Proteomic analysis reveals distinct cerebrospinal fluid signatures across genetic frontotemporal dementia subtypes

Authors: Aitana Sogorb-Esteve[1,2]*, Sophia Weiner[3]*, Joel Simrén[3]*, Imogen J Swift[1,2], Martina Bocchetta[2,4], Emily G. Todd[2], David M. Cash[1,2], Arabella Bouzigues[2], Lucy L. Russell[2], Phoebe H. Foster[2], Eve Ferry-Bolder[2], John C. van Swieten[5], Lize C. Jiskoot[5], Harro Seelaar[5], Raquel Sanchez-Valle[6], Robert Laforce[7], Caroline Graff[8,9], Daniela Galimberti[10,11], Rik Vandenberghe[12,13,14], Alexandre de Mendonça[15], Pietro Tiraboschi[16], Isabel Santana[17,18], Alexander Gerhard[19,20], Johannes Levin[21,22,23], Sandro Sorbi[24,25], Markus Otto[26,27], Florence Pasquier[28,29,30], Simon Ducharme[31,32], Chris R. Butler[33,34], Isabelle Le Ber[35,36,37], Elizabeth Finger[38], Maria Carmela Tartaglia[39], Mario Masellis[40], James B. Rowe[41], Matthis Synofzik[42,43], Fermin Moreno[44,45], Barbara Borroni[46,47], GENFI consortium, Kaj Blennow[3,48,49,50], Henrik Zetterberg[1,3,48,51,52,53]*, Jonathan D. Rohrer[2]*, Johan Gobom[3,48]*.

*The authors have contributed equally to the work.

[+]Consortium author list shown in Appendix

Affiliations:

[1]UK Dementia Research Institute at University College London; WC1N 3BG, London, UK

[2]Dementia Research Centre, UCL Queen Square Institute of Neurology, University College London; WC1N 3BG, London, UK

[3]Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg; 431 39, Mölndal, Sweden

[4]Centre for Cognitive and Clinical Neuroscience, Division of Psychology, Department of Life Sciences, College of Health, Medicine and Life Sciences, Brunel University, UB8 3PH, London, UK

[5]Department of Neurology, Erasmus Medical Centre; 3015 GD, Rotterdam, Netherlands

[6]Alzheimer's disease and Other Cognitive Disorders Unit, Neurology Service, Hospital Clínic, Institut d'Investigacións Biomèdiques August Pi I Sunyer, University of Barcelona; 08036, Barcelona, Spain

[7]Clinique Interdisciplinaire de Mémoire, Département des Sciences Neurologiques, CHU de Québec, and Faculté de Médecine, Université Laval; G1V 0A6, Québec, QC, Canada

[8]Center for Alzheimer Research, Division of Neurogeriatrics, Department of Neurobiology, Care Sciences and Society, Bioclinicum, Karolinska Institutet; 171 64, Solna, Sweden

[9]Unit for Hereditary Dementias, Theme Aging, Karolinska University Hospital; 171 77, Solna, Sweden

[10]Fondazione Ca' Granda, IRCCS Ospedale Policlinico; 20122, Milan, Italy

[11]University of Milan, Centro Dino Ferrari; 20122, Milan, Italy

[12]Laboratory for Cognitive Neurology, Department of Neurosciences, KU Leuven; 3000, Leuven, Belgium
[13]Neurology Service, University Hospitals Leuven; 3000, Leuven, Belgium
[14]Leuven Brain Institute, KU Leuven; 3000, Leuven, Belgium
[15]Faculty of Medicine, University of Lisbon; 1649-028, Lisbon, Portugal
[16]Fondazione IRCCS Istituto Neurologico Carlo Besta; 20133, Milano, Italy
[17]University Hospital of Coimbra (HUC), Neurology Service, Faculty of Medicine, University of Coimbra; 3004-531, Coimbra, Portugal
[18]Center for Neuroscience and Cell Biology, Faculty of Medicine, University of Coimbra; 3004-531, Coimbra, Portugal
[19]Division of Neuroscience and Experimental Psychology, Wolfson Molecular Imaging Centre, University of Manchester; M20 3LJ, Manchester, UK
[20]Departments of Geriatric Medicine and Nuclear Medicine, University of Duisburg-Essen; 4514, Germany
[21]Department of Neurology, Ludwig-Maximilians Universität München; 80539, Munich, Germany
[22]German Center for Neurodegenerative Diseases (DZNE); 81377, Munich, Germany
[23]Munich Cluster of Systems Neurology (SyNergy); 81377, Munich, Germany
[24]Department of Neurofarba, University of Florence; 50139, Florence, Italy
[25]IRCCS Fondazione Don Carlo Gnocchi; 50143, Florence, Italy
[26]Department of Neurology, University of Ulm; 89081, Ulm, Germany
[27]Martin-Luther-University Hospital of Halle-Wittenberg, Department of Neurology, 06120, Halle (Saale), Germany
[28]University of Lille; 59000 France
[29]Inserm 1172, Lille; 59000 France
[30]CHU, CNR-MAJ, Labex Distalz, LiCEND Lille; France
[31]Department of Psychiatry, McGill University Health Centre, McGill University; H4A 3J1, Montreal, Québec, Canada
[32]McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University; H3A 0G4, Montreal, Québec, Canada
[33]Nuffield Department of Clinical Neurosciences, Medical Sciences Division, University of Oxford; OX3 9DU, Oxford, UK
[34]Department of Brain Sciences, Imperial College London; W12 0NN, London, UK
[35]Sorbonne Université, Paris Brain Institute – Institut du Cerveau – ICM, Inserm U1127, CNRS UMR 7225, AP-HP - Hôpital Pitié-Salpêtrière; 75013, Paris, France
[36]Centre de référence des démences rares ou précoces, IM2A, Département de Neurologie, AP-HP - Hôpital Pitié-Salpêtrière; 75013, Paris, France
[37]Département de Neurologie, AP-HP - Hôpital Pitié-Salpêtrière ; Paris, France
[38]Department of Clinical Neurological Sciences, University of Western Ontario; N6A 5A5, London, ON, Canada

[39]Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto; M5S 1A8, Toronto, ON, Canada
[40]Sunnybrook Health Sciences Centre, Sunnybrook Research Institute, University of Toronto; M4N 3M5, Toronto, Canada
[41]Department of Clinical Neurosciences, University of Cambridge; CB2 3EB, Cambridge, UK
[42]Division Translational Genomics of Neurodegenerative Diseases, Hertie-Institute for Clinical Brain Research and Center of Neurology, University of Tübingen; 72076, Tübingen, Germany
[43]Center for Neurodegenerative Diseases (DZNE); 72076, Tübingen, Germany
[44]Cognitive Disorders Unit, Department of Neurology, Donostia Universitary Hospital; 20014, San Sebastian, Spain
[45]Neuroscience Area, Biodonostia Health Research Institute; 20014, San Sebastian, Gipuzkoa, Spain
[46]Department of Clinical and Experimental Sciences, University of Brescia; 25123, Italy
[47]Department of Continuity of Care and Frialy, ASST Spedali Civili Brescia; 25123, Italy
[48]Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital; SE-43180, Mölndal, Sweden
[49]Institut du Cerveau et de la Moelle épinière (ICM), Pitié-Salpêtrière Hospital, Sorbonne Université, 75013, Paris, France
[50]University of Science and Technology of China and First Affiliated Hospital of USTC, Hefei, Anhui, P.R. China
[51]Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square; WC1N 3BG, London, UK
[52]Hong Kong Center for Neurodegenerative Diseases, Clear Water Bay; Hong Kong, China
[53]Wisconsin Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, University of Wisconsin-Madison; 53792, Madison, WI, USA

Corresponding author: Aitana Sogorb-Esteve a.sogorb-esteve@ucl.ac.uk

Abstract

We used an untargeted mass spectrometric approach, tandem mass tag (TMT) proteomics, for the identification of proteomic signatures in genetic frontotemporal dementia (FTD). A total of 238 cerebrospinal fluid (CSF) samples from the Genetic FTD Initiative (GENFI) were analysed, including 107 presymptomatic (44 *C9orf72*, 38 *GRN*, 25 *MAPT*) and 55 symptomatic (27 *C9orf72*, 17 *GRN*, 11 *MAPT*) mutation carriers as well as 76 mutation-negative controls ('non-carriers'). We found shared and distinct proteomic alterations in each genetic form of FTD. Among the proteins significantly altered in symptomatic mutation carriers compared to non-carriers, we found a set of proteins including neuronal pentraxin 2 (NPTX2) and fatty acid binding protein 3 (FABP3) shared across all three genetic forms, as well as in patients with Alzheimer's disease from previously published datasets. We observed differential changes in lysosomal proteins among symptomatic mutation carriers with marked abundance decreases in *MAPT* carriers, but not other carriers. Further, we identified mutation-associated proteomic changes already evident in presymptomatic mutation carriers. Weighted gene co-expression network analysis combined with gene ontology annotation revealed clusters of proteins enriched in neurodegeneration and glial responses, as well as synapse-, or lysosome-related proteins indicating that these are the central biological processes affected in genetic FTD. These clusters correlated with measures of disease severity and associated with cognitive decline. This study revealed distinct proteomic changes in the CSF of patients with genetic FTD, providing insights into the pathological processes involved in the disease. Additionally, we identified proteins that warrant further exploration as diagnostic and prognostic biomarker candidates.

One sentence summary: Both distinct and common cerebrospinal fluid proteomic signatures were observed in the different genetic forms of frontotemporal dementia.

Main Text:

INTRODUCTION

Frontotemporal dementia (FTD) is an umbrella term referring to a group of progressive neurodegenerative disorders, which typically present with behavioral changes (behavioral variant, (bvFTD), language problems (primary progressive aphasia, PPA), or motor impairment (either FTD with amyotrophic lateral sclerosis or FTD with parkinsonism)) *(1)*. Although less common than Alzheimer's disease (AD), dementia with Lewy bodies and vascular dementia, FTD is a leading cause of early onset dementia *(2)*. The underlying molecular basis of FTD is complex, but most cases can be attributed to a frontotemporal lobar degeneration (FTLD) pathology, with cellular inclusions of tau, TAR DNA-binding protein 43 (TDP-43) or FET proteins (FUS (Fused in sarcoma), EWS (Ewing sarcoma) and TAF15 (TATA-binding associated factor 15)) *(3)*. Unlike AD, around a third of FTD cases have a genetic cause, with the most common mutations occurring in three genes: *GRN* (progranulin) and *C9orf72* (chromosome 9 open reading frame 72), both of which are typically accompanied by an underlying TDP-43 proteinopathy, as well as *MAPT* (microtubule-associated protein tau), manifesting as tauopathy *(1, 4)*.

In FTD, the complex relationship between clinical presentations and underlying molecular pathology poses a challenge for its diagnosis and treatment. AD can be viewed as a successful example of how the introduction of cerebrospinal fluid (CSF) biomarker-assisted diagnosis has led to recent therapeutic advances *(5)* with the potential to revolutionise its treatment. In the case of FTD, however, the historic lack of biomarkers, as well as the complex relationship between clinical symptomatology and underlying pathophysiology have so far hampered such advancements. Nonetheless, there are biomarkers that show promise also in the context of FTD. Neurofilament light chain (NfL) has emerged as a promising, although disease-nonspecific, biomarker in differentiating FTD from primary psychiatric causes of behavioral symptoms *(6)* and, due to plasma NfL increasing in concentration in the presymptomatic phase of genetic FTD, also as a biomarker to detect neurodegeneration onset and disease intensity *(7)*. Although there are indications that group-level concentrations of NfL are highest (at least in plasma) in *GRN* carriers *(8)*, NfL cannot be used to identify the underlying pathology. For

this purpose, mutation- or pathology-specific biomarkers are needed, with current examples being limited, such as low plasma/CSF progranulin as indication of an underlying *GRN* mutation resulting in haploinsufficiency *(9, 10)*; or promising new results on TDP-43 or 3R/4R tau protein in plasma-derived extracellular vesicles *(11)*, that need further replication. Due to the lack of an *antemortem* gold standard for FTLD-tau and TDP-43 pathologies, sporadic FTD is likely not an ideal model to develop novel biomarkers at present. In familial FTD, however, the clear relationship between genetic mutation and resulting pathology may provide a context that allows the identification of such markers.

Previous studies using antibody-based methods *(12–14)* or mass spectrometric techniques *(15, 16)* have identified several FTD biomarker candidates, including neurofilament light, medium (NfM), and heavy (NfH), neuronal pentraxins, chitinase-3-like protein 1 (CHI3L1, also known as YKL-40), and ubiquitin carboxy-terminal hydrolase L1 (UCHL1). However, none of these proteins have proven specific for either FTLD or its subtypes, with similar alterations being seen in other neurodegenerative disorders, such as AD, Creutzfeldt-Jakob disease or ALS *(14, 17–21)*.

In this study, we adopted an untargeted proteomics approach, using high-resolution mass spectrometry combined with tandem mass tag (TMT), to measure CSF proteins in a large, well-characterised genetic FTD cohort: the Genetic FTD Initiative (GENFI) study. We aimed to measure changes in low-abundant proteins not previously implicated in FTD to identify proteomic signatures of symptomatic groups carrying the most common genetic mutations causing FTD and therefore potentially distinguish specific underlying pathologies. Furthermore, we explored CSF proteomic changes that may identify mutation carriers at the presymptomatic stage of the disease, as has been done previously in autosomal dominant AD *(22)*. Lastly, we investigated alterations of biological pathways in FTD, as mirrored in the CSF proteome, and their association with relevant clinical parameters and cognitive decline.

RESULTS

We analyzed a total of 238 CSF samples from 71 *C9orf72* expansion carriers, 55 *GRN* mutation carriers, and 36 *MAPT* mutation carriers, including both presymptomatic and symptomatic carriers in each group, as well as 76 asymptomatic non-carriers (Table 1).

Key methodological information of this study is summarized in Fig. 1, and specific descriptions for each analysis are detailed in the Materials and Methods and Supplementary Methods sections. Having prepared and analyzed all study samples using protocols previously described and developed by our laboratory *(23–25)*, we initially explored differential protein abundances among symptomatic groups to assess wide-spread CSF proteomic changes in the context of different underlying pathologies and compared those to AD. Next, we utilised linear models to discern mutation-associated proteins already changed at the presymptomatic disease stage. Furthermore, employing weighted gene co-expression network analysis (WGCNA), we aimed to elucidate pathophysiological features associated with genetic mutations, as well as the cross-sectional correlations of protein networks with measures of cognitive function and brain volume. Finally, to investigate the prognostic properties of protein networks, we assessed their association with cognitive decline.

