *et al*. Optimizing Patient Care and Research: The Amsterdam Dementia Cohort. *J Alzheimer's Dis* 41, 313–327 (2014).

17. Konijnenberg, E. *et al.* The EMIF-AD PreclinAD study: study design and baseline cohort overview. *Alzheimer's Research & Therapy* 10, S85 (2018).

18. Legdeur, N. *et al.* Resilience to cognitive impairment in the oldest-old: design of the EMIF-AD 90+ study. *Bmc Geriatr* 18, 289 (2018).

19. Ritchie, C. W. *et al.* The European Prevention of Alzheimer's Dementia (EPAD) Longitudinal Cohort Study: Baseline Data Release V500.0. *J Prev Alzheimer's Dis* 7, 8–13 (2020).

20. Leon, M. J. D. *et al.* The nonlinear relationship between cerebrospinal fluid Aβ42 and tau in preclinical Alzheimer's disease. *PLoS ONE* 13, e0191240 (2018).

21. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–791 (1999).

22. Karch, C. M., Jeng, A. T. & Goate, A. M. Extracellular Tau Levels Are Influenced by Variability in Tau That Is Associated with Tauopathies*. *J Biol Chem* 287, 42751–42762 (2012).

23. Sato, C. *et al.* Tau Kinetics in Neurons and the Human Central Nervous System. *Neuron* 97, 1284-1298.e7 (2018).

24. Busche, M. A. *et al*. Clusters of Hyperactive Neurons Near Amyloid Plaques in a Mouse Model of Alzheimer's Disease. *Science (New York, N.Y.)* 321, 1686–1689 (2008).

25. Kuchibhotla, K. V., Lattarulo, C. R., Hyman, B. T. & Bacskai, B. J. Synchronous Hyperactivity and Intercellular Calcium Waves in Astrocytes in Alzheimer Mice. *Science* 323, 1211–1215 (2009).

26. Pooler, A. M., Phillips, E. C., Lau, D. H. W., Noble, W. & Hanger, D. P. Physiological release of endogenous tau is stimulated by neuronal activity. *EMBO reports* 14, 389–394 (2013).

27. Yamada, K. *et al.* Neuronal activity regulates extracellular tau in vivo. *The Journal of Experimental Medicine* 211, 387–393 (2014).

28. Kaeser, S. A. *et al.* CSF p-tau increase in response to Aβ-type and Danish-type cerebral amyloidosis and in the absence of neurofibrillary tangles. *Acta Neuropathol* 143, 287–290 (2022).

29. Cirrito, J. R. *et al.* Synaptic Activity Regulates Interstitial Fluid Amyloid-β Levels In Vivo. *Neuron* 48, 913–922 (2005).

30. Kamenetz, F. *et al.* APP Processing and Synaptic Function. *Neuron* 37, 925–937 (2003).

31. Zott, B. *et al.* A vicious cycle of β amyloid-dependent neuronal hyperactivation. *Science (New York, N.Y.)* 365, 559–565 (2019).

32. Tsai, M.-C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science (New York, N.Y.)* 329, 689–693 (2010).

33. Ballas, N. & Mandel, G. The many faces of REST oversee epigenetic programming of neuronal genes. *Current Opinion in Neurobiology* 15, 500–506 (2005).

34. Lu, T. *et al.* REST and stress resistance in ageing and Alzheimer's disease. *Nature* 507, 448–454 (2014).

35. Meyer, K. *et al.* REST and Neural Gene Network Dysregulation in iPSC Models of Alzheimer's Disease. *Cell reports* 26, 1112-1127.e9 (2019).

36. Caldwell, A. B. *et al.* Dedifferentiation and neuronal repression define familial Alzheimer's disease. *Science Advances* 6, eaba5933 (2020).

37. Otero-Garcia, M. *et al.* Single-soma transcriptomics of tangle-bearing neurons in Alzheimer's disease reveals the signatures of tau-associated synaptic dysfunction. 2, 370–53 (2020).

38. Yuan, P. *et al.* PLD3 affects axonal spheroids and network defects in Alzheimer's disease. *Nature* 1–10 (2022) doi:10.1038/s41586-022-05491-6.

39. Sadleir, K. R. *et al.* Presynaptic dystrophic neurites surrounding amyloid plaques are sites of microtubule disruption, BACE1 elevation, and increased Aβ generation in Alzheimer's disease. *Acta Neuropathologica* 132, 235–256 (2016).

40. McQuade, A. *et al.* Gene expression and functional deficits underlie TREM2-knockout microglia responses in human models of Alzheimer's disease. *Nat Commun* 11, 5370 (2020).

41. Cosker, K. *et al.* Microglial signalling pathway deficits associated with the patient derived R47H TREM2 variants linked to AD indicate inability to activate inflammasome. *Sci Rep-uk* 11, 13316 (2021).

42. Zhao, P. *et al.* LILRB2-mediated TREM2 signaling inhibition suppresses microglia functions. *Mol Neurodegener* 17, 44 (2022).

43. Strooper, B. D. & Karran, E. The Cellular Phase of Alzheimer's Disease. *Cell* 164, 603–615 (2016).

44. Romero-Molina, C., Garretti, F., Andrews, S. J., Marcora, E. & Goate, A. M. Microglial efferocytosis: Diving into the Alzheimer's disease gene pool. *Neuron* 110, 3513–3533 (2022).

45. Zhao, R., Hu, W., Tsai, J., Li, W. & Gan, W.-B. Microglia limit the expansion of β-amyloid plaques in a mouse model of Alzheimer's disease. *Mol Neurodegener* 12, 47 (2017).

46. Gratuze, M., Leyns, C. E. G. & Holtzman, D. M. New insights into the role of TREM2 in Alzheimer's disease. 1–16 (2018) doi:10.1186/s13024-018-0298-9.

47. Yuan, P. *et al.* TREM2 Haplodeficiency in Mice and Humans Impairs the Microglia Barrier Function Leading to Decreased Amyloid Compaction and Severe Axonal Dystrophy. *Neuron* 90, 724–739 (2016).