After outlier exclusion and removal of proteins with high missingness, we identified and obtained quantitative information for 1981 CSF proteins. First, we compared our TMT dataset to existing biomarker data from the same sample cohort. TMT CSF neurofilament light chain (NEFL; henceforth used interchangeably with protein abbreviation, NfL) measurements strongly correlated with plasma NfL measurements acquired on the Single molecule array (Simoa, Quanterix) platform (R=0.62, P<0.001; fig. S1A). The relative protein abundances of 14-3-3 epsilon (referred to as YWHAE) (R=0.39, P<0.001; fig. S1B), neuronal pentraxin 2 (NPTX2) (R=0.8, P<0.001; fig. S1C), and neuronal pentraxin receptor (NPTXR) (R=0.68, P<0.001; fig. S1D) also correlated significantly with previous data from the same cohort, acquired using targeted mass spectrometric analysis *(26)*. The strong correlations of TMT relative protein abundances with measures acquired on two independent platforms indicate good analytical precision of our results.

Cerebrospinal fluid proteomes differ across symptomatic FTD mutation carrier groups

Linear regression analysis, including age and sex as covariates, was used to perform group comparisons between non-carriers and symptomatic mutation carriers. In the case of symptomatic *MAPT* mutation carriers, 58 proteins significantly differed in abundance compared with non-carriers (Fig. 2A; Table S1A), whereas the abundance of 138 and 385 proteins was significantly altered in symptomatic *GRN* (Fig. 2B; Table S1B) and *C9orf72* mutation carriers (Fig. 2C; Table S1C) compared to non-carriers, respectively (P_{adjust}<0.05).

Next, to strengthen our findings, we compared our results to those from an external cohort consisting of symptomatic *GRN* carriers (n=11) and healthy non-carriers (n=12) *(27)* whose CSF proteome was measured with label-free mass spectrometry. Most proteins were commonly quantified in both studies, of which 73 proteins were significantly changed in both datasets ($P_{unadjusted}$<0.05) (fig. S2, Table S2), with $\log_2$ fold changes being strongly correlated between the studies (R=0.87, P< 0.001) (fig. S2).

From the 25 hits that presented the largest $\log_2$ fold changes in each symptomatic mutation group (Table S3), a list of proteins was compiled (excluding overlap between groups) denoting corresponding protein abundance fold changes compared with non-carriers in a heatmap (Fig. 2D; Table S3). As expected, the three neurofilaments, NfL, NfM and NfH, alongside YKL-40 (CHI3L1), exhibited the greatest fold change in abundance across most symptomatic groups when compared to non-carriers, with NfL abundances being up to 7.4-times higher in symptomatic *GRN* carriers in comparison to non-carriers. Other proteins showing a notable positive fold change in symptomatic mutation carriers included the spectrins (SPTBN1, SPTAN1) as well as UCHL1 (ubiquitin C-terminal hydrolase 1), FABP3 (fatty acid binding protein 3), PEA15 (Proliferation and apoptosis adaptor protein 15) and several 14-3-3 proteins (YWHAZ, YWHAG, YWHAE). Proteins that were lower in abundance across symptomatic mutation carriers compared with non-carriers included the synaptic proteins NPTXR, NPTX2 and NPTX1, as well as PCSK2 (proprotein convertase subtilisin/kexin type 2) and PENK (proenkephalin). Furthermore, GRN relative abundance levels was lower in *GRN*

mutation carriers. Most proteins showed the same directionality of abundance fold change across the three mutation carrier groups except for a few proteins. These include GRN, which showed opposite direction of change in symptomatic *C9orf72* and *GRN* carriers (both P_{adjust}<0.05), and the lysosomal proteins deoxyribonuclease 2 (DNASE2) and phospholipase B domain containing 2 (PLBD2), which were selectively decreased in symptomatic *MAPT* carriers.

Proteomic similarities and differences between genetic FTD and sporadic AD

Since some of the proteins quantified in this study are expected to change also in other neurodegenerative disorders, we compared the summary statistics of our differential abundance analyses of symptomatic FTD mutation carrier groups with summary statistics of previously published TMT proteomics datasets from two distinct AD studies: the European Medical Information Framework (EMIF) CSF study *(25)* and a CSF proteomics study performed by Higginbotham and colleagues *(28)*. Of the about 1192 proteins quantifiable in all three studies, (Fig. 3A; Table S4A, Table S4B and Table S5), only 6 were significantly changed in all groups (P_{adjust}<0.05) (YWHAZ, YWHAG, UCHL1, NPTXR, NPTX2 and FABP3, Fig. 3, B and C; Table S5). Conversely, many proteins were distinctly changed in each FTD mutation carrier group (Fig. 3, B and C, Table S5), with more widespread changes being found in symptomatic *C9orf72* carriers (calretinin [CALB2], sortilin 1 [SORT1] and roundabout guidance receptor 1 [ROBO1]) compared with *GRN* (transmembrane protein 132A [TMEM132A], ring finger protein [RNF13] and chitinase 3 like 2 [CHI3L2]) and *MAPT* (hexosaminidase subunit alpha [HEXA], semaphorin 6A [SEMA6A] and cathepsin D [CTSD]) carriers. Proteins shared between *C9orf72* and *GRN* carriers included many proteins involved in lysosomal processes (GRN, cathepsin S [CTSS], lysosomal-associated membrane protein 1 [LAMP1]). Proteins uniquely changed in both AD studies included neurogranin (NRGN) and SPARC related modular calcium binding 1 (SMOC1), both previously shown to increase in response to amyloid pathology *(29)*. Only two proteins were distinctly changed in all symptomatic FTD mutation carrier groups (CD44 and follistatin like 4 [FSTL4]), likely reflecting the different processes involved in these disease-causing mutations.

Mutation-associated proteomic changes are evident in presymptomatic disease mutation carriers

Having compared proteomic alterations of symptomatic FTD subtypes and their overlap with AD, we next set out to determine changes in protein abundances associated with a specific genetic background, regardless of affectation (presence or absence of symptoms). The presence of symptoms is expected to coincide with diverse neurodegenerative processes impacting the CSF proteome and obscuring potential mutation-related changes. Thus, to investigate proteomic alterations attributable to each underlying pathogenic mutation, we (i) fitted linear models combining all study participants, testing the effect of genetic mutation on protein abundances while adjusting for affectation (Fig. 4, Table S6) and (ii) compared CSF proteomes of presymptomatic individuals with non-carriers for each genetic group separately (fig. S3-S8, Table S7A-S7C). This approach yielded several proteins strongly associated with either *C9orf72* (Fig. 4A), *GRN* (Fig. 4B) or *MAPT* (Fig. 4C) mutation status, of which the top five proteins for each association were chosen for visual display. Standardised β coefficients indicate the strength of the association respectively and are depicted in a forest plot for ease of comparison. The protein most strongly associated with *C9orf72* mutation status was CALB2 (Fig. 4A, standardised β=0.77, P_{adjust}<0.01), which could also be found among the top changed proteins in the analysis of presymptomatic *C9orf72* carriers vs. non-carriers (Fig. S5 and S8). Numerous proteins found to be associated with *C9orf72*, such as glucose-6-phosphate isomerase (GPI), hexokinase 1 (HK1) (Fig. 4A, fig. S8), and phosphoglycerate kinase 1 (PGK1) (fig. S8) are key enzymes of the glycolysis pathway, hinting at early metabolic disturbances. The proteins CALB2, HK1, and PGK1 demonstrated a stepwise increase in abundance from non-carriers over presymptomatic to symptomatic *C9orf72* carriers (fig. S8), further underlining their implication in *C9orf72*-related disease processes.

Reflecting the GRN haploinsufficiency, the protein most strongly associated with *GRN* mutation status was GRN itself, (Fig. 4B, fig. S4 and S7, standardised β=-1.59, P_{adjust}<0.01), followed by NAGA (alpha-N-acetylgalactosaminidase) (standardised β=0.71, P_{adjust}=0.04) and RNF13 (standardised β=0.64, P_{adjust}=0.09). RNF13, though narrowly failing to reach the significance threshold of 0.05 after multiple testing correction in the combined analysis, was found to be

significantly changed in the presymptomatic *GRN* carrier vs. non-carrier analysis (fig. S4, P_{adjust}=0.03) and increased in abundance across the *GRN* disease continuum (fig. S7). The proteins most strongly associated with *MAPT* mutation status were PEA15 (Fig. 4C, standardised β=0.9, P_{adjust}<0.01) and SEMA6A (standardised β=-0.82, P_{adjust}<0.05). PEA15 was also significantly altered in the comparison between presymptomatic *MAPT* carriers vs. non-carriers (fig. S3, P_{adjust}=0.02) and increased in abundance from the presymptomatic to symptomatic disease stage (fig. S6). Due to concerns of family membership adversely affecting our results, we conducted sensitivity analyses adopting the same linear models as in the main analysis but including one member from each family. These analyses presented similar results (Table S8, A to E).

Protein networks reveal pathology-specific pathophysiological alterations and correlate with clinical parameters

Having studied the proteomic signatures of each genetic group, we further explored the biological processes implicated in these proteomic changes by performing WGCNA (fig. S9 toS23). WGCNA is an analysis tool aimed at reducing the complexity of a proteomics dataset by breaking it down into gene ontology (GO)-annotated protein clusters. These protein modules consist of highly co-correlated proteins likely reflecting similar biological processes. We identified a total of 14 protein modules, including a group of 645 proteins that could not be assigned to any of the modules and a module containing contaminants from the laboratory environment. The modules varied in size from 14 to 349 proteins with a median module size of 52 proteins (Table S9). We determined the biological relevance of each protein module utilising GO analysis of its constituent proteins and selected the most representative term for module annotation (fig. S12 to S23). Furthermore, we identified the hub proteins of each module, indicating the proteins most strongly correlating (R>0.7) with the module's first principal component (Eigenprotein value), as most representative and important proteins of the respective module.

Figure 5A shows a selection of six protein modules and their corresponding Eigenprotein values (representative abundance values) plotted across all genetic groups as well as non-carriers. One module, which we termed 'Core markers' of neurodegenerative disease, consisted of 15 proteins and

was most strongly increased in abundance in each genetic group at the symptomatic stages when compared to non-carriers. The strong difference between non-carriers and presymptomatic *MAPT* carriers is largely influenced by age. It included YWHAG, NEFL, CHI3L1, NEFM and YWHAZ as hub proteins. These proteins were also among the top hits in the differential abundance analysis and had the highest fold change in symptomatic mutation carriers compared with non-carriers (Fig. 2D). As expected, many proteins belonging to the 'Core markers' module were also seen among the proteins overlapping between the three genetic forms and were found to be altered in the CSF of AD patients in the EMIF and Higginbotham studies (Fig. 3B and 3C).

Correlating the 'Core marker' Eigenprotein values with clinical parameters in both presymptomatic and symptomatic mutation carriers (Fig. 5B) revealed a strong positive association of the module with both plasma NfL (R=0.86, P_{adjust}<0.0001) and the National Alzheimer's Coordinating Center's Frontotemporal Lobar Degeneration plus clinical dementia rating sum of boxes (FTLD-CDR-SOB) disease severity scores (R=0.67, P_{adjust}<0.0001) as well as a negative association with MMSE scores (R=-0.53, P_{adjust}<0.0001) and regional brain volumes. The 'Core markers' module also positively correlated with estimated years until disease onset (EYO) in presymptomatic individuals (R=0.68, P_{adjust}<0.0001).

Besides the 'Core markers' module, Eigenprotein values for both the 'Actin binding' module and the 'Stress response' module were higher across symptomatic mutation carrier groups (albeit not statistically significant), suggesting common pathophysiological alterations in these processes (Fig. 5A). Both modules, along with the 'Glycosaminoglycan processing' module (Fig. S24D), showed a similar correlation pattern to the 'Core markers' module.

Conversely, the 'Synapse' module, containing proteins such as CHGB, SHISA6, CADM3, CADM1 and GPR158 showed lower Eigenprotein values in all symptomatic mutation carrier groups compared with non-carriers, although changes were not significant. Its correlation pattern with clinical parameters was inverse compared to the 'Core markers' module, exhibiting negative correlations with

age, plasma NfL and FTLD-CDR-SOB scores, and positive correlations with MMSE scores and brain volumes (Fig. 5B), similarly to the 'Neuronal development' and the 'Extracellular matrix 1' modules (fig. S24B and E). The 'Neuronal development' module contained several proteins considered to be markers of synaptic loss (NPTX2 and NPTXR, among others) and was significantly lower in symptomatic *C9orf72* carriers (P_{adjust}<0.05).

We also identified a module associated with lysosomal proteins ('Lysosome' module), for which Eigenprotein values were selectively decreased in symptomatic *MAPT* mutation carriers compared to non-carriers (P_{adjust}<0.05). They were also slightly decreased in presymptomatic *MAPT* individuals, albeit without statistical significance (P=0.79). The hub proteins were determined to be SIAE (sialic acid acetylesterase), hexosaminidase subunit beta (HEXB), HEXA, DNASE2 and PLBD2, all of which are implicated in lysosomal processes. These specific changes in *MAPT* mutation carriers in DNASE2 and PLBD2 were already evident in the heatmap (Fig. 2D) contrasting symptomatic mutation carrier groups. Other lysosomal proteins found to be commonly changed in *GRN* and *C9orf72* carriers (LAMP1 and CTSS, Fig. 3) were not part of the 'Lysosome' module, suggesting different subpopulations of lysosomal proteins, which might be reflective of distinct biological processes. The 'Lysosome' module did not correlate with markers of neurodegeneration, cognitive decline or brain atrophy.

The 'Immune response' module contained proteins related to the complement pathway and the immune system. For all symptomatic groups, there was a visible trend of increase in these clusters when compared to non-carriers, however, these differences were not statistically significant (P>0.05). This module showed similar correlation patterns with clinical features to the 'Core markers' module.

Protein networks associate with cognitive decline in mutation carriers

To evaluate the prognostic properties of protein networks, the module Eigenprotein values of mutation carriers with cognitive evaluation at the time of lumbar puncture (LP) (n=146, mean number of annual visits = 2.7, range 1-5) were modelled with FTLD-CDR-SOB score as outcome. In

agreement with analyses of cross-sectional cognitive scores, higher 'Core markers' Eigenprotein values were most strongly associated with higher FTLD-CDR-SOB scores, reflecting poorer cognitive outcomes (standardised β=0.83, P<0.001; Fig. 6A). A similar but less prominent pattern was seen for the 'Actin binding' module (standardised β=0.50, P<0.001; Fig. 6B). Conversely, lower Eigenvalues of the 'Synapse' module were associated with increasing FTLD-CDR-SOB scores (standardised β=-0.49, P<0.001; Fig. 6C). This indicated that lower 'Synapse' Eigenprotein values were associated with worse cognitive outcomes. Further, 'Semaphorin signalling', 'Neuronal development', 'Extracellular Matrix 1', 'Lysosome' and 'Immune response' module Eigenprotein values were also significantly (all P<0.05) associated with cognitive decline (Fig. 6D; for full model output, see Table S10).