48. Boon, B. D. C. *et al.* The coarse-grained plaque: a divergent Aβ plaque-type in early-onset Alzheimer's disease. *Acta Neuropathologica* 1–20 (2020) doi:10.1007/s00401-020-02198-8.

49. Gratuze, M. *et al.* Impact of TREM2R47H variant on tau pathology-induced gliosis and neurodegeneration. *Journal of Clinical Investigation* 1–43 (2020) doi:10.1172/jci138179.

50. Sheng, L. *et al.* Microglial Trem2 induces synaptic impairment at early stage and prevents amyloidosis at late stage in APP/PS1 mice. *Faseb J* 33, 10425–10442 (2019).

51. Filipello, F. *et al.* The Microglial Innate Immune Receptor TREM2 Is Required for Synapse Elimination and Normal Brain Connectivity. *Immunity* 48, 979-991.e8 (2018).

52. Qin, Q., Wang, M., Yin, Y. & Tang, Y. The Specific Mechanism of TREM2 Regulation of Synaptic Clearance in Alzheimer's Disease. *Front Immunol* 13, 845897 (2022).

53. Stoiljkovic, M., Gutierrez, K. O., Kelley, C., Horvath, T. L. & Hajós, M. TREM2 Deficiency Disrupts Network Oscillations Leading to Epileptic Activity and Aggravates Amyloid-β-Related Hippocampal Pathophysiology in Mice. *J Alzheimer's Dis* 88, 837–847 (2022).

54. Lengerich, B. van *et al.* A TREM2-activating antibody with a blood–brain barrier transport vehicle enhances microglial metabolism in Alzheimer's disease models. *Nat Neurosci* 1–14 (2023) doi:10.1038/s41593-022-01240-0.

55. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* 169, 1276-1290.e17 (2017).

56. Huang, Y. *et al.* Microglia use TAM receptors to detect and engulf amyloid beta plaques. *Nat Immunol* 22, 586–594 (2021).

57. Heneka, M. T. *et al.* NLRP3 is activated in Alzheimer's disease and contributes to pathology in APP/PS1 mice. *Nature* 493, 674–678 (2013).

58. Venegas, C. *et al.* Microglia-derived ASC specks cross-seed amyloid-β in Alzheimer's disease. *Nature* 552, 355–361 (2017).

59. Ransohoff, R. M. Specks of insight into Alzheimer's disease. *Nature* 552, 342–343 (2017).

60. Ising, C. *et al.* NLRP3 inflammasome activation drives tau pathology. *Nature* 1–24 (2019) doi:10.1038/s41586-019-1769-z.

61. Stevens, B. *et al.* The Classical Complement Cascade Mediates CNS Synapse Elimination. *Cell* 131, 1164–1178 (2007).

62. Hong, S. *et al.* Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science (New York, N.Y.)* 352, 712–716 (2016).

63. Jiang, S. & Bhaskar, K. Dynamics of the Complement, Cytokine, and Chemokine Systems in the Regulation of Synaptic Function and Dysfunction Relevant to Alzheimer's Disease. *Journal of Alzheimer's Disease* 57, 1123–1135.

64. Paolicelli, R. C. *et al.* Synaptic Pruning by Microglia Is Necessary for Normal Brain Development. *Science* 333, 1456–1458 (2011).

65. Hu, B. *et al.* Insights Into the Role of CSF1R in the Central Nervous System and Neurological Disorders. *Front Aging Neurosci* 13, 789834 (2021).

66. Gorlovoy, P., Larionov, S., Pham, T. T. H. & Neumann, H. Accumulation of tau induced in neurites by microglial proinflammatory mediators. *Faseb J* 23, 2502–2513 (2009).

67. Maphis, N. *et al.* Reactive microglia drive tau pathology and contribute to the spreading of pathological tau in the brain. *Brain : a journal of neurology* 138, 1738–1755 (2015).

68. Tsai, A. P. *et al.* PLCG2 is associated with the inflammatory response and is induced by amyloid plaques in Alzheimer's disease. *Genome Med* 14, 17 (2022).

69. Tábuas-Pereira, M., Santana, I., Guerreiro, R. & Brás, J. Alzheimer's Disease Genetics: Review of Novel Loci Associated with Disease. *Curr Genetic Medicine Reports* 8, 1–16 (2020).

70. Sidhu, R., Gatt, A., Fratta, P., Lashley, T. & Bampton, A. HnRNP K mislocalisation in neurons of the dentate nucleus is a novel neuropathological feature of neurodegenerative disease and ageing. *Neuropath Appl Neuro* 48, e12793 (2022).

71. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum Genet* 135, 851–867 (2016).

72. Costa, C. J. & Willis, D. E. To the end of the line: Axonal mRNA transport and local translation in health and neurodegenerative disease. *Developmental Neurobiology* 78, 209–220 (2017).

73. Kavanagh, T., Halder, A. & Drummond, E. Tau interactome and RNA binding proteins in neurodegenerative diseases. *Mol Neurodegener* 17, 66 (2022).

74. Triantopoulou, N. & Vidaki, M. Local mRNA translation and cytoskeletal reorganization: Mechanisms that tune neuronal responses. *Front Mol Neurosci* 15, 949096 (2022).

75. Melamed, Z. *et al.* Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. *Nat Neurosci* 22, 180–190 (2019).

76. Yu, Q.-S., Feng, W.-Q., Shi, L.-L., Niu, R.-Z. & Liu, J. Integrated Analysis of Cortex Single-Cell Transcriptome and Serum Proteome Reveals the Novel Biomarkers in Alzheimer's Disease. *Brain Sci* 12, 1022 (2022).

77. Moore, D. L. *et al.* KLF Family Members Regulate Intrinsic Axon Regeneration Ability. *Science* 326, 298–301 (2009).

78. Wei, Y. *et al.* Multiple direct interactions of TBP with the MYC oncoprotein. *Nat Struct Mol Biol* 26, 1035–1043 (2019).

79. Majd, S., Power, J. & Majd, Z. Alzheimer's Disease and Cancer: When Two Monsters Cannot Be Together. *Front Neurosci-switz* 13, 155 (2019).