DISCUSSION

The present study offers a detailed and untargeted account of the CSF proteomic signatures in genetic FTD. By including participants from the well-characterised GENFI cohort, with presymptomatic and symptomatic carriers of pathogenic mutations in the three genes comprising the overwhelming majority of genetic FTD, we covered most of the clinical continuum as well as its underlying genetic causes. Our analytical approach allowed us to uncover proteomic changes beyond known CSF and blood biomarkers, such as NfL, GFAP, and progranulin, suggesting potential pathology- and FTD-specific biomarkers.

To assess both differences and similarities across the FTD spectrum, we explored the proteome of each genetic group through separate analyses. Employing differential protein abundance analysis, we found several proteins that were altered in all symptomatic mutation carriers. Among these proteins, many of the top hits were neuronal proteins known to be increased in CSF in several neurodegenerative diseases, including neurofilaments (NfL, NfM and NfH), and 14-3-3 proteins (YWHAZ, YWHAG) *(17, 18, 30)*. NfL (both when measured in CSF and plasma) has especially been suggested to be of diagnostic, prognostic and theragnostic value in FTD, as both this and other studies find large fold changes (seemingly most pronounced in *GRN* carriers) compared with healthy controls and even other brain-related conditions, which bears important implications for differential diagnoses *(7, 30, 31)*. The decreased relative abundances of neuronal pentraxins (NPTX1, NPTX2) and their receptor (NPTXR), previously reported to be decreased in genetic FTD *(15, 26, 32)* and other neurodegenerative diseases *(18, 33)*, further emphasises the presence of synaptic changes in FTD. These markers displayed a similar fold change in the study of FTD-*GRN* by Pesämaa et al. *(27)*, which we used to validate our findings. In addition, changes shared between groups included proteins recently suggested to be associated with astrocytic and microglial responses in AD as well as FTD-TDP brains, such as rab GDP dissociation inhibitor alpha (GDI1), FABP3, and CD44 *(34)*. Although not significantly changed in either the EMIF or Higginbotham study, CD44 antigen has been shown to play a role in neuroinflammation in AD, in relation to disease associated microglia (DAM) *(34, 35)* and their communication with astrocytes *(36)*, as well as in *GRN* deficient animal models *(37)*.

Despite not being specific to glial responses in FTD, the clear increases seen in symptomatic FTD suggest that CD44 may be a promising fluid-based marker to index such glial changes in future trials. Of note, GDI1, FABP3, and CD44 were also identified as microglial activation-dependent markers in the study by Pesämaa et al. *(27)*.

Many of the proteins found to be clearly altered in all groups of symptomatic carriers were assigned to the 'Core markers' module in the protein network analysis (YWHAG, NfL, UCHL-1, FABP3, CHI3L1, CD44). Several of these 'Core marker' proteins (FABP3, UCHL1 and YWHAG, among others) were also shown to be changed in abundance in the CSF of AD patients, as evident by the EMIF and Higginbotham studies *(25, 28)*. Together, these findings support the strong neurodegenerative and glial component of both diseases and highlight that, despite AD and FTD being separate disease entities, they appear to share common downstream pathophysiological features. Of note, the 'Core markers' also reflected disease severity and imaging measures of neurodegeneration and proved to be the protein network most closely linked to cognitive decline and estimated years until disease onset highlighting the prognostic value of markers reflecting neurodegenerative and neuroinflammatory processes.

Besides the 'Core markers' module, we identified several other protein modules seemingly altered across several groups in the FTD spectrum compared with non-carriers, with constituent proteins relating to the synapse ('Synapse' module), in line with results shown in the heatmap (Fig. 2), actin binding and stress response. Lower relative abundances of the 'Synapse' Eigenprotein values also predicted cognitive decline in mutation carriers. Although these protein networks strongly correlated with clinical and neuroimaging features, abundance differences compared with non-carriers were rendered non-significant, likely due to their association with age. Of note, the synaptic protein neurogranin (NRGN) as well as the extracellular matrix protein SMOC1 were not altered in any of the groups of symptomatic genetic FTD mutation carriers but altered in both AD studies. This is in line with previous research *(29)* and suggests their specificity for amyloid-related changes in AD. Conversely, the protein FSTL4 was found to be changed in all groups of symptomatic FTD mutation

carriers but not AD, hinting at its potential specificity for FTD. Knowledge is still limited on the extent to which this protein is associated with neurodegenerative disorders, although one small study reported lower protein abundances of FSTL4 in patients with ALS *(38)*.

As this study aimed to look beyond proteomic alterations shared between FTD subtypes as well as AD, we also investigated the differences across genetic forms of FTD and their expected underlying pathologies. We identified lysosomal proteins with the potential to separate processes implicated in *MAPT* mutation carriers but not the other two groups. Decreased DNSAE2 and PLBD2 (which show divergent patterns in symptomatic *GRN* and *C9orf72* carriers as shown in Fig. 2B) appear to be specifically related to the presence of tau pathology, without the amyloid background observed in AD (as evident in Fig. 3). This was further supported by PLBD2 and DNASE2 being among the hub proteins in the 'lysosome' module, driving the marked Eigenprotein value decrease in *MAPT* mutation carriers (Fig. 5G). Indeed, evidence suggests that PLBD2 and DNASE2 play a role in lysosomal processes *(39–42)*. These results were unexpected, given the evidence of lysosomal dysfunction in *GRN* mutation carriers, but not in *MAPT* mutation carriers, due to the role of progranulin in the endolysosomal pathway *(43, 44)*. Nonetheless, tau protein has been previously implicated in the trafficking of autophagic vesicles and autolysosome fusion *(45–47)*, suggesting that a reduction of proteins related to the endolysosomal pathway in *MAPT* may indicate a potential dysregulation in this system. Lysosomal acid phosphatase 2 (ACP2) *(48)*, a member of the 'lysosome' module, was found to be downregulated in presymptomatic *MAPT* carriers, which aligns with the changes seen in the protein networks found in symptomatic carriers. This dysregulation might be different from that observed in *C9orf72* and *GRN* carriers, in which there was a selective increase for some lysosomal proteins (LAMP1 and CTSS) not belonging to the 'Lysosome' module and thus displaying a different correlation pattern.

In analyses stratifying groups by mutation irrespective of symptomatology, we observed a stepwise abundance increase across the disease continuum in PEA15, an astroglial protein associated with glial responses *(34)*, being more strongly associated with *MAPT* mutation carriership than with *GRN* and

C9orf72. In *GRN* carriers, the expected decrease in GRN concentrations was observed *(10)* in both presymptomatic and symptomatic *GRN* carriers. Further, we found increased concentrations of RNF13 in both presymptomatic and symptomatic *GRN* carriers, which might be reflecting an underlying alteration in the ubiquitin system *(49)*, not as well captured in *MAPT* and *C9orf72* carriers. Further, we discovered several proteins that were changed in *C9orf72* expansion carriers, including PGK1 which was elevated not only in presymptomatic carriers in comparison with non-carriers but also showed a stepwise increase across the disease continuum. In addition, CALB2 as well as HK1 were elevated in presymptomatic *C9orf72 c*arriers and, like PGK1, their relative abundances appeared to increase with disease progression. Both HK1 and PGK1 are key enzymes of the glycolysis pathway suggesting that a dysregulation of the glucose metabolism might be an early feature of *C9orf72*-related FTD *(50)*. HK1 and CALB2 were also selected as two of the top proteins in analyses comparing mutation carriers irrespective of underlying symptomatology, indicating their stronger association with a *C9orf72* mutation.

This study has limitations. The identification of a lower number of proteins that were changed in *MAPT* mutation carriers in comparison with *GRN* and *C9orf72* mutation carriers may be biased due to the lower number of participants in this group. *C9orf72* seems to be the most common genetic cause of FTD worldwide, followed by *GRN* and then *MAPT (1)*, and this trend is reflected in the recruitment of the GENFI study. Due to the structure of participant recruitment in the GENFI cohort, some participants from the same family were included in the study. Family members may share genetic and environmental factors to a greater degree than the general population, which may bias the results. However, we conducted sensitivity analyses that included only one member from each family, with comparable results.

Although genetic FTD offers the unique advantage of linking proteomic changes to pathological alterations *antemortem*, specifically distinguishing between tau and TDP-43 pathology, it cannot be excluded that observed proteomic changes are, in fact, specific to the underlying genetic mutation and not necessarily transferable to the resulting pathology in sporadic FTD.

Further, although both the EMIF and Higginbotham *et al.* studies employed similar statistical and mass spectrometric methods, it is likely that some of the differences seen between studies are due to varying power to detect proteomic alterations.

Finally, the age-difference between symptomatic carriers and non-carriers may have resulted in age influencing the interpretation of results. However, including age as a covariate in all relevant analyses is likely to mostly mitigate this potential issue.

To conclude, this study explored the CSF proteome in genetic FTD and found distinct changes occurring already in presymptomatic mutation carriers indicating early lysosomal dysfunction and alterations in proteins involved in glucose metabolism, with more widespread proteomic differences during the symptomatic stage of the disease. We found that proteomic profiles largely overlapped between the different causes of FTD as well as with AD, especially with respect to synaptic loss, glial responses and neurodegenerative processes. Furthermore, we discovered that certain lysosomal proteins are strongly associated with *MAPT*-mutation carriers, hinting at their potential value in distinguishing underlying FTD pathologies. Taken together, our results can inform the development of targeted assays that could be of value in clinical scenarios as well as in research aiming to better understand these diseases.

MATERIALS AND METHODS

Study design

The objective of this study was to explore the CSF proteomic signatures of the three most common genetic pathogenic mutations in FTD. To this end, 238 CSF samples from an ongoing case-control study cohort of genetic FTD, the GENFI cohort, were employed. Participants of the GENFI cohort were recruited from 14 GENFI centres, distributed across Europe and Canada, since 2012. One CSF sample per participant, generally obtained upon the first (baseline) visit, was included in the present cross-sectional study. The samples were randomised for measurement and the researchers were blinded for genetic status and genetic mutations at the time of the experiment. No prior calculations were performed to determine cohort size; all available samples were included in the study. The presence of batch effects and sample outliers were investigated using hierarchical clustering and principal component analysis before and after normalization. The proteomic experiments were conducted in one replicate. No participants were excluded. London Queen Square Ethics committee as well as local ethics committees at each site approved the study. The study complies with the Declaration of Helsinki. All participants provided written informed consent at enrolment including consent to publication. This study adhered to the STROBE reporting guidelines for observational studies.

Participants and sample collection

Participants were recruited from the GENFI study, which includes individuals with a diagnosis of FTD due to a pathogenic mutation in *MAPT*, *GRN*, or *C9orf72* (symptomatic mutation carriers), at-risk first-degree relatives (presymptomatic mutation carriers), and non-carriers (mutation-negative first-degree relatives from the same families). Demographics of the cohort are described in Table 1. Participants were assessed using a standardised history and examination and were classified as symptomatic if they met consensus diagnostic criteria *(51, 52)*. The CDR Dementia Staging Instrument with National Alzheimer Coordinating Centre Frontotemporal Lobar Degeneration component (CDR® plus NACC FTLD) was used to assess disease severity, and the CDR® plus NACC FTLD sum of boxes (SOB) was used for quantitative analyses in this paper. Participants

underwent Volumetric T1-weighted MRI scans. More details on clinical evaluation and imaging can be found in Supplementary Methods.

CSF collection and sample preparation

CSF was collected in polypropylene tubes through a lumbar puncture and centrifuged to remove insoluble material and cells. Supernatants were aliquoted and stored at -80 °C within 2 hours after collection. CSF samples (25 µL) were reduced by the addition of Tris(2)-carboxyethylphosphine (TCEP) in sodium deoxycholate (DOC), and triethylammonium bicarbonate (TEAB) to a final concentration of 5 mM TCEP (1% DOC, 100 mM TEAB). Following incubation at 55 °C for one hour, samples were equilibrated to room temperature (RT). Carbamidomethylation was performed by adding iodoacetamide to a concentration of 10 mM and subsequently incubating the reaction mixture in the dark for 30 min at RT. Trypsin (100 µg per vial; Promega) was dissolved in resuspension buffer (Promega) and 1.5 µg were added for overnight digestion at 37 °C. The following day, TMTpro reagents (TMT 18plex, Thermo Fisher, 5 mg) were dissolved in 200 µL acetonitrile (ACN) having been equilibrated to RT. Samples were randomised across TMT sets and TMT labelling was performed by adding 10 µL of TMT reagent to each sample. Per set, a global internal standard (GIS; pool of all cohort samples) was included as the last TMT channel (135N) for reference and normalisation. The reaction mixture was incubated for one hour under constant agitation and afterwards the labelling process was quenched by the addition of hydroxylamine to a final concentration of 0.2% (v/v). Following an incubation period of 30 min, samples were combined into 18-plex sets and subsequently acidified with 0.5 M hydrochloric acid to precipitate DOC as well as diluted with 0.1% trifluoroacetic acid (TFA). To remove DOC, TMT sets were centrifuged at 4000*g for 15 min at 4 °C and the resulting supernatant was subjected to desalting by solid phase extraction (SPE). Desalting was performed on reversed-phase C18 cartridges (Sep-Pak C18 light) with a vacuum manifold. The columns were first washed with 2*1000 µL 0.1% TFA in 80% ACN and then equilibrated with 2*1000 µL 0.1% TFA. After sample loading, the column was again washed twice with 1000 µL 0.1% TFA and finally peptides were eluted with 0.1% TFA, 80% ACN. The eluate was split into three aliquots of equal volume, dried by vacuum centrifugation, and stored at -20 °C.

572 Plasma NfL and other CSF marker measurements are detailed in Supplementary Methods.

573

574 **Offline high-pH reverse phase HPLC sample fractionation**

575 Offline high-pH HPLC fractionation was performed on an UltiMate™ 3000 Nano LC system. Each
576 TMT set aliquot was dissolved in 22 µL of 2.5 mM NH_4OH of which 20 µL were injected to be
577 separated on an XBridge BEH C18 column (pore size: 130 Å, inner diameter: 4.6 mm). Peptide
578 elution was accomplished using the following gradient: Buffer B was increased from 1% to 45% over
579 a 65-minute period (flow rate of 100 µL/min), while Buffer C was maintained at 10% (Buffer A: H_2O,
580 Buffer B: 84% ACN, Buffer C: 25 mM NH_4OH). Resulting fractions were collected circling over two
581 rows in a 96-well microtiter plate at 1 min intervals, yielding 24 concatenated fractions. Subsequent
582 column cleaning was performed at 90% B and 10% C for 10 minutes followed by an equilibration at
583 1% B and 10% C for 10 minutes. All fractions were subjected to vacuum centrifugation and stored
584 dry at -20 °C until subsequent LC-MS analysis.

585

586 **Liquid chromatography-mass spectrometry (LC-MS)**

587 Fractions were dissolved in 50 µL 0.05% TFA, 0.1% bovine serum albumin (loading buffer) and
588 loaded on a nano-LC (Ultimate RSLC Nano, Thermo Scientific) equipped with a C18 trap column
589 (PepMap Acclaim 300 µm mm * 5 mm, Thermo Scientific) and C18 separation column (PepMap
590 Acclaim 75 µm * 500 mm, Thermo Scientific), connected to an Orbitrap Fusion™ Lumos™ Tribrid™
591 mass spectrometer (Thermo Scientific), fitted with an Easy Spray Source and a high-field asymmetric
592 waveform ion mobility spectrometry (FAIMS) unit for spatial ion separation. Peptides were separated
593 according to the following gradient: 5 min, 4% B; 6 min, 10% B; 74 min, 40% B; 75 min, 100% B
594 (Buffer A: 0.1% FA; Buffer B: 84% ACN, 0.1% FA). In the positive ion mode, alternating MS/MS
595 cycles (cycle time = 1.5 s) were performed at compensation voltages (CV) of CV=-70 V, CV = -50 V.
596 A full Orbitrap MS scan was recorded with the parameters specified as follows: R = 120 k, AGC
597 target = 100%, max injection time = 50 ms. The full MS scan was then followed by data dependent
598 Orbitrap MS/MS scans set to the following parameters: R = 50 k, AGC target = 200%, max. injection
599 time = 120 ms, isolation window = 0.7 m/z, activation type = HCD.

Statistical analysis

All statistical analyses were performed with R version 4.1.2. For basic demographic variables, Omnibus Kruskal-Wallis tests were performed for continuous variables, whereas Fisher's exact tests were used for categorical variables. Unless otherwise specified, Spearman correlations were used to test associations between continuous variables. To assess differentially abundant proteins across the diagnostic groups, linear regression models were built with the $\log_2$-transformed value of the measured protein abundance as dependent variable, testing the effect of the diagnostic group, and adjusting for both age and sex as covariates. Resulting *P*-values were adjusted with the Benjamini-Hochberg procedure to account for multiple testing. Statistical significance (α) was set at a two-sided $P<0.05$. To ensure a minimum number of observations per group, proteins with a high fraction of missing values (>75% of participants) were excluded from the regression analysis. Additionally, group-wise outlier removal of protein measurements (+/- 1.5*IQR) was performed prior to regression analysis as the presence of outliers can severely affect resulting test statistics potentially increasing the rate of false negatives in the initial biomarker discovery phase. For all subsequent statistical analyses as well as boxplots shown in this paper, outliers were not removed. Linear models (also adjusted for age and sex) including only one member from each family were performed in comparisons when more than 5 participants were available in both groups. To identify mutation-specific signatures, linear models were fitted including protein abundance as a dependent variable while evaluating the effect of each mutation group including affectation (absence/presence of symptoms) as well as age and sex as covariates. To identify subsets of co-correlated proteins relating to pathophysiological features of genetic FTD, we performed network analysis (WGCNA) followed by GO annotation of the output modules. The prognostic properties of protein networks were evaluated using linear mixed effects models. The specifics of each of these methods are described in Supplementary Methods.

628

629

630

631

632

633

634

635

SUPPLEMENTARY MATERIALS

Supplementary methods

Fig. S1 – S25

References (53-56)

Data File S1 (Table S1-S10)

MDAR Reproducibility checklist

REFERENCES and Notes

1. C. V. Greaves, J. D. Rohrer, An update on genetic frontotemporal dementia. *J Neurol* **266**, 2075–2086 (2019).

2. S. Hendriks, K. Peetoom, C. Bakker, W. M. Van Der Flier, J. M. Papma, R. Koopmans, F. R. J. Verhey, M. De Vugt, S. Köhler, A. Withall, J. L. Parlevliet, Ö. Uysal-Bozkir, R. C. Gibson, S. M. Neita, T. R. Nielsen, L. C. Salem, J. Nyberg, M. A. Lopes, J. C. Dominguez, M. F. De Guzman, A. Egeberg, K. Radford, T. Broe, M. Subramaniam, E. Abdin, A. C. Bruni, R. Di Lorenzo, K. Smith, L. Flicker, M. O. Mol, M. Basta, D. Yu, G. Masika, M. S. Petersen, L. Ruano, Global Prevalence of Young-Onset Dementia: A Systematic Review and Meta-analysis. *JAMA Neurol* **78**, 1080–1090 (2021).

3. I. R. A. Mackenzie, M. Neumann, Molecular neuropathology of frontotemporal dementia: insights into disease mechanisms from postmortem studies. *J Neurochem* **138**, 54–70 (2016).

4. J. D. Rohrer, R. Guerreiro, J. Vandrovcova, J. Uphill, D. Reiman, J. Beck, A. M. Isaacs, A. Authier, R. Ferrari, N. C. Fox, I. R. A. MacKenzie, J. D. Warren, R. De Silva, J. Holton, T. Revesz, J. Hardy, S. Mead, M. N. Rossor, The heritability and genetics of frontotemporal lobar degeneration. *Neurology* **73**, 1451–1456 (2009).

5. C. H. van Dyck, Anti-Amyloid-β Monoclonal Antibodies for Alzheimer's Disease: Pitfalls and Promise. *Biol Psychiatry* **83**(4), 311-319 (2017).

6. M. R. Al Shweiki, P. Steinacker, P. Oeckl, B. Hengerer, A. Danek, K. Fassbender, J. Diehl-Schmid, H. Jahn, S. Anderl-Straub, A. C. Ludolph, C. Schönfeldt-Lecuona, M. Otto, Neurofilament light chain as a blood to differentiate psychiatric disorders from behavioural variant frontotemporal dementia. *J Psychiatr Res* **113**, 137–140 (2019).

7. J. C. Rojas, P. Wang, A. M. Staffaroni, C. Heller, Y. Cobigo, A. Wolf, S. Y. M. Goh, P. A. Ljubenkov, H. W. Heuer, J. C. Fong, J. B. Taylor, E. Veras, L. Song, A. Jeromin, D. Hanlon, L. Yu, A. Khinikar, R. Sivasankaran, A. Kieloch, M. A. Valentin, A. M. Karydas, L. L. Mitic, R. Pearlman, J. Kornak, J. H. Kramer, B. L. Miller, K. Kantarci, D. S. Knopman, N. Graff-Radford, L. Petrucelli, R. Rademakers, D. J. Irwin, M. Grossman, E. M. Ramos, G. Coppola, M. F. Mendez, Y. Bordelon, B. C. Dickerson, N. Ghoshal, E. D. Huey, I. R. Mackenzie, B. S. Appleby, K. Domoto-Reilly, G. Y. R. Hsiung, A. W. Toga, S. Weintraub, D. I. Kaufer, D. Kerwin, I. Litvan, C. U. Onyike, A. Pantelyat, E. D. Roberson, M. C. Tartaglia, T. Foroud, W. Chen, J. Czerkowicz, D. L. Graham, J. C. van Swieten, B. Borroni, R. Sanchez-Valle, F. Moreno, R. Laforce, C. Graff, M. Synofzik, D. Galimberti, J. B. Rowe, M. Masellis, E. Finger, R. Vandenberghe, A. de Mendonça, F. Tagliavini, I. Santana, S. Ducharme, C. R. Butler, A. Gerhard, J. Levin, A. Danek, M. Otto, S. Sorbi, D. M. Cash, R. S. Convery, M. Bocchetta, M. Foiani, C. V. Greaves, G. Peakman, L. Russell, I. Swift, E. Todd, J. D. Rohrer, B. F. Boeve, H. J. Rosen, A. L. Boxer, Plasma Neurofilament Light for Prediction of Disease Progression in Familial Frontotemporal Lobar Degeneration. *Neurology* **96**, e2296–e2312 (2021).

8. A. M. Staffaroni, M. Quintana, B. Wendelberger, H. W. Heuer, L. L. Russell, Y. Cobigo, A. Wolf, S. Y. M. Goh, L. Petrucelli, T. F. Gendron, C. Heller, A. L. Clark, J. C. Taylor, A. Wise, E. Ong, L. Forsberg, D. Brushaber, J. C. Rojas, L. VandeVrede, P. Ljubenkov, J. Kramer, K. B. Casaletto, B. Appleby, Y. Bordelon, H. Botha, B. C. Dickerson, K. Domoto-Reilly, J. A. Fields, T. Foroud, R. Gavrilova, D. Geschwind, N. Ghoshal, J. Goldman, J. Graff-Radford, N. Graff-Radford, M. Grossman, M. G. H. Hall, G. Y. Hsiung, E. D. Huey, D. Irwin, D. T. Jones, K. Kantarci, D. Kaufer, D. Knopman, W. Kremers, A. L. Lago, M. I. Lapid, I. Litvan, D. Lucente, I. R. Mackenzie, M. F. Mendez, C. Mester, B. L. Miller, C. U. Onyike, R. Rademakers, V. K. Ramanan, E. M. Ramos, M. Rao, K. Rascovsky, K. P. Rankin, E. D. Roberson, R. Savica, M. C. Tartaglia, S. Weintraub, B. Wong, D. M. Cash, A. Bouzigues, I. J. Swift, G. Peakman, M. Bocchetta, E. G. Todd, R. S. Convery, J. B. Rowe, B. Borroni, D. Galimberti, P. Tiraboschi, M. Masellis, E. Finger, J. C. van Swieten, H. Seelaar, L. C. Jiskoot, S. Sorbi, C. R. Butler, C. Graff, A. Gerhard, T. Langheinrich, R. Laforce, R. Sanchez-Valle, A. de Mendonça, F. Moreno, M. Synofzik, R. Vandenberghe, S. Ducharme, I. Le Ber, J. Levin, A. Danek, M. Otto, F. Pasquier, I. Santana, J. Kornak, B. F. Boeve, H. J. Rosen, J. D. Rohrer, A. L. Boxer, Temporal order of clinical and biomarker changes in familial frontotemporal dementia. *Nat Med* **28**, 2194–2206 (2022).

9. N. Finch, M. Baker, R. Crook, K. Swanson, K. Kuntz, R. Surtees, G. Bisceglio, A. Rovelet-Lecrux, B. Boeve, R. C. Petersen, D. W. Dickson, S. G. Younkin, V. Deramecourt, J. Crook, N. R. Graff-Radford, R. Rademakers, Plasma progranulin levels predict progranulin mutation status in frontotemporal dementia patients and asymptomatic family members. *Brain* **132**, 583–591 (2009).

10. L. H. H. Meeter, H. Patzke, G. Loewen, E. G. P. Dopper, Y. A. L. Pijnenburg, R. Van Minkelen, J. C. Van Swieten, Progranulin Levels in Plasma and Cerebrospinal Fluid in Granulin Mutation Carriers. *Dement Geriatr Cogn Dis Extra* **6**, 330–340 (2016).

11. M. Chatterjee, S. Özdemir, C. Fritz, W. Möbius, L. Kleineidam, E. Mandelkow, J. Biernat, C. Doğdu, O. Peters, N. C. Cosma, X. Wang, L. S. Schneider, J. Priller, E. Spruth, A. A. Kühn, P. Krause, T. Klockgether, I. R. Vogt, O. Kimmich, A. Spottke, D. C. Hoffmann, K. Fliessbach, C. Miklitz, C. McCormick, P. Weydt, B. Falkenburger, M. Brandt, R. Guenther, E. Dinter, J. Wiltfang, N. Hansen, M. Bähr, I. Zerr, A. Flöel, P. J. Nestor, E. Düzel, W. Glanz, E. Incesoy, K. Bürger, D. Janowitz, R. Perneczky, B. S. Rauchmann, F. Hopfner, O. Wagemann, J. Levin, S. Teipel, I. Kilimann, D. Goerss, J. Prudlo, T. Gasser, K. Brockmann, D. Mengel, M. Zimmermann, M. Synofzik, C. Wilke, J. Selma-González, J. Turon-Sans, M. A. Santos-Santos, D. Alcolea, S. Rubio-Guerra, J. Fortea, Á. Carbayo, A. Lleó, R. Rojas-García, I. Illán-Gala, M. Wagner, I. Frommann, S. Roeske, L. Bertram, M. T. Heneka, F. Brosseron, A. Ramirez, M. Schmid, R. Beschorner, A. Halle, J. Herms, M. Neumann, N. R. Barthélemy, R. J. Bateman, P. Rizzu, P. Heutink, O. Dols-Icardo, G. Höglinger, A. Hermann, A. Schneider, Plasma extracellular vesicle tau and TDP-43 as diagnostic biomarkers in FTD and ALS. *Nature Medicine* **30**, 1771–1783 (2024).