80. Tan, M.-S., Yu, J.-T. & Tan, L. Bridging integrator 1 (BIN1): form, function, and Alzheimer's disease. *Trends Mol Med* 19, 594–603 (2013).

81. Damián-Zamacona, S. *et al.* Cell survival regulation during receptor-mediated endocytosis of chemically-modified lipoproteins associated to the formation of an Amphiphysin 2 (Bin1)/c-Myc complex. *Biochem Bioph Res Co* 505, 365–371 (2018).

82. Mehta, P. R., Brown, A.-L., Ward, M. E. & Fratta, P. The era of cryptic exons: implications for ALS-FTD. *Mol Neurodegener* 18, 16 (2023).

83. Liddelow, S. A. Development of the choroid plexus and blood-CSF barrier. 1–13 (2015) doi:10.3389/fnins.2015.00032/abstract.

84. Johanson, C. *et al.* The Distributional Nexus of Choroid Plexus to Cerebrospinal Fluid, Ependyma and Brain. *Toxicol Pathol* 39, 186–212 (2011).

85. Choi, J. D. *et al.* Choroid Plexus Volume and Permeability at Brain MRI within the Alzheimer Disease Clinical Spectrum. *Radiology* 212400 (2022) doi:10.1148/radiol.212400.

86. Serot, J.-M., Béné, M.-C., Foliguet, B. & Faure, G. C. Morphological alterations of the choroid plexus in late-onset Alzheimer's disease. *Acta Neuropathol* 99, 105–108 (2000).

87. Čarna, M. *et al.* Pathogenesis of Alzheimer's disease: Involvement of the choroid plexus. *Alzheimer's Dementia* (2023) doi:10.1002/alz.12970.

88. Cui, J. *et al.* Inflammation of the Embryonic Choroid Plexus Barrier following Maternal Immune Activation. *Dev Cell* 55, 617-628.e6 (2020).

89. Wang, Y. *et al.* IL-34 is a tissue-restricted ligand of CSF1R required for the development of Langerhans cells and microglia. *Nat Immunol* 13, 753–760 (2012).

90. Mizuno, T. *et al.* Interleukin-34 Selectively Enhances the Neuroprotective Effects of Microglia to Attenuate Oligomeric Amyloid-β Neurotoxicity. *Am J Pathology* 179, 2016–2027 (2011).

91. Yang, A. C. *et al.* A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* 1–8 (2022) doi:10.1038/s41586-021-04369-3.

92. Bossaerts, L., Cacace, R. & Broeckhoven, C. V. The role of ATP-binding cassette subfamily A in the etiology of Alzheimer's disease. *Mol Neurodegener* 17, 31 (2022).

93. Ando, K. *et al.* PICALM and Alzheimer's Disease: An Update and Perspectives. *Cells* 11, 3994 (2022).

94. Kadry, H., Noorani, B. & Cucullo, L. A blood–brain barrier overview on structure, function, impairment, and biomarkers of integrity. *Fluids Barriers Cns* 17, 69 (2020).

95. Sweeney, M. D., Sagare, A. P. & Zlokovic, B. V. Blood–brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nature reviews. Neurology* 14, 133–150 (2018).

96. Gonzales-Marrero, I. Choroid plexus dysfunction impairs beta-amyloid clearance in a triple transgenic mouse model of Alzheimer's disease. *Front Cell Neurosci* 9, 17 (2015).

97. Kanekiyo, T. & Bu, G. The low-density lipoprotein receptor-related protein 1 and amyloid-β clearance in Alzheimer's disease. *Front Aging Neurosci* 6, 93 (2014).

98. Strazielle, N. & Ghersi-Egea, J. F. Physiology of Blood–Brain Interfaces in Relation to Brain Disposition of Small Compounds and Macromolecules. *Mol Pharmaceut* 10, 1473–1491 (2013).

99. Ruzali, W. A. W., Kehoe, P. G. & Love, S. LRP1 expression in cerebral cortex, choroid plexus and meningeal blood vessels: Relationship to cerebral amyloid angiopathy and APOE status. *Neurosci Lett* 525, 123–128 (2012).

100. Zhao, Z. *et al.* Central role for PICALM in amyloid-β blood-brain barrier transcytosis and clearance. *Nature Neuroscience* 18, 978–987 (2015).

101. Chitu, V., Gokhan, Ş., Nandi, S., Mehler, M. F. & Stanley, E. R. Emerging Roles for CSF-1 Receptor and its Ligands in the Nervous System. *Trends Neurosci* 39, 378–393 (2016).

102. Zuroff, L. R. *et al.* Effects of IL-34 on Macrophage Immunological Profile in Response to Alzheimer's-Related Aβ42 Assemblies. *Front Immunol* 11, 1449 (2020).

103. Dayon, L. *et al.* Proteomes of Paired Human Cerebrospinal Fluid and Plasma: Relation to Blood–Brain Barrier Permeability in Older Adults. *Journal of Proteome Research* 18, 1162–1174 (2019).

104. Zlokovic, B. V. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. 1–16 (2011) doi:10.1038/nrn3114.

105. Bell, R. D. *et al.* Pericytes Control Key Neurovascular Functions and Neuronal Phenotype in the Adult Brain and during Brain Aging. *Neuron* 68, 409–427 (2010).

106. Armulik, A. *et al.* Pericytes regulate the blood–brain barrier. *Nature* 468, 557–561 (2010).

107. Carrano, A. *et al.* Neuroinflammation and Blood-Brain Barrier Changes in Capillary Amyloid Angiopathy. *Neurodegener Dis* 10, 329–331 (2012).

108. Garcia, F. J. *et al.* Single-cell dissection of the human brain vasculature. *Nature* 1–11 (2022) doi:10.1038/s41586-022-04521-7.

109. Procter, T. V., Williams, A. & Montagne, A. Interplay between brain pericytes and endothelial cells in dementia. *Am J Pathology* 191, 1917–1931 (2021).

110. Montagne, A. *et al.* Pericyte degeneration causes white matter dysfunction in the mouse central nervous system. *Nat Med* 24, 326–337 (2018).