12. S. Bergström, L. Öijerstedt, J. Remnestål, J. Olofsson, A. Ullgren, H. Seelaar, J. C. van Swieten, M. Synofzik, R. Sanchez-Valle, F. Moreno, E. Finger, M. Masellis, C. Tartaglia, R. Vandenberghe, R. Laforce, D. Galimberti, B. Borroni, C. R. Butler, A. Gerhard, S. Ducharme, J. D. Rohrer, A. Månberg, C. Graff, P. Nilsson, L. Jiskoot, J. B. Rowe, A. de Mendonça, F. Tagliavini, I. Santana, I. Le Ber, J. Levin, A. Danek, M. Otto, G. Frisoni, R. Ghidoni, S. Sorbi, F. Pasquier, V. Jelic, C. Andersson, S. Afonso, M. R. Almeida, S. Anderl-Straub, A. Antonell, S. Archetti, A. Arighi, M. Balasa, M. Barandiaran, N. Bargalló, R. Bartha, B. Bender, A. Benussi, L. Benussi, V. Bessi, G. Binetti, S. Black, M. Bocchetta, S. Borrego-Ecija, J. Bras, R. Bruffaerts, M. Cañada, V. Cantoni, P. Caroppo, D. Cash, M. Castelo-Branco, R. Convery, T. Cope, G. Di Fede, A. Díez, D. Duro, C. Fenoglio, C. Ferrari, C. B. Ferreira, N. Fox, M. Freedman, G. Fumagalli, A. Gabilondo, R. Gasparotti, S. Gauthier, S. Gazzina, G. Giaccone, A. Gorostidi, C. Greaves, R. Guerreiro, C. Heller, T. Hoegen, B. Indakoetxea, L. Jiskoot, H. O. Karnath, R. Keren, T. Langheinrich, M. J. Leitão, A. Lladó, G. Lombardi, S. Loosli, C. Maruta, S. Mead, L. Meeter, G. Miltenberger, R. van Minkelen, S. Mitchell, K. Moore, B. Nacmias, J. Nicholas, J. Olives, S. Ourselin, A. Padovani, J. Panman, J. M. Papma, G. Peakman, M. Pievani, Y. Pijnenburg, C. Polito, E. Premi, S. Prioni, C. Prix, R. Rademakers, V. Redaelli, T. Rittman, E. Rogaeva, P. Rosa-Neto, G. Rossi, M. Rosser, B. Santiago, E. Scarpini, S. Schönecker, E. Semler, R. Shafei, C. Shoesmith, M. Tábuas-Pereira, M. Tainta, R. Taipa, D. Tang-Wai, D. L. Thomas, P. Thompson, H. Thonberg, C. Timberlake, P. Tiraboschi, E. Todd, P. Van Damme, M. Vandenbulcke, M. Veldsman, A. Verdelho, J. Villanua, J. Warren, C. Wilke, I. Woollacott, E. Wlasich, H. Zetterberg, M. Zulaica, A panel of CSF proteins separates genetic frontotemporal dementia from presymptomatic mutation carriers: a GENFI study. *Mol Neurodegener* **16**, 79 (2021).

13. J. Remnestål, L. Öijerstedt, A. Ullgren, J. Olofsson, S. Bergström, K. Kultima, M. Ingelsson, L. Kilander, M. Uhlén, A. Månberg, C. Graff, P. Nilsson, Altered levels of CSF proteins in patients with FTD, presymptomatic mutation carriers and non-carriers. *Transl Neurodegener* **9**, 27 (2020).

14. C. E. Teunissen, N. Elias, M. J. A. Koel-Simmelink, S. Durieux-Lu, A. Malekzadeh, T. V. Pham, S. R. Piersma, T. Beccari, L. H. H. Meeter, E. G. P. Dopper, J. C. van Swieten, C. R. Jimenez, Y. A. L. Pijnenburg, Novel diagnostic cerebrospinal fluid biomarkers for pathologic subtypes of frontotemporal dementia identified by proteomics. *Alzheimers Dement (Amst)* **2**, 86–94 (2016).

15. E. L. van der Ende, L. H. Meeter, C. Stingl, J. G. J. van Rooij, M. P. Stoop, D. A. T. Nijholt, R. Sanchez-Valle, C. Graff, L. Öijerstedt, M. Grossman, C. McMillan, Y. A. L. Pijnenburg, R. Laforce, G. Binetti, L. Benussi, R. Ghidoni, T. M. Luider, H. Seelaar, J. C. van Swieten, Novel CSF biomarkers in genetic frontotemporal dementia identified by proteomics. *Ann Clin Transl Neurol* **6**, 698–707 (2019).

16. N. Mattsson, U. Rüetschi, Y. A. L. Pijnenburg, M. A. Blankenstein, V. N. Podust, S. Li, I. Fagerberg, L. Rosengren, K. Blennow, H. Zetterberg, Novel cerebrospinal fluid biomarkers of axonal degeneration in frontotemporal dementia. *Mol Med Rep* **1**, 757–761 (2008).

17. P. Barschke, P. Oeckl, P. Steinacker, M. H. D. R. Al Shweiki, J. H. Weishaupt, G. B. Landwehrmeyer, S. Anderl-Straub, P. Weydt, J. Diehl-Schmid, A. Danek, J. Kornhuber, M. L. Schroeter, J. Prudlo, H. Jahn, K. Fassbender, M. Lauer, E. L. Van Der Ende, J. C. Van Swieten, A. E. Volk, A. C. Ludolph, M. Otto, Different CSF protein profiles in amyotrophic lateral sclerosis and frontotemporal dementia with C9orf72 hexanucleotide repeat expansion. *J Neurol Neurosurg Psychiatry* **91**, 503–511 (2020).

18. J. Nilsson, J. Gobom, S. Sjödin, G. Brinkmalm, N. J. Ashton, J. Svensson, P. Johansson, E. Portelius, H. Zetterberg, K. Blennow, A. Brinkmalm, Cerebrospinal fluid biomarker panel for synaptic dysfunction in Alzheimer's disease. Alzheimer's & Dementia: Diagnosis, *Assessment & Disease Monitoring* **13**, e12179 (2021).

19. F. Llorens, K. Thüne, W. Tahir, E. Kanata, D. Diaz-Lucena, K. Xanthopoulos, E. Kovatsi, C. Pleschka, P. Garcia-Esparcia, M. Schmitz, D. Ozbay, S. Correia, Â. Correia, I. Milosevic, O. Andréoletti, N. Fernández-Borges, I. M. Vorberg, M. Glatzel, T. Sklaviadis, J. M. Torres, S. Krasemann, R. Sánchez-Valle, I. Ferrer, I. Zerr, YKL-40 in the brain and cerebrospinal fluid of neurodegenerative dementias. *Mol Neurodegener* **12**, 83 (2017).

20. P. Oeckl, P. Weydt, D. R. Thal, J. H. Weishaupt, A. C. Ludolph, M. Otto, Proteomics in cerebrospinal fluid and spinal cord suggests UCHL1, MAP2 and GPNMB as biomarkers and underpins importance of transcriptional pathways in amyotrophic lateral sclerosis. *Acta Neuropathol* **139**, 119–134 (2020).

21. G. Ary, H. Sich, C. Larence, J. G. Ibbs, E. H. L. Ee, G. H. Arrington, The 14-3-3 Brain Protein in Cerebrospinal Fluid as a Marker for Transmissible Spongiform Encephalopathies, *N Engl J Med.* **335**, 924–930 (1996).

22. S. Palmqvist, S. Janelidze, Y. T. Quiroz, H. Zetterberg, F. Lopera, E. Stomrud, Y. Su, Y. Chen, G. E. Serrano, A. Leuzy, N. Mattsson-Carlgren, O. Strandberg, R. Smith, A. Villegas, D. Sepulveda-Falla, X. Chai, N. K. Proctor, T. G. Beach, K. Blennow, J. L. Dage, E. M. Reiman, O. Hansson, Discriminative Accuracy of Plasma Phospho-tau217 for Alzheimer Disease vs Other Neurodegenerative Disorders. *JAMA* **324**, 772–781 (2020).

23. N. K. Magdalinou, A. J. Noyce, R. Pinto, E. Lindstrom, J. Holmén-Larsson, M. Holtta, K. Blennow, H. R. Morris, T. Skillbäck, T. T. Warner, A. J. Lees, I. Pike, M. Ward, H. Zetterberg, J. Gobom, Identification of candidate cerebrospinal fluid biomarkers in parkinsonism using quantitative proteomics. *Parkinsonism Relat Disord* **37**, 65–71 (2017).

24. S. Weiner, M. Sauer, P. J. Visser, B. M. Tijms, E. Vorontsov, K. Blennow, H. Zetterberg, J. Gobom, Optimized sample preparation and data analysis for TMT proteomic analysis of cerebrospinal fluid applied to the identification of Alzheimer's disease biomarkers. *Clin Proteomics* **19**, 13 (2022).

25. B. M. Tijms, J. Gobom, L. Reus, I. Jansen, S. Hong, V. Dobricic, F. Kilpert, M. Ten Kate, F. Barkhof, M. Tsolaki, F. R. J. Verhey, J. Popp, P. Martinez-Lage, R. Vandenberghe, A. Lleó, J. L. Molinuevo, S. Engelborghs, L. Bertram, S. Lovestone, J. Streffer, S. Vos, I. Bos, K. Blennow, P. Scheltens, C. E. Teunissen, H. Zetterberg, P. J. Visser, Pathophysiological subtypes of Alzheimer's disease based on cerebrospinal fluid proteomics. *Brain* **143**, 3776–3792 (2020).

26. A. Sogorb-Esteve, J. Nilsson, I. J. Swift, C. Heller, M. Bocchetta, L. L. Russell, G. Peakman, R. S. Convery, J. C. van Swieten, H. Seelaar, B. Borroni, D. Galimberti, R. Sanchez-Valle, R. Laforce, F. Moreno, M. Synofzik, C. Graff, M. Masellis, M. C. Tartaglia, J. B. Rowe, R. Vandenberghe, E. Finger, F. Tagliavini, I. Santana, C. R. Butler, S. Ducharme, A. Gerhard, A. Danek, J. Levin, M. Otto, S. Sorbi, I. Le Ber, F. Pasquier, J. Gobom, A. Brinkmalm, K. Blennow, H. Zetterberg, J. D. Rohrer, A. Nelson, A. Bouzigues, C. V. Greaves, D. Cash, D. L. Thomas, E. Todd, H. Benotmane, J. Nicholas, K. Samra, R. Shafei, C. Timberlake, T. Cope, T. Rittman, A. Benussi, E. Premi, R. Gasparotti, S. Archetti, S. Gazzina, V. Cantoni, A. Arighi, C. Fenoglio, E. Scarpini, G. Fumagalli, V. Borracci, G. Rossi, G. Giaccone, G. Di Fede, P. Caroppo, P. Tiraboschi, S. Prioni, V. Redaelli, D. Tang-Wai, E. Rogaeva, M. Castelo-Branco, M. Freedman, R. Keren, S. Black, S. Mitchell, C. Shoesmith, R. Bartha, R. Rademakers, J. Poos, J. M. Papma, L. Giannini, R. van Minkelen, Y. Pijnenburg, B. Nacmias, C. Ferrari, C. Polito, G. Lombardi, V. Bessi, M. Veldsman, C. Andersson, H. Thonberg, L. Öijerstedt, V. Jelic, P. Thompson, T. Langheinrich, A. Lladó, A. Antonell, J. Olives, M. Balasa, N. Bargalló, S. Borrego-Ecija, A. de Mendonça, A. Verdelho, C. Maruta, C. B. Ferreira, G. Miltenberger, F. S. do Couto, A. Gabilondo, A. Gorostidi, J. Villanua, M. Cañada, M. Tainta, M. Zulaica, M. Barandiaran, P. Alves, B. Bender, C. Wilke, L. Graf, A. Vogels, M. Vandenbulcke, P. Van Damme, R. Bruffaerts, K. Poesen, P. Rosa-Neto, S. Gauthier, A. Camuzat, A. Brice, A. Bertrand, A. Funkiewiez, D. Rinaldi, D. Saracino, O. Colliot, S. Sayah, C. Prix, E. Wlasich, O. Wagemann, S. Loosli, S. Schönecker, T. Hoegen, J. Lombardi, S. Anderl-Straub, A. Rollin, G. Kuchcinski, M. Bertoux, T. Lebouvier, V. Deramecourt, B. Santiago, D. Duro, M. J. Leitão, M. R. Almeida, M. Tábuas-Pereira, S. Afonso, Differential impairment of cerebrospinal fluid synaptic biomarkers in the genetic forms of frontotemporal dementia. *Alzheimers Res Ther* **14**, 118 (2022).

27. I. Pesämaa, S. A. Müller, S. Robinson, A. Darcher, D. Paquet, H. Zetterberg, S. F. Lichtenthaler, C. Haass, A microglial activity state biomarker panel differentiates FTD-granulin and Alzheimer's disease patients from controls. *Mol Neurodegener* **18**, 1–18 (2023).

28. L. Higginbotham, L. Ping, E. B. Dammer, D. M. Duong, M. Zhou, M. Gearing, C. Hurst, J. D. Glass, S. A. Factor, E. C. B. Johnson, I. Hajjar, J. J. Lah, A. I. Levey, N. T. Seyfried, Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic. *Alzheimer's disease. Sci. Adv* **6**, eaaz9360 (2020).

29. E. Portelius, B. Olsson, K. Höglund, N. C. Cullen, H. Kvartsberg, U. Andreasson, H. Zetterberg, Å. Sandelius, L. M. Shaw, V. M. Y. Lee, D. J. Irwin, M. Grossman, D. Weintraub, A. Chen-Plotkin, D. A. Wolk, L. McCluskey, L. Elman, J. McBride, J. B. Toledo, J. Q. Trojanowski, K. Blennow, Cerebrospinal fluid neurogranin concentration in neurodegeneration: relation to clinical phenotypes and neuropathology. *Acta Neuropathol* **136**, 363–376 (2018).

30. N. J. Ashton, S. Janelidze, A. Al Khleifat, A. Leuzy, E. L. van der Ende, T. K. Karikari, A. L. Benedet, T. A. Pascoal, A. Lleó, L. Parnetti, D. Galimberti, L. Bonanni, A. Pilotto, A. Padovani, J. Lycke, L. Novakova, M. Axelsson, L. Velayudhan, G. D. Rabinovici, B. Miller, C. Pariante, N. Nikkheslat, S. M. Resnick, M. Thambisetty, M. Schöll, G. Fernández-Eulate, F. J. Gil-Bea, A. López de Munain, A. Al-Chalabi, P. Rosa-Neto, A. Strydom, P. Svenningsson, E. Stomrud, A. Santillo, D. Aarsland, J. C. van Swieten, S. Palmqvist, H. Zetterberg, K. Blennow, A. Hye, O. Hansson, A multicentre validation study of the diagnostic value of plasma neurofilament light. *Nat Commun* **12**, 3400 (2021).

31. Illán-Gala I, Alcolea D, Montal V, Dols-Icardo O, Muñoz L, de Luna N, Turón-Sans J, Cortés-Vicente E, Sánchez-Saudinós MB, Subirana A, Sala I, Blesa R, Clarimón J, Fortea J, Rojas-García R, Lleó A. CSF sAPPβ, YKL-40, and NfL along the ALS-FTD spectrum. *Neurology* **91**, e1619-e1628 (2018).