111. Mäe, M. A. *et al.* Single-Cell Analysis of Blood-Brain Barrier Response to Pericyte Loss. *Circ Res* 128, e46–e62 (2021).

112. Daneman, R., Zhou, L., Kebede, A. A. & Barres, B. A. Pericytes are required for blood–brain barrier integrity during embryogenesis. *Nature* 468, 562–566 (2010).

113. Kelleher, J. *et al.* Patient-Specific iPSC Model of a Genetic Vascular Dementia Syndrome Reveals Failure of Mural Cells to Stabilize Capillary Structures. *Stem Cell Rep* 13, 817–831 (2019).

114. Sagare, A. P. *et al.* Pericyte loss influences Alzheimer-like neurodegeneration in mice. *Nat Commun* 4, 2932 (2013).

115. Ma, D. *et al.* TGF-β induced by interleukin-34-stimulated microglia regulates microglial proliferation and attenuates oligomeric amyloid β neurotoxicity. *Neurosci Lett* 529, 86–91 (2012).

116. Ristori, E., Donnini, S. & Ziche, M. New Insights Into Blood-Brain Barrier Maintenance: The Homeostatic Role of β-Amyloid Precursor Protein in Cerebral Vasculature. *Front Physiol* 11, 1056 (2020).

117. Greenberg, S. M. *et al.* Cerebral amyloid angiopathy and Alzheimer disease — one peptide, two pathways. *Nature Publishing Group* 1–13 (2019) doi:10.1038/s41582-019-0281-2.

118. Sweeney, M. D., Zhao, Z., Montagne, A., Nelson, A. R. & Zlokovic, B. V. Blood-Brain Barrier: From Physiology to Disease and Back. *Physiological Reviews* 99, 21–78 (2019).

119. Sperling, R. A. *et al.* 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. *Alzheimer's & Dementia* 7, 280–292 (2011).

120. Petersen, R. C. *et al.* Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology* 56, 303 (1999).

121. Mckhann, G. M. *et al.* 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. *Alzheimer's & Dementia* 7, 263–269 (2011).

122. McKhann, G. *et al.* 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, 939–939 (1984).

123. Ossenkoppele, R. *et al.* Research Criteria for the Behavioral Variant of Alzheimer Disease. *Jama Neurol* 79, (2022).

124. Campo, M. del *et al.* Recommendations to standardize preanalytical confounding factors in Alzheimer's and Parkinson's disease cerebrospinal fluid biomarkers: an update. *Biomarkers in Medicine* 6, 419–430 (2012).

125. Weiner, S. *et al.* Optimized sample preparation and data analysis for TMT proteomic analysis of cerebrospinal fluid applied to the identification of Alzheimer's disease biomarkers. *Clin Proteom* 19, 13 (2022).

126. Seyfried, N. T. *et al.* A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease. *Cell Systems* 4, 60-72.e4 (2017).

127. Plubell, D. L. *et al.* Extended Multiplexing of Tandem Mass Tags (TMT) Labeling Reveals Age and High Fat Diet Specific Proteome Changes in Mouse Epididymal Adipose Tissue*. *Mol Cell Proteomics* 16, 873–890 (2017).

128. Tijms, B. M. *et al.* Unbiased Approach to Counteract Upward Drift in Cerebrospinal Fluid Amyloid-β 1–42 Analysis Results. *Clinical Chemistry* 64, 576–585 (2018).

129. Mulder, C. *et al.* Amyloid- (1-42), Total Tau, and Phosphorylated Tau as Cerebrospinal Fluid Biomarkers for the Diagnosis of Alzheimer Disease. *Clinical Chemistry* 56, 248–253 (2010).

130. Ingala, S. *et al.* Application of the ATN classification scheme in a population without dementia: Findings from the EPAD cohort. *Alzheimer's Dementia* 17, 1189–1204 (2021).

131. Tomassen, J. *et al.* Abnormal cerebrospinal fluid levels of amyloid and tau are associated with cognitive decline over time in cognitively normal older adults: A monozygotic twin study. *Alzheimer's Dementia Transl Res Clin Interventions* 8, (2022).

132. Legdeur, N. *et al.* Associations of brain pathology, cognitive and physical markers with age in cognitively normal individuals aged 60-102 years. *The journals of gerontology. Series A, Biological sciences and medical sciences* (2019) doi:10.1093/gerona/glz180.

133. Willemse, E. A. J. *et al.* Diagnostic performance of Elecsys immunoassays for cerebrospinal fluid Alzheimer's disease biomarkers in a nonacademic, multicenter memory clinic cohort: The ABIDE project. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 10, 563–572 (2018).

134. Wesenhagen, K. E. J. *et al.* P-tau subgroups in AD relate to distinct amyloid production and synaptic integrity profiles. *Alzheimer's Res Ther* 14, 95 (2022).

135. Pelkmans, W. *et al.* Amyloid-β, cortical thickness, and subsequent cognitive decline in cognitively normal oldest-old. *Annals of Clinical and Translational Neurology* (2021) doi:10.1002/acn3.51273.

136. Saunders, T. S. *et al.* Interactions between apolipoprotein E, sex, and amyloid□beta on cerebrospinal fluid p□tau levels in the European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS). *Alzheimer's Dementia* 18, (2022).

137. Tesi, N. *et al.* Centenarian controls increase variant effect sizes by an average twofold in an extreme case–extreme control analysis of Alzheimer's disease. *Eur J Hum Genet* 27, 244–253 (2019).

138. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590, 290–299 (2021).

139. Dicks, E. *et al.* Gray matter network measures are associated with cognitive decline in mild cognitive impairment. *Neurobiology of Aging* 61, 198–206 (2018).

140. Lorenzini, L. *et al.* The Open-Access European Prevention of Alzheimer's Dementia (EPAD) MRI dataset and processing workflow. *Neuroimage Clin* 35, 103106 (2022).

141. Fischl, B. FreeSurfer. *NeuroImage* 62, 774–781 (2012).

142. Duits, F. H. *et al.* Synaptic proteins in CSF as potential novel biomarkers for prognosis in prodromal Alzheimer's disease. *Alzheimer's Research & Therapy* 10, 387 (2018).

143. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* 11, 367 (2010).

144. Thomas, P. D. *et al.* PANTHER: Making genome□scale phylogenetics accessible to all. *Protein Sci* 31, 8–22 (2022).

145. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14, 128–14 (2013).

146. Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* 89, 37–53 (2016).

147. Rouillard, A. D. *et al.* The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* 2016, baw100 (2016).

148. Otero-Garcia, M. *et al.* Molecular signatures underlying neurofibrillary tangle susceptibility in Alzheimer's disease. *Neuron* 110, 2929-2948.e8 (2022).

Acknowledgements/funding

BMT: ZonMW VIDI #09150171910068 ; the Dutch L'Óreal-UNESCO fellowship 2022 for women in science;
PJV: ZonMW Redefining AD #733050824 received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking (EPAD grant n° 115736; EPND grant n°101034344), and from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement number 101034344 (EPND). The IMI JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.
LV: received research support from ZonMW, Alzheimer Nederland and Stichting Dioraphte.
WF is supported by the Pasman stichting. WF is recipient of OTAPA, a collaboration project which is co-funded by the PPP Allowance made available by health-Holland, Top Sector Life Sciences & Health to stimulate public-private partnerships and Brain Research Center (grant no. LSHM19051). WF is recipient of JPND-funded E-DADS (ZonMW project # 733051106). WF is recipient of ABOARD, which is a public-private partnership receiving funding from ZonMW (#73305095007) and Health~Holland, Topsector Life Sciences & Health (PPP-allowance; #LSHM20106).
VV is supported by JPND-funded E-DADS project (ZonMW project #733051106).
LMR was funded by a ZonMW Memorabel fellowship (#10510022110012).
SL is recipient of ZonMW funding (#733050512). Genotyping of the Dutch case-control samples was performed in the context of EADB (European Alzheimer DNA biobank) funded by the JPco-fuND FP-829-029 (ZonMW projectnumber 733051061.
CET is supported by the European Commission (Marie Curie International Training Network, grant agreement No 860197 (MIRIADE), Innovative Medicines Initiatives 3TR (Horizon 2020, grant no 831434) EPND (IMI 2 Joint Undertaking (JU), grant No. 101034344) and JPND (bPRIDE), National MS Society (Progressive MS alliance), Alzheimer Association, Health

Holland, the Dutch Research Council (ZonMW), Alzheimer Drug Discovery Foundation, The Selfridges Group Foundation, Alzheimer Netherlands. CT is recipient of ABOARD, which is a public-private partnership receiving funding from ZonMW (#73305095007) and Health~Holland, Topsector Life Sciences & Health (PPP-allowance; #LSHM20106).

Research of the Alzheimer center Amsterdam is part of the neurodegeneration research program of Amsterdam Neuroscience. Alzheimer Center Amsterdam is supported by Stichting Alzheimer Nederland and Stichting VUmc fonds. For part of this work the Dutch national e-infrastructure was used with the support of the SURF Cooperative using grant no. EINF-2044.
Mass spectrometry-based proteomic analyses were performed by FBerven at the Proteomics Unit at the University of Bergen (PROBE). This facility is a member of the National Network of Advanced Proteomics Infrastructure (NAPI), which is funded by the Research Council of Norway (INFRASTRUKTUR-program project number: 295910).

Competing interests:
PJV and BMT co-inventors on patent of CSF proteomic subtypes (published under US2022196683A1, owner Stichting VUmc).
EVM is co-founder of ADx NeuroSciences, while JG is an employee of ADx NeuroSciences.
FB reports editorial fees from Springer, consulting fees from Biogen, IXICO Ltd and Combinostics, steering committee and DSMB compensation from Prothena, USC-ATRI, Merck, Biogen and grants from Roche, Merck, Biogen, IMI-EU, GE Healthcare, UK MS Society.
LV received consulting fees from Roch and Olink, all paid to Amsterdam UMC.
CET has a collaboration contract with ADx Neurosciences, Quanterix and Eli Lilly, performed contract research or received grants from AC-Immune, Axon Neurosciences, BioConnect, Bioorchestra, Brainstorm Therapeutics, Celgene, EIP Pharma, Eisai, Fujirebio, Grifols, Instant Nano Biosensors, Merck, Novo Nordisk, PeopleBio, Roche, Siemens, Toyama, Vivoryon.
She is editor of Alzheimer Research and Therapy, and serves on editorial boards of Medidact Neurologie/Springer, and Neurology: Neuroimmunology & Neuroinflammation.