32. E. L. Van Der Ende, M. Xiao, D. Xu, J. M. Poos, J. L. Panman, L. C. Jiskoot, L. H. Meeter, E. G. P. Dopper, J. M. Papma, C. Heller, R. Convery, K. Moore, M. Bocchetta, M. Neason, G. Peakman, D. M. Cash, C. E. Teunissen, C. Graff, M. Synofzik, F. Moreno, E. Finger, R. Sánchez-Valle, R. Vandenberghe, R. Laforce, M. Masellis, M. C. Tartaglia, J. B. Rowe, C. R. Butler, S. Ducharme, A. Gerhard, A. Danek, J. Levin, Y. A. L. Pijnenburg, M. Otto, B. Borroni, F. Tagliavini, A. De Mendonca, I. Santana, D. Galimberti, H. Seelaar, J. D. Rohrer, P. F. Worley, J. C. Van Swieten, Neuronal pentraxin 2: A synapse-derived CSF biomarker in genetic frontotemporal dementia. *J Neurol Neurosurg Psychiatry* **91**, 612-621 (2020).

33. J. Nilsson, J. Constantinescu, B. Nellgård, P. Jakobsson, W. S. Brum, J. Gobom, L. Forsgren, K. Dalla, R. Constantinescu, H. Zetterberg, O. Hansson, K. Blennow, D. Bäckström, A. Brinkmalm, Cerebrospinal Fluid Biomarkers of Synaptic Dysfunction are Altered in Parkinson's Disease and Related Disorders. *Mov Disord* **38**, 267–277 (2023).

34. E. C. B. Johnson, E. B. Dammer, D. M. Duong, L. Ping, M. Zhou, L. Yin, L. A. Higginbotham, A. Guajardo, B. White, J. C. Troncoso, M. Thambisetty, T. J. Montine, E. B. Lee, J. Q. Trojanowski, T. G. Beach, E. M. Reiman, V. Haroutunian, M. Wang, E. Schadt, B. Zhang, D. W. Dickson, N. Ertekin-Taner, T. E. Golde, V. A. Petyuk, P. L. De Jager, D. A. Bennett, T. S. Wingo, S. Rangaraju, I. Hajjar, J. M. Shulman, J. J. Lah, A. I. Levey, N. T. Seyfried, Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat Med* **26**, 769-780 (2020).

35. S. Rangaraju, E. B. Dammer, S. A. Raza, P. Rathakrishnan, H. Xiao, T. Gao, D. M. Duong, M. W. Pennington, J. J. Lah, N. T. Seyfried, A. I. Levey, Identification and therapeutic modulation of a pro-inflammatory subset of disease-associated-microglia in Alzheimer's disease. *Mol Neurodegener* **13**, 24 (2018).

36. C. J. Garwood, L. E. Ratcliffe, J. E. Simpson, P. R. Heath, P. G. Ince, S. B. Wharton, Review: Astrocytes in Alzheimer's disease and other age-associated dementias: a supporting player with a central role. *Neuropathol Appl Neurobiol* **43**, 281–298 (2017).

37. M. Huang, E. Modeste, E. Dammer, P. Merino, G. Taylor, D. M. Duong, Q. Deng, C. J. Holler, M. Gearing, D. Dickson, N. T. Seyfried, T. Kukar, Network analysis of the progranulin-deficient mouse brain proteome reveals pathogenic mechanisms shared in human frontotemporal dementia caused by GRN mutations. *Acta Neuropathol Commun* **8**, 163 (2020).

38. S. Oh, Y. Jang, C. H. Na, Discovery of Biomarkers for Amyotrophic Lateral Sclerosis from Human Cerebrospinal Fluid Using Mass-Spectrometry-Based Proteomics. *Biomedicines* **11**, 1250 (2023).

39. R. Mao, Y. Wang, F. Wang, L. Zhou, S. Yan, S. Lu, W. Shi, Y. Zhang, Identification of Four Biomarkers of Human Skin Aging by Comprehensive Single Cell Transcriptome, Transcriptome, and Proteomics. *Front Genet* **13**, 881051 (2022).

40. K. Kawane, H. Fukuyama, G. Kondoh, J. Takeda, Y. Ohsawa, Y. Uchiyama, S. Nagata, Requirement of DNase II for definitive erythropoiesis in the mouse fetal liver. *Science* **292**, 1546–1549 (2001).

41. H. D. Shin, B. L. Park, H. S. Cheong, H. S. Lee, J. B. Jun, S. C. Bae, DNase II polymorphisms associated with risk of renal disorder among systemic lupus erythematosus patients. *J Hum Genet* **50**, 107–111 (2005).

42. P. A. Keyel, Dnases in health and disease. *Dev Biol* **429**, 1–11 (2017).

43. J. Root, P. Merino, A. Nuckols, M. Johnson, T. Kukar, Lysosome dysfunction as a cause of neurodegenerative diseases: Lessons from frontotemporal dementia and amyotrophic lateral sclerosis. *Neurobiol Dis* **154**, 105360 (2021).

44. C. J. Holler, G. Taylor, Q. Deng, T. Kukar, Intracellular Proteolysis of Progranulin Generates Stable, Lysosomal Granulins that Are Haploinsufficient in Patients with Frontotemporal Dementia Caused by GRN Mutations. eNeuro 4 (2017).

45. M. Hutton, C. L. Lendon, P. Rizzu, M. Baker, S. Froelich, H. Houlden, S. Pickering-Brown, S. Chakraverty, A. Isaacs, A. Grover, J. Hackett, J. Adamson, S. Lincoln, D. Dickson, P. Davies, R. C. Petersen, M. Stevens, E. de Graaff, E. Wauters, J. van Baren, M. Hillebrand, M. Joosse, J. M. Kwon, P. Nowotny, L. K. Che, J. Norton, J. C. Morris, L. A. Reed, J. Trojanowski, H. Basun, L. Lannfelt, M. Neystat, S. Fahn, F. Dark, T. Tannenberg, P. R. Dodd, N. Hayward, J. B. Kwok, P. R. Schofield, A. Andreadis, J. Snowden, D. Craufurd, D. Neary, F. Owen, B. A. Oostra, J. Hardy, A. Goate, J. van Swieten, D. Mann, T. Lynch, P. Heutink, Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* **393**, 702–5 (1998).

46. F. Lim, F. Hernández, J. J. Lucas, P. Gómez-Ramos, M. A. Morán, J. Ávila, FTDP-17 mutations in tau transgenic mice provoke lysosomal abnormalities and Tau filaments in forebrain. *Mol Cell Neurosci* **18**, 702–714 (2001).

47. C. D. Pacheco, M. J. Elrick, A. P. Lieberman, Tau normal function influences Niemann-Pick type C disease pathogenesis in mice and modulates autophagy in NPC1-deficient cells. *Autophagy* **5**, 548–550 (2009).

48. N. Ashtari, X. Jiao, M. Rahimi-Balaei, S. Amiri, S. E. Mehr, B. Yeganeh, H. Marzban, Lysosomal Acid Phosphatase Biosynthesis and Dysfunction: A Mini Review Focused on Lysosomal Enzyme Dysfunction in Brain. *Curr Mol Med* **16**, 439–446 (2016).

49. Q. Zhang, Y. Meng, L. Zhang, J. Chen, D. Zhu, RNF13: a novel RING-type ubiquitin ligase over-expressed in pancreatic cancer. *Cell Res* **19**, 348–357 (2009).

50. J. M. Berg, J. L. Tymoczko, L. Stryer, Biochemistry, Fifth Edition. Freeman W. H. And Company, Ed. (New York, ed. 5th, 2002).

51. M. L. Gorno-Tempini, A. E. Hillis, S. Weintraub, A. Kertesz, M. Mendez, S. F. Cappa, J. M. Ogar, J. D. Rohrer, S. Black, B. F. Boeve, F. Manes, N. F. Dronkers, R. Vandenberghe, K. Rascovsky, K. Patterson, B. L. Miller, D. S. Knopman, J. R. Hodges, M. M. Mesulam, M. Grossman, Classification of primary progressive aphasia and its variants. *Neurology* **76**, 1006–1014 (2011).

52. K. Rascovsky, J. R. Hodges, D. Knopman, M. F. Mendez, J. H. Kramer, J. Neuhaus, J. C. Van Swieten, H. Seelaar, E. G. P. Dopper, C. U. Onyike, A. E. Hillis, K. A. Josephs, B. F. Boeve, A. Kertesz, W. W. Seeley, K. P. Rankin, J. K. Johnson, M. L. Gorno-Tempini, H. Rosen, C. E. Prioleau-Latham, A. Lee, C. M. Kipps, P. Lillo, O. Piguet, J. D. Rohrer, M. N. Rossor, J. D. Warren, N. C. Fox, D. Galasko, D. P. Salmon, S. E. Black, M. Mesulam, S. Weintraub, B. C. Dickerson, J. Diehl-Schmid, F. Pasquier, V. Deramecourt, F. Lebert, Y. Pijnenburg, T. W. Chow, F. Manes, J. Grafman, S. F. Cappa, M. Freedman, M. Grossman, B. L. Miller, Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* **134**, 2456–2477 (2011).

53. M. J. Cardoso, R. Wolz, M. Modat, N. C. Fox, D. Rueckert, S. Ourselin, Geodesic information flows. *Med Image Comput Comput Assist Interv* **15**, 262–270 (2012).

54. I. B. Malone, K. K. Leung, S. Clegg, J. Barnes, J. L. Whitwell, J. Ashburner, N. C. Fox, G. R. Ridgway, Accurate automatic estimation of total intracranial volume: a nuisance variable with less nuisance. *Neuroimage* **104**, 366–372 (2015).

55. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 1–13 (2008).

56. U. Raudvere, L. Kolberg, I. Kuzmin, T. Arak, P. Adler, H. Peterson, J. Vilo, g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191–W198 (2019).

Acknowledgments

We thank Ida Pesämaa for discussions around the work with microglial activation-dependent markers, from which the published results were used as an external validation cohort of this study.

Funding

This work is supported by Race Against Dementia fellowship, supported by Alzheimer's Research UK (ARUK-RADF2021A-003 to AS-E). The UK Dementia Research Institute which receives its funding from DRI Ltd, funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK to AS-E. The Dementia Research Centre is supported by Alzheimer's Research UK, Alzheimer's Society, Brain Research UK, and The Wolfson Foundation.
Co-authors in the manuscript are supported by the Gothenburg Medical Society (Göteborgs Läkaresällskap, #GLS-988641 to JS). HZ is a Wallenberg Scholar and a Distinguished Professor at the Swedish Research Council supported by grants from the Swedish Research Council (#2023-00356; #2022-01018 and #2019-02397 to HZ), the European Union's Horizon Europe research and innovation programme under grant agreement No 101053962 to HZ, Swedish State Support for Clinical Research (#ALFGBG-71320 to HZ), the Alzheimer Drug Discovery Foundation (ADDF), USA (#201809-2016862 to HZ), the AD Strategic Fund and the Alzheimer's Association (#ADSF-21-831376-C, #ADSF-21-831381-C, #ADSF-21-831377-C, and #ADSF-24-1284328-C to HZ), the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States (NEuroBioStand, #22HLT07 to HZ), the Bluefield Project, Cure Alzheimer's Fund, the Olav Thon Foundation, the Erling-Persson Family Foundation, Familjen Rönströms Stiftelse, Stiftelsen för Gamla Tjänarinnor, Hjärnfonden, Sweden (#FO2022-0270 to HZ), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860197 (MIRIADE), the European Union Joint Programme – Neurodegenerative Disease Research (JPND2021-00694 to HZ), the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre, and the UK Dementia Research Institute at UCL (UKDRI-1003 to HZ).

KB is supported by the Swedish Research Council (#2017-00915 and #2022-00732 to KB), the Swedish Alzheimer Foundation (#AF-930351, #AF-939721 and #AF-968270 to KB), Hjärnfonden, Sweden (#FO2017-0243 and #ALZ2022-0006 to KB), the Swedish state under the agreement between the Swedish government and the County Councils, the ALF-agreement (#ALFGBG-715986 and #ALFGBG-965240 to KB), the European Union Joint Program for Neurodegenerative Disorders (JPND2019-466-236 to KB), the Alzheimer's Association 2021 Zenith Award (ZEN-21-848495 to KB), and the Alzheimer's Association 2022-2025 Grant (SG-23-1038904 QC to KB).
JCVS was supported by the Dioraphte Foundation grant 09-02-03-00, Association for Frontotemporal Dementias Research Grant 2009, Netherlands Organization for Scientific Research grant HCMI 056-13-018, ZonMw Memorabel (Deltaplan Dementie, project number 733 051 042 to JCVS), Alzheimer Nederland and the Bluefield Project. FM received funding from the Tau Consortium and the Center for Networked Biomedical Research on Neurodegenerative Disease. RS-V is supported by Alzheimer's Research UK Clinical Research Training Fellowship (ARUK-CRF2017B-2 to RS-V) and has received funding from Fundació Marató de TV3, Spain (grant no. 20143810 to RS-V). DG received support from the EU Joint Programme-Neurodegenerative Disease Research and the Italian Ministry of Health (PreFrontALS) grant 733051042 to DG. CG received funding from EU Joint Programme-Neurodegenerative Disease Research -Prefrontals VR Dnr 529-2014-7504, VR 2015-02926 and 2018-02754 to CG, the Swedish FTD Inititative-Schörling Foundation, Alzheimer Foundation, Brain Foundation and Stockholm County Council ALF. MM has received funding from a Canadian Institute of Health Research operating grant and the Weston Brain Institute and Ontario Brain Institute. JBR has received funding from the Welcome Trust (220258 to JBR), the Bluefield Project, and is supported by the Cambridge University Centre for Frontotemporal Dementia, the Medical Research Council (MC_UU_00030/14; MR/T033371/1) and the National Institute for Health Research Cambridge Biomedical Research Centre (NIHR203312).. EF has received funding from a Canadian Institute of Health Research grant #327387 to EF. RV has received funding from the Mady Browaeys Fund for Research into Frontotemporal Dementia. JL received funding for this work by the Deutsche Forschungsgemeinschaft German Research Foundation under Germany's Excellence Strategy within the framework of the Munich Cluster for Systems Neurology (EXC 2145 SyNergy—

ID 390857198 to JL). MO has received funding from Germany's Federal Ministry of Education and Research (BMBF). JDR is supported by the Bluefield Project and the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre and has received funding from an MRC Clinician Scientist Fellowship (MR/M008525/1 to JDR) and a Miriam Marks Brain Research UK Senior Fellowship. DA received funding by the Fondation Recherche Alzheimer, and the Swiss National Science Foundation (project CRSK-3_196354/1 to DA). JG is supported by Alzheimerfonden (AF-980746 to JG) and Stiftelsen för Gamla tjänarinnor (2022-01324 to JG). Several authors of this publication are members of the European Reference Network for Rare Neurological Diseases (ERN-RND) - Project ID No 739510 to JCVS, MS, RV, AdM, MO, RV and JDR. This work was also supported by the EU Joint Programme—Neurodegenerative Disease Research GENFI-PROX grant [2019-02248; to JDR, MO, BB, CG, JCVS and MS], and by the Clinician Scientist programme "PRECISE.net" funded by the Else Kröner-Fresenius-Stiftung (to CW, DM, and MS).