Table

Table 1 Comparison of subtypes on clinical characteristics

Characteristic	Controls N=187	Subtype 1 N=137	Subtype 2 N=124	Subtype 3 N=24	Subtype 4 N=78	Subtype 5 N=56
Cognitive state						
Normal cognition	187 (100%)	51 (37%)	32 (26%)	1 (4%) b,c	15 (19%) b	8 (14%) b
MCI	0	37 (27%)	28 (23%)	3 (12%)	16 (21%)	19 (34%)
Dementia	0	49 (36%)	64 (52%) b	20 (83%) b,c	47 (60%)	29 (52%) d
Age	64.01 (11.83)	64.71 (6.82)	69.38 (8.35)a,b	64.46 (8.73)c	64.28 (8.09)c	66.16 (8.11)c
Men, n (%)	111 (59%)	62 (45%)a	61 (49%)	10 (42%)	51 (65%)b,c	41 (73%)b,c,d
Years of education, mean (SD)	12.4 (3.2)	12.1 (3.4)	11.2 (3.2)a	11.7 (2.9)	11.9 (3.4)	11.8 (3.3)
>1 *APOE* e4 allele, n (%)	51 (28%)	88 (68%)a	73 (62%)a	15 (65%)a	47 (64%)a	40 (74%)a
AD PRS, mean (SD)	5.6 (0.37)	5.8 (0.41)a	5.8 (0.31)a	6.0 (0.41)a,c	5.8 (0.46)a	5.8 (0.34)a
CSF total tau, pg/ml mean (SD)	199 (88)	592 (340)a	765 (447)a,b	882 (367)a,b	301 (166)a,b,c,d	469 (297)a,b,c
CSF p-tau181, Z score mean (SD)	0 (0.99)	3.4 (2.5)a	5.1 (3.1) a,b	5.0 (2.5) a,b	0.6 (1.4)a,b, c,d	2.1 (2.3) a,b,c,
CSF BACE1, pg/ml mean (SD)	1931.9 (643.49)	2203.8 (479.07)a	2478.49 (687.98) a,b	2185.33 (573.92) a,c	1391.36 (323.74)a, b,c,d	1819.55 (560.55)b,c,d,e
CSF Abeta40, pg/ml mean (SD)	7135.12 (2134.68)	7825.31 (1722.1)a	8519.82 (2264.92) a,b	6817.46 (1783.25)b,c	4610.78 (1286.96)a,b,c,d	5943.02 (1543.58)a,b,c
CSF NRGN, pg/ml mean (SD)	317.49 (148.07)	488.01 (178.95)a	634.44 (300.5)a,b	561 (175.88)a,c	244.09 (97.64)a,b,c,d	370.91 (166.83)b,c,d,e
CSF NEFL, pg/ml mean (SD)	360.1 (275.1)	447.01 (187.34)a	620.08 (341.41)a,b	630.12 (293.84)a,b	453.61 (292.84)a,c,d	594.16 (371.67)a,b,e
CSF VAMP2, pg/ml mean (SD)	162.17 (70.43)	196.32 (61.39)a	233.44 (79.91)a,b	188.6 (61.77)c	100.1 (39.8)a,b,c,d	141.94 (52.35)a,b,c,d,
Microbleed count on MRI, mean (SD)	0.91 (2.55)	1.89 (9.80)	1.16 (3.74)	1.65 (3.18)	2.07 (8.02)	4.40 (17.94)a,e

a is differs from controls with p<0.05, b is differs from subtype 1 with p<.05, c is differs from subtype 2 with p<0.05, d is differs from subtype 3 with p<.05, e is differs from subtype 4 with p<.05.

MCI is mild cognitive impairment, SD is standard deviation, *APOE* is Apolipoprotein E, AD is Alzheimer's disease, PRS is polygenic risk score, CSF is cerebrospinal fluid, MRI is magnetic resonance imaging.

CSF tau levels were measured with Innotest, and predicted from proteomics data when missing or measured with another assay.

CSF p-tau181 levels were normalised according to controls within cohort and assay.

Missing values (n) for: Years of education (n=6), APOE genotype (n=26), AD PRS (n=68), CSF ptau (n=26), CSF NEFL (n=3), MRI microbleeds (n=163).

See supplemental 4 for additional comparisons stratified for cognitive state, and adjusted for sex and age.

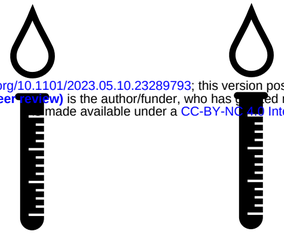
Figure legends

Figure 1. Schematic overview of subtype discovery within AD patients. 1) Cerebrospinal fluid (CSF) samples from 198 controls and 419 individuals were selected and 2) analysed with Alzheimer's disease (AD) were analysed with tandem mass tag mass spectrometry to obtain untargeted proteomics. 3) Protein levels were then compared between controls and AD to select proteins associated with AD. 4) Within the AD group for proteins related to AD, data driven clustering was performed with non-negative matrix factorisation (NMF) (5). 6) The resulting patient subgroups were then molecularly characterised based on their corresponding proteomic signatures, and compared on clinical and biological characteristics.

Figure 2. **A** Patient subtypes projected to UMAP space. **B** CSF protein levels averaged across individuals within subtypes. **C** Cell type specificity signatures for proteins associated with AD subtypes for proteins with increased levels on the top row, and decreased levels in the bottom row. Most left circle diagram shows all cell types associated with a subtype combined (most left circle diagrams). Proteins that could not be assigned to a specific cell type were not plotted (missing bar to 100% in first column). Circle diagrams to the right zoom into subcategories of specific celltypes (neurons, glia, immune cells and endothelial cells). Cell type specificity was determined according to the Human Protein Atlas. **D** Top transcription factors associated with subtypes from the CHEA and ENCODE databases. **E** GO biological pathways associated with subtypes (see supplemental material for all pathways). **F** AD genetic risk factors associated with specific subtypes, white indicates not significant. In all figures S1 is subtype 1 (hyperplasticity), S2 is subtype 2 (innate immune activation), S3 is subtype 3 (RNA dysregulation), S4 is subtype 4 (choroid plexus dysfunction), and S5 is subtype 5 (blood-brain barrier dysfunction). Supplementary tables 5, 6, 7, 8a and 8b list all proteins, pathways, transcription, and genetic factors tested with their statistical metrics.

Figure 3. **A** Hippocampal volume compared between subtypes. **B** Choroid plexus volume compared between subtypes. **C** Cortical atrophy associated with AD subtypes as compared to controls. **D** Clinical progression from MCI to dementia according to subtype (left; excluding subtype 3 due to n=2), and time from dementia to death according to subtypes (right). All atrophy measures are based on individuals with dementia only. See supplemental tables 9, 10a and 10b.