Authors contributions

AS-E, JDR and JG designed the study. SW, AS-E, and JS performed experiments. JG supervised the analytical study. SW analyzed the data, AS-E and JS assisted with data analyses as well as visualisation. AS-E, SW, and JS wrote the manuscript; JG, JDR, KB and HZ revised the manuscript and contributed to data interpretation. The rest of GENFI consortium authors have contributed with the recruitment of participants and processing of samples at their sites, as well as sending the samples and providing corresponding demographic data of their participants. All authors read and approved the final manuscript.

Conflicts of interest

HZ has served at scientific advisory boards andas a consultant for Abbvie, Acumen, Alector, Alzinova, ALZPath, Amylyx, Annexon, Apellis, Artery Therapeutics, AZTherapies, Cognito

Therapeutics, CogRx, Denali, Eisai, LabCorp, Merry Life, Nervgen, Novo Nordisk, Optoceutics, Passage Bio, Pinteon Therapeutics, Prothena, Red Abbey Labs, reMYND, Roche, Samumed, Siemens Healthineers, Triplet Therapeutics, and Wave, has given lectures in symposia sponsored by Alzecure, Biogen, Cellectricon, Fujirebio, Lilly, Novo Nordisk, and Roche, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program (outside submitted work).

KB has served as a consultant and at advisory boards for Acumen, ALZPath, BioArctic, Biogen, Eisai, Lilly, Moleac Pte. Ltd, Novartis, Ono Pharma, Prothena, Roche Diagnostics, and Siemens Healthineers; has served at data monitoring committees for Julius Clinical and Novartis; has given lectures, produced educational materials and participated in educational programs for AC Immune, Biogen, Celdara Medical, Eisai and Roche Diagnostics; and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program, outside the work presented in this paper.

MS has received consultancy honoraria from Ionis, UCB, Prevail, Orphazyme, Biogen, Servier, Reata, GenOrph, AviadoBio, Biohaven, Zevra, Lilly, and Solaxa, all unrelated to the present manuscript. JBR has provided consultancy or advisory board input to Alector, Asceneuron, Astronautx, Astex, CumulusNeuro, Cerevance, Clinical Ink, Curasen, Eisai, WAVE, unrelated to the current work. SD has provided paid consultancy to QuRALIS, Eisai, Eli Lilly and has received speaker fees from Eisai. SD is a SDMB member of Aviado Bio, IntelGenX. RV's institution has a clinical trial agreement (RV as PI) with Alector, AviadioBio, Denali, EliLilly, J&J, UCB. RV's institution has a consultancy agreement (RV as DSMB chair) with AC Immune. LR is a consultant for Prevail Therapeutics.

Data and materials availability

All data associated with this study are present in the paper or supplementary materials. Proteomic data underlying the results of this study have been deposited in the Dryad repository (DOI: 10.5061/dryad.r7sqv9snk). Individual identifiers have not been included to protect the privacy of study participants but can be requested from the authors. No custom code was used for data analysis;

respective R packages have been indicated where applicable. The data used for external validation of the *GRN* results as well as for AD-FTD comparisons are publicly available in the supplementary sections of the respective publications by Pesämaa *et al*., Tijms *et al.* and Higginbotham *et al*.

In addition to members of GENFI who are co-authors, the following members are collaborators who have contributed to the study design, data analysis, and interpretation: David L Thomas[2], Thomas Cope[41], Timothy Rittman[41], Alberto Benussi[46], Enrico Premi[47], Roberto Gasparotti[46], Silvana Archetti[47], Stefano Gazzina[47], Valentina Cantoni[46], Andrea Arighi[10, 11], Chiara Fenoglio[10, 11], Elio Scarpini[10, 11], Giorgio Fumagalli[10, 11], Vittoria Borracci[10, 11], Giacomina Rossi[16], Giorgio Giaccone[16], Giuseppe Di Fede[16], Paola Caroppo[16], Pietro Tiraboschi[16], Sara Prioni[16], Veronica Radaaelli[16], David Tang-Wai[54], Ekaterina Rogaeva[39], Michel Castelo-Branco[17], Morris Freedman[55], Ron Keren[54], Sandra Black[40], Sara Mitchell[40], Christen Shoesmith[38], Robart Bartha[56, 57], Rosa Rademakers[58], Jackie Poos[5], Janne M. Papma[5], Lucia Giannini[5], Rick can Minkelen[59], Yolande Pijnenburg[60], Benedetta Nacmias[61], Camilla Ferrari[61], Cristina Polito[62], Gemma Lombardi[61], Valentina Bessi[61], Michele Veldsman[33], Christin Andersson[63], Hakan Thonberg[8], Linn Öijerstedt[8, 9], Vesna Jelic[64], Paul Thompson[19], Tobias Langheinrich[19, 65], Abert Lladó[6], Anna Antonell[6], Jaume Olives[6], Mircea Balasa[6], Nuria Bargalló[66], Sergi Borrego-Écija[6], Ana Verdelho[67], Carolina Maruta[68], Catarina B. Ferreira[69], Gabriel Miltenberger[15], Frederico Simões do Couto[70], Alazne Gabilondo[44, 45], Jorge Villanua[71], Marta Cañada[72], Mikel Tainta[45], Miren Zulaica[45], Myriam Barandiaran[44, 45], Patricial Alves[45, 73], Benjamin Bender[74], Carlo Wilke[42, 43], Lisa Graf[42], Annick Vogels[75], Mathieu Vandenbulcke[76, 77], Philip van Damme[13, 78], Rose Buffaerts[79, 80], Koen Poesen[81], Pedro Rosa-Neto[82], Serge Gauthier[83], Agnès Camuzat[35], Alexis Brice[35, 36], Anne Bertrand[35, 84, 85], Aurélie Funkiewiez[35, 36], Daisy Rinaldi[35, 36], Dario Saracino[35, 36], Olivier Colliot[35, 84], Sabrina Sayah[35], Catharina Prix[21], Elisabeth Wlasich[21], Olivia Wagemann[21], Sandra Loosli[21], Sonja Schönecker[21], Tobias Hoegen[21], Jolina Lombardi[26], Sarah Anderl-Straub[26], Adeline Rollin[30], Gregory Kuchcinski[28, 30], Maxime Bertoux[29, 30], Thibaud Lebouvier[28, 29, 30], Vincent Deramecourt[28, 29, 30], Beatriz Santiago[17], Diana Duro[17], Maria João Leitão[18], Maria Rosario Almeida[17], Miguel Tábuas-Pereira[17], Sónia Afonso[86].

Affiliations 1 to 53 can be found on the first page of the paper.

[54]The University Health Network, Krembil Research Institute, M5T 0S8, Toronto, Canada

[55]Baycrest Health Sciences, Rotman Research Institute, University of Toronto, M5S 3E6, Toronto, Canada

[56]Department of Medical Biophysics, The University of Western Ontario, N6A 5C1, London, Ontario, Canada

[57]Centre for Functional and Metabolic Mapping, Robarts Research Institute, The University of Western Ontario, N6A 5B7, London, Ontario, Canada

[58]Center for Molecular Neurology, University of Antwerp, 2650, Edegem, Belgium

[59]Department of Clinical Genetics, Erasmus Medical Center, 3000, Rotterdam, Netherlands

[60]Amsterdam University Medical Centre, Amsterdam VUmc, 1081 HV, Amsterdam, Netherlands

[61]Department of Neuroscience, Psychology, Drug Research and Child Health, University of Florence, 50139, Florence, Italy

[62]Department of Biomedical, Experimental and Clinical Sciences “Mario Serio”, Nuclear Medicine Unit, University of Florence, 50139, Florence, Italy

[63]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, 171 65, Solna, Sweden

[64]Division of Clinical Geriatrics, Karolinska Institutet, Stockholm, 171 65, Solna, Sweden

[65]Manchester Centre for Clinical Neurosciences, Department of Neurology, Salford Royal NHS Foundation Trust, M6 8HD, Manchester, UK

[66]Imaging Diagnostic Center, Hospital Clínic, 08036, Barcelona, Spain

[67]Department of Neurosciences and Mental Health, Centro Hospitalar Lisboa Norte - Hospital de Santa Maria & Faculty of Medicine, University of Lisbon, 1649-028, Lisbon, Portugal

[68]Laboratory of Language Research, Centro de Estudos Egas Moniz, Faculty of Medicine, University of Lisbon, 2829 – 511, Lisbon, Portugal

[69]Laboratory of Neurosciences, Faculty of Medicine, University of Lisbon, 1649-028, Lisbon, Portugal

[70]Faculdade de Medicina, Universidade Católica Portuguesa, 2635-631, Rio de Mouro, Portugal71. OSATEK, University of Donostia, San Sebastian, 20014, Gipuzkoa, Spain

[72]CITA Alzheimer, San Sebastian, 20009, Gipuzkoa, Spain

[73]Department of Educational Psychology and Psychobiology, Faculty of Education, International University of La Rioja, 26006, Logroño, Spain

[74]Department of Diagnostic and Interventional Neuroradiology, University of Tübingen, 72074, Tübingen

[75]Department of Human Genetics, KU Leuven, 3000, Leuven, Belgium

[76]Geriatric Psychiatry Service, University Hospitals Leuven, 3000, Leuven, Belgium

[77]Neuropsychiatry, Department of Neurosciences, KU Leuven, 3000, Leuven, Belgium

[78]Laboratory for Neurobiology, VIB-KU Leuven Centre for Brain Research, 3000, Leuven, Belgium

[79]Department of Biomedical Sciences, University of Antwerp, 2000, Antwerp, Belgium

[80]Biomedical Research Institute, Hasselt University, 3500 Hasselt, Belgium

[81]Laboratory for Molecular Neurobiomarker Research, KU Leuven, 3000, Leuven, Belgium

[82]Translational Neuroimaging Laboratory, McGill Centre for Studies in Aging, McGill University, H3A 2B4, Montreal, Québec, Canada

[83]Alzheimer Disease Research Unit, McGill Centre for Studies in Aging, Department of Neurology & Neurosurgery, McGill University, H4H 1R2, Montreal, Québec, Canada

[84]Inria, Aramis project-team, 75013, Paris, France

[85]Centre pour l'Acquisition et le Traitement des Images, Institut du Cerveau et la Moelle, 75013, Paris, France

[86]Instituto Ciencias Nucleares Aplicadas a Saude, Universidade de Coimbra, 3000-548, Coimbra, Portugal

1117

Figures and tables

Tables

Table 1. Baseline demographic characteristics of the GENFI cohort. Abbreviations: MMSE, mini mental state examination; FTLD-CDR-SOB, frontotemporal lobar degeneration national Alzheimer's disease coordinating centre + clinical dementia rating sum of boxes; NfL, neurofilament light.

Characteristic	Overall, N = 238[1]	Non-carrier, N = 76[1]	Presymptomatic *C9orf72*, N = 44[1]	Presymptomatic *GRN*, N = 38[1]	Presymptomatic *MAPT*, N = 25[1]	Symptomatic *C9orf72*, N = 27[1]	Symptomatic *GRN*, N = 17[1]	Symptomatic *MAPT*, N = 11[1]	*P*-value[2]
Age, years	48 (38, 58)	43 (38, 53)	43 (33, 50)	50 (37, 56)	42 (33, 46)	58 (55, 70)	64 (58, 67)	63 (59, 66)	<0.001
Sex, male	108 (45%)	33 (43%)	19 (43%)	18 (47%)	9 (36%)	16 (59%)	8 (47%)	5 (45%)	0.8
Education, years	15 (12, 16)	15 (12, 17)	14 (12, 16)	15 (13, 16)	15 (13, 16)	13 (11, 14)	14 (9, 15)	13 (12, 16)	0.019
Plasma NfL, pg/mL	8 (6, 15)	7 (5, 10)	8 (6, 10)	8 (5, 10)	6 (5, 9)	40 (21, 55)	44 (37, 69)	20 (18, 23)	<0.001
MMSE	30.0 (28.0, 30.0)	30.0 (29.0, 30.0)	30.0 (29.0, 30.0)	30.0 (29.0, 30.0)	30.0 (29.0, 30.0)	26.0 (20.3, 28.8)	23.0 (20.5, 28.0)	24.5 (17.8, 27.0)	<0.001
CDR FTLD SOB	0.0 (0.0, 1.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.5)	0.0 (0.0, 0.0)	0.0 (0.0, 0.5)	11.5 (4.8, 15.5)	10.0 (4.8, 13.0)	7.5 (3.3, 10.6)	<0.001

[1] Median (IQR); n (%)
[2] Kruskal-Wallis rank sum test; Fisher's exact test

Figure Legends

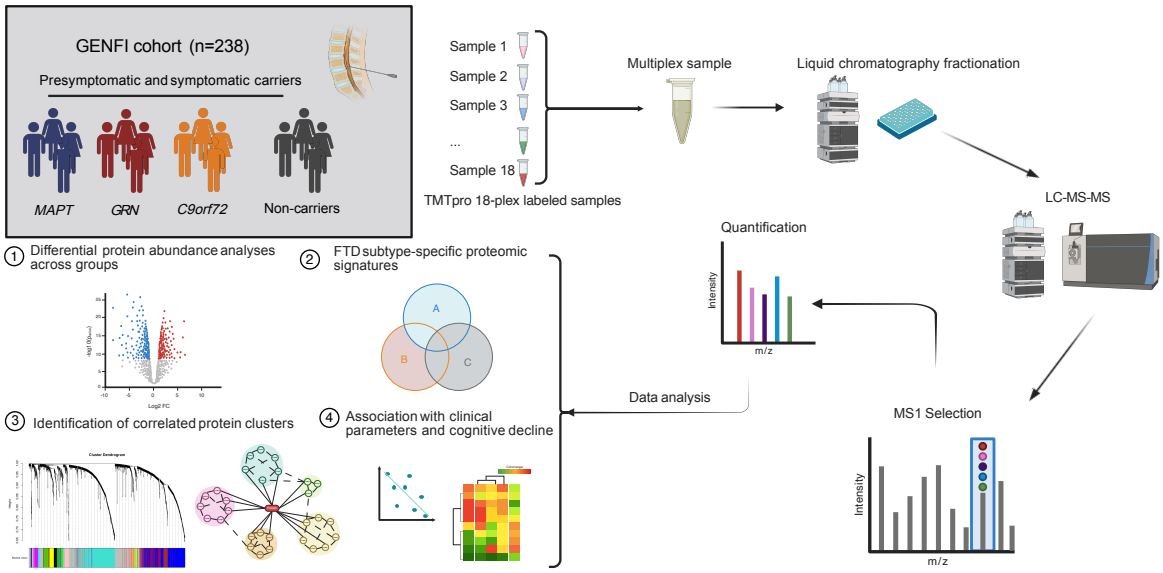


Fig. 1. Key information about participants, proteomics workflow and data analysis.