1) Cerebrospinal fluid samples

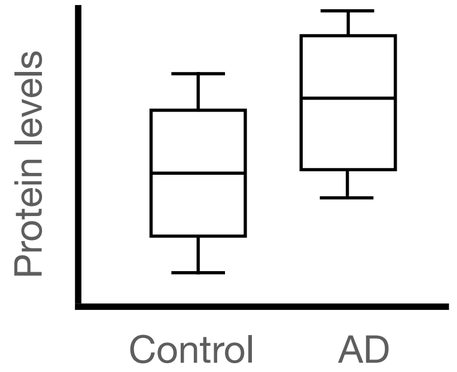


187 Controls 419 AD

2)

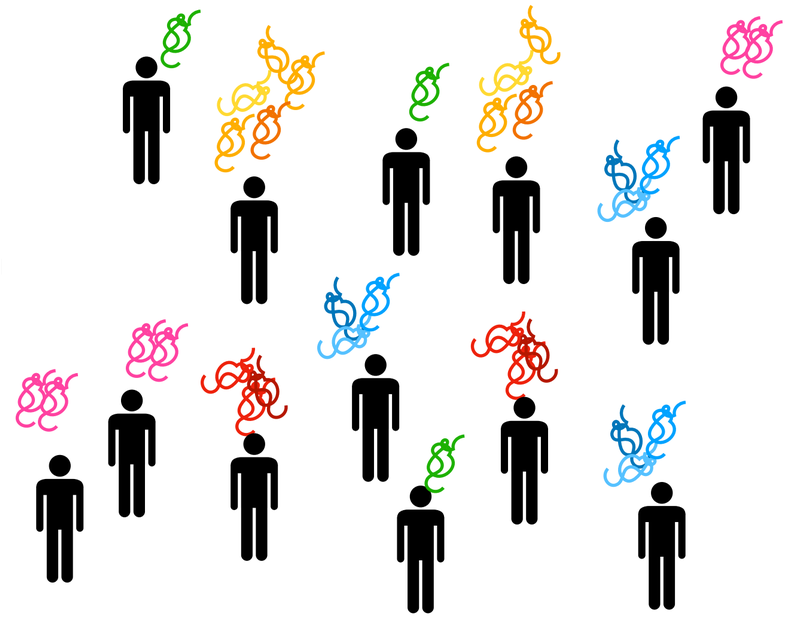
Tandem Mass Tag MS proteomics

3)



Select protein levels related to AD

4)



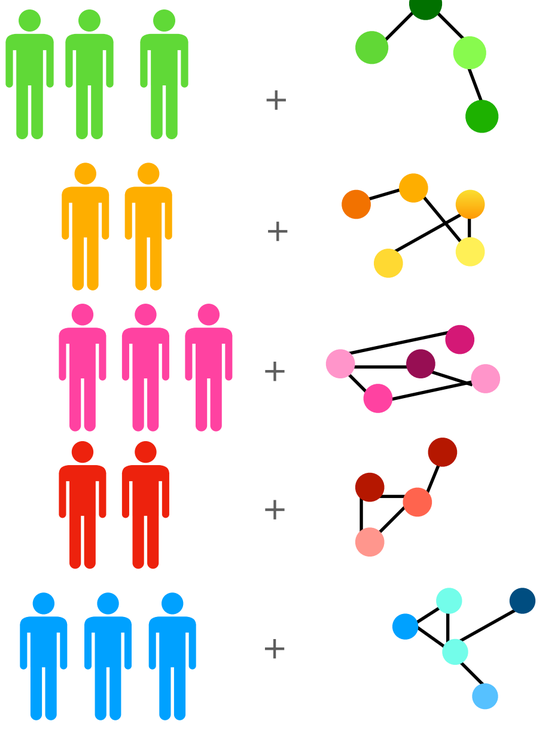
AD patient samples with AD related proteins

5)