This figure shows the TMT tandem mass spectrometry (MS/MS) proteomics workflow, in which samples were pre-processed and labelled with 18 different isobaric TMTpro tags (TMTpro 18-plex) and combined into multiplex samples to allow for relative quantification and simultaneous analysis of the 18 individual samples. This process was then repeated until all 238 samples were labelled with isobaric tags. Next, each multiplex sample was fractionated using offline high-pH liquid chromatography (HP-LC) to reduce sample complexity, and each fraction was subsequently analyzed by LC-MS/MS. The data analysis conceptually consisted of four steps: 1) Investigating differences in protein abundances in mutation carriers compared with non-carriers and 2) determining FTD-subtype specific proteomic signatures employing linear models, 3) protein network analysis to investigate mutation and pathology-specific pathophysiological features as well as finally 4) correlating these protein clusters with clinical parameters and cognitive decline to discern clinically relevant changes. Abbreviations: TMT, tandem mass tag; FTD, frontotemporal dementia; MS, mass spectrometry; *MAPT*, microtubule associated protein tau; *GRN*, progranulin; *C9orf72*, chromosome 9 open reading frame 72; GENFI, GENetic Frontotemporal dementia Initiative.

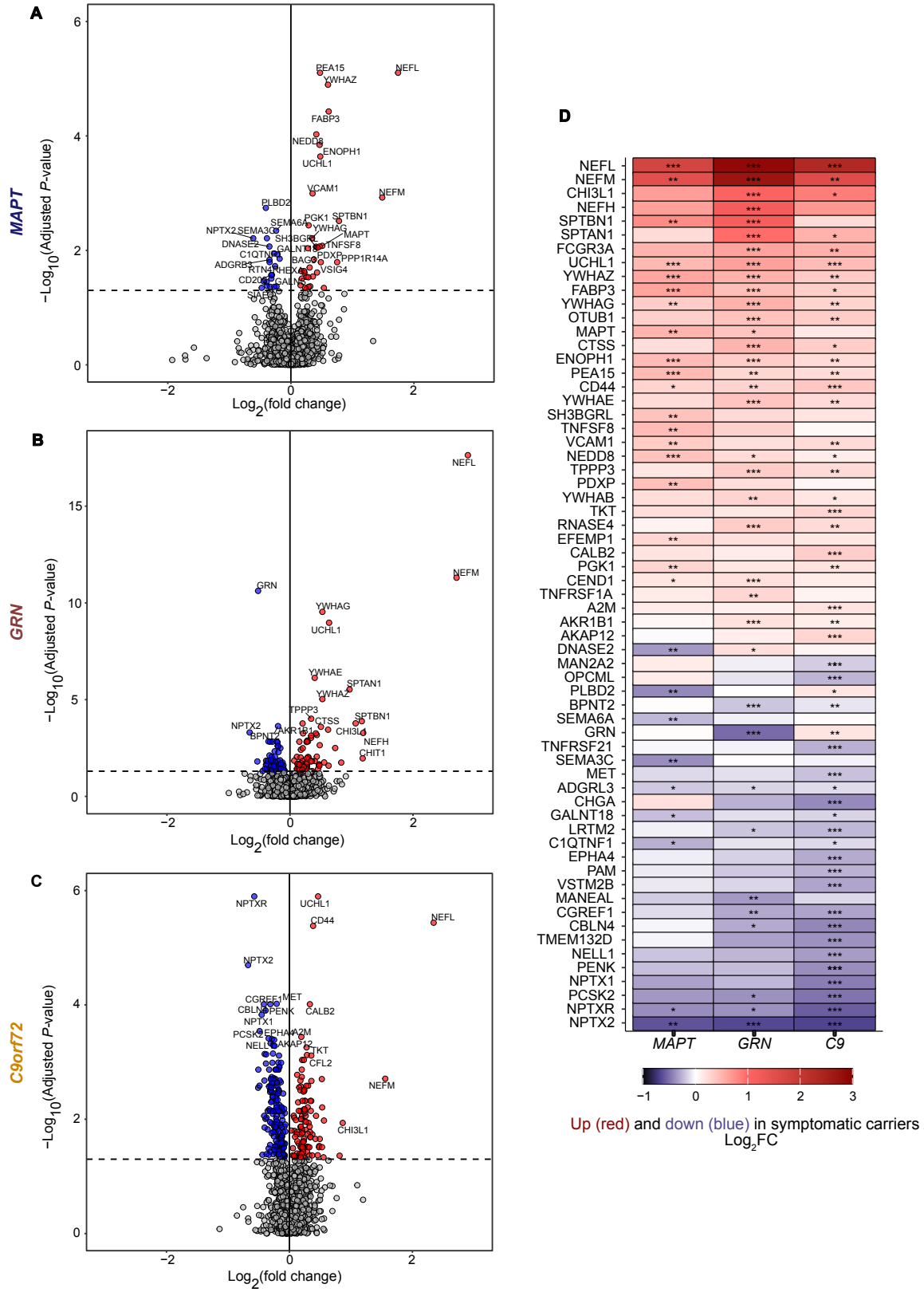
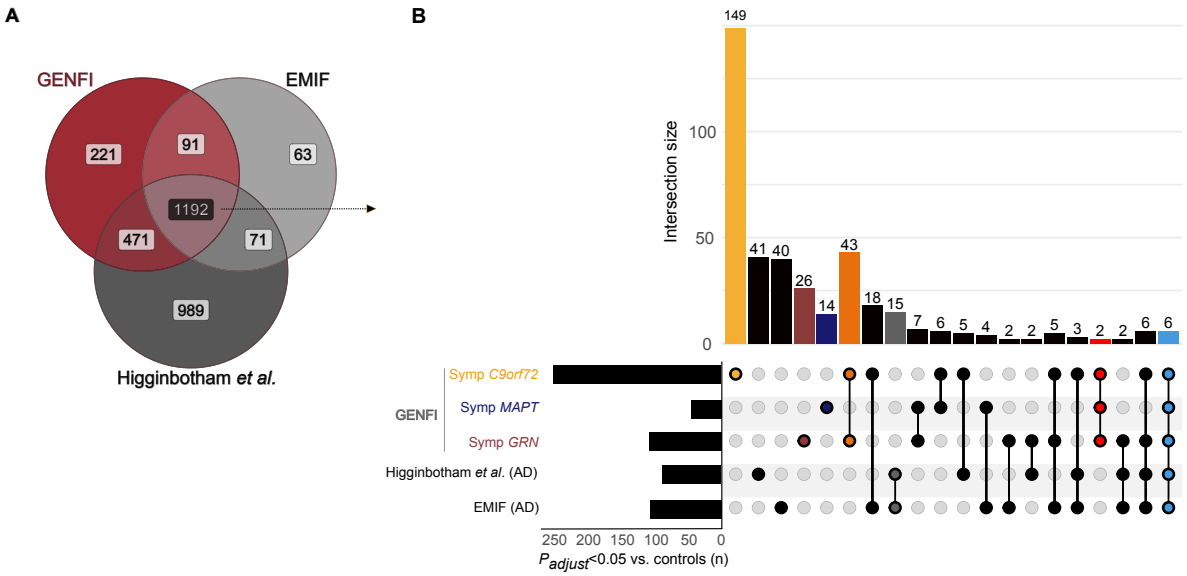


Fig. 2. Volcano plots and heat map displaying top protein hits in symptomatic mutation carriers vs non-carriers.

(A – C) Volcano plots showing proteomic differences in symptomatic *MAPT* (A), *GRN* (B) and *C9orf72* (C) mutation carriers based on linear regression analysis with age and sex as covariates. Differences were considered significant if Benjamini-Hochberg (false discovery rate [FDR]) adjusted *P*-values were <0.05. (D) The heatmap displays the 25 proteins in each group that had the lowest FDR-adjusted *P*-values in linear regression analysis, resulting in 62 proteins when accounting for overlapping proteins among groups. The $\log_2$ fold abundance change between non-carriers and the respective mutation carrier group is colour-coded; proteins higher or lower in abundance in symptomatic mutation carriers vs. non-carriers are shown in red and blue, respectively. Note that not all proteins listed in D were significantly altered in all groups. * P_{adjust}<0.05, ** P_{adjust}<0.01, *** P_{adjust}<0.001. Details on exact *P*-values and $\log_2$ fold change can be found in Table S1A-S1C.



C

Symptomatic *C9orf72*	Symptomatic *MAPT*	Symptomatic *GRN*	All GENFI and both AD
SORT1, CALB2, ROBO1 (...)	HEXA, SEMA6A, CTSD (...)	TMEM132A, RNF13,CHI3L2 (...)	YWHAZ, YWHAG, UCHL1, NPTXR, NPTX2, FABP3
TDP-43 (*GRN*/*C9orf72*)	AD	All GENFI	
GRN, CTSS, LAMP1 (...)	NRGN, SMOC1, STMN1 (...)	FSTL4, CD44	

Fig. 3. Cross cohort comparisons of symptomatic genetic FTD with AD.

(A) Venn diagram with proteins measured in GENFI, the EMIF cohort *(25)* and Higginbotham *et al. (28)*. The overlap (n=1192) represents proteins quantified in all studies. (B) Upset plot of differentially expressed proteins (FDR-adjusted P<0.05) for symptomatic *C9orf72*, *MAPT*, *GRN* mutation carriers and for patients with AD from the Higginbotham and EMIF cohort. The upper, vertical bars show the number of differentially expressed proteins exclusive to one patient group or shared between groups. The left horizontal bars represent the total number of proteins with P_{adjust}<0.05 comparing each group with control individuals. Intersections of clinical interest are color-coded. Intersections only containing 1 protein are not displayed in the figure. (C) Selection of proteins in intersections from the upset plot in panel B that are of clinical interest, as well as proteins specifically altered in one group. Proteins included in each of these intersections, as well as those not displayed, can be found in Table S5. Abbreviations: GENFI, GENetic Frontotemporal dementia Initiative; EMIF, European Medical Information Framework; AD, Alzheimer's disease.

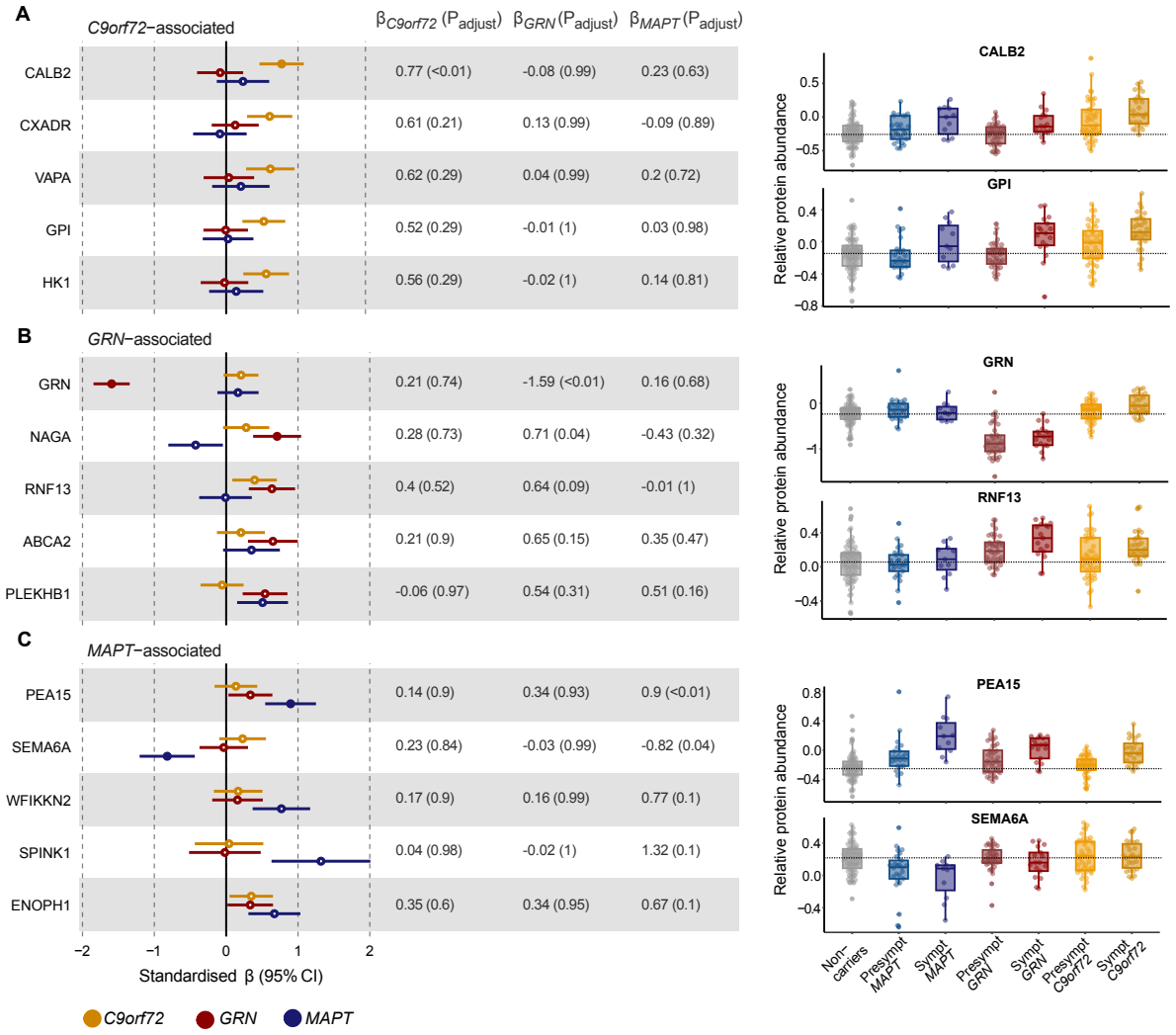