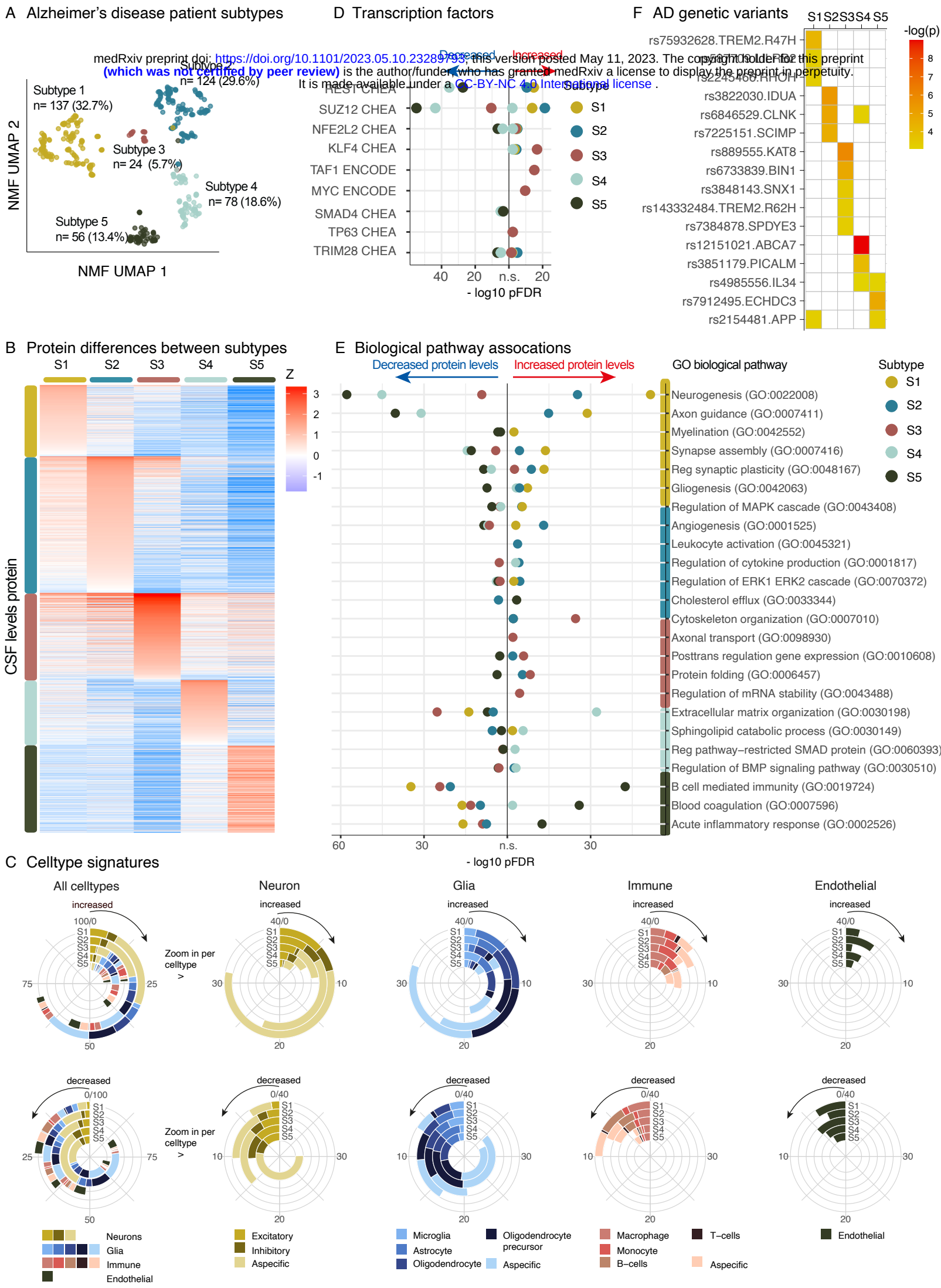
NMF dual clustering

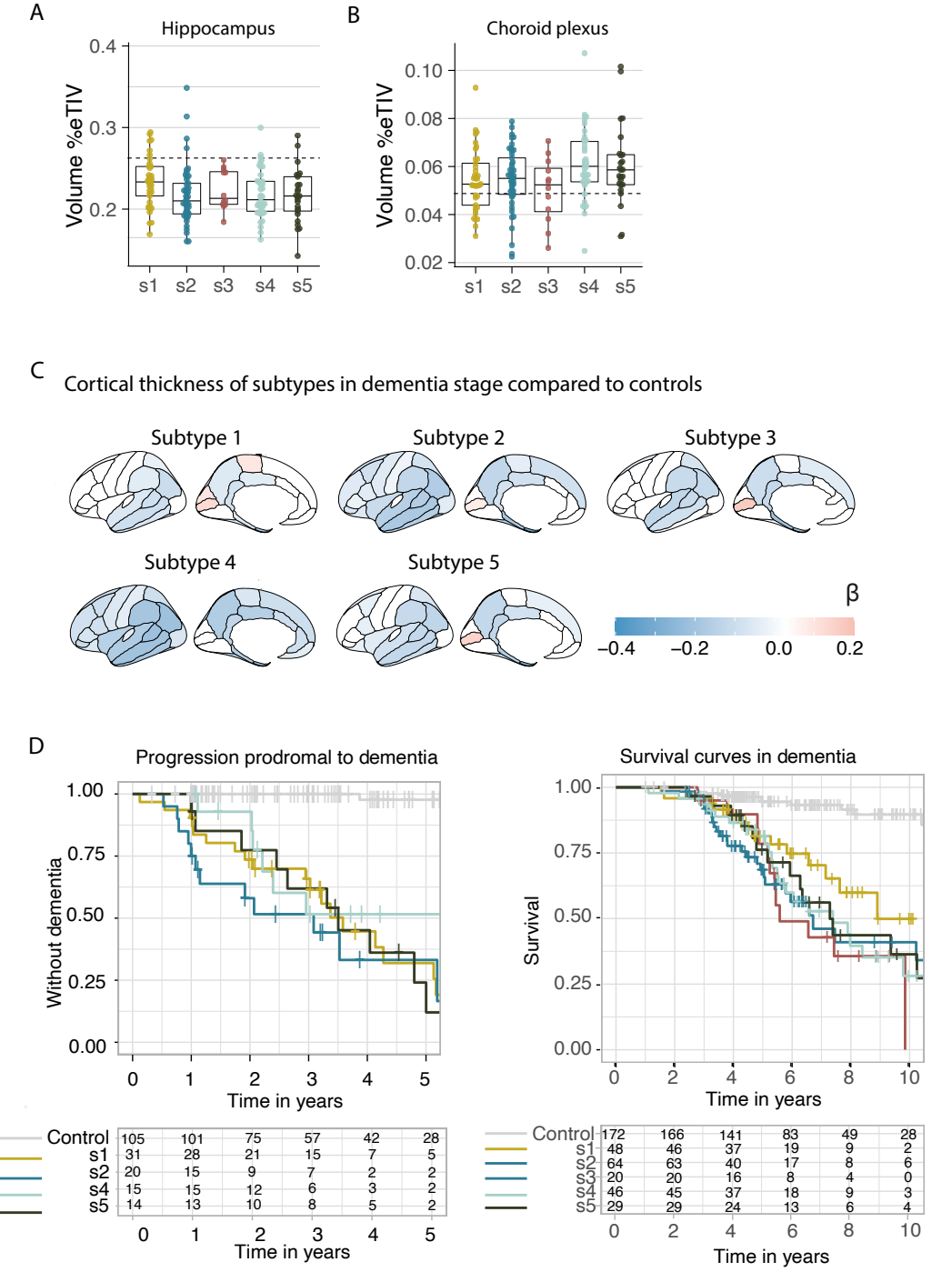
Data driven clustering *within* AD

6)



Patient subgroups + Molecular signatures




A
Hippocampus
Volume %eTIV
0.4
0.3
0.2
s1
s2
s3
s4
s5
B
Choroid plexus
Volume %eTIV
0.10
0.08
0.06
0.04
0.02
s1
s2
s3
s4
s5
C
Cortical thickness of subtypes in dementia stage compared to controls
Subtype 1
Subtype 2
Subtype 3
Subtype 4
Subtype 5
β
−0.4
−0.2
0.0
0.2
D
Progression prodromal to dementia
Without dementia
1.00
0.75
0.50
0.25
0.00
0
1
2
3
4
5
Time in years
Control 105 101 75 57 42 28
s1 31 28 21 15 7 5
s2 20 15 9 7 2 2
s4 15 15 12 6 3 2
s5 14 13 10 8 5 2
0
1
2
3
4
5
Time in years
Survival curves in dementia
Survival
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10
Time in years
Control 172 166 141 83 49 28
s1 48 46 37 19 9 2
s2 64 63 40 17 8 6
s3 20 20 16 8 4 0
s4 46 45 37 18 9 3
s5 29 29 24 13 6 4
0
2
4
6
8
10
Time in years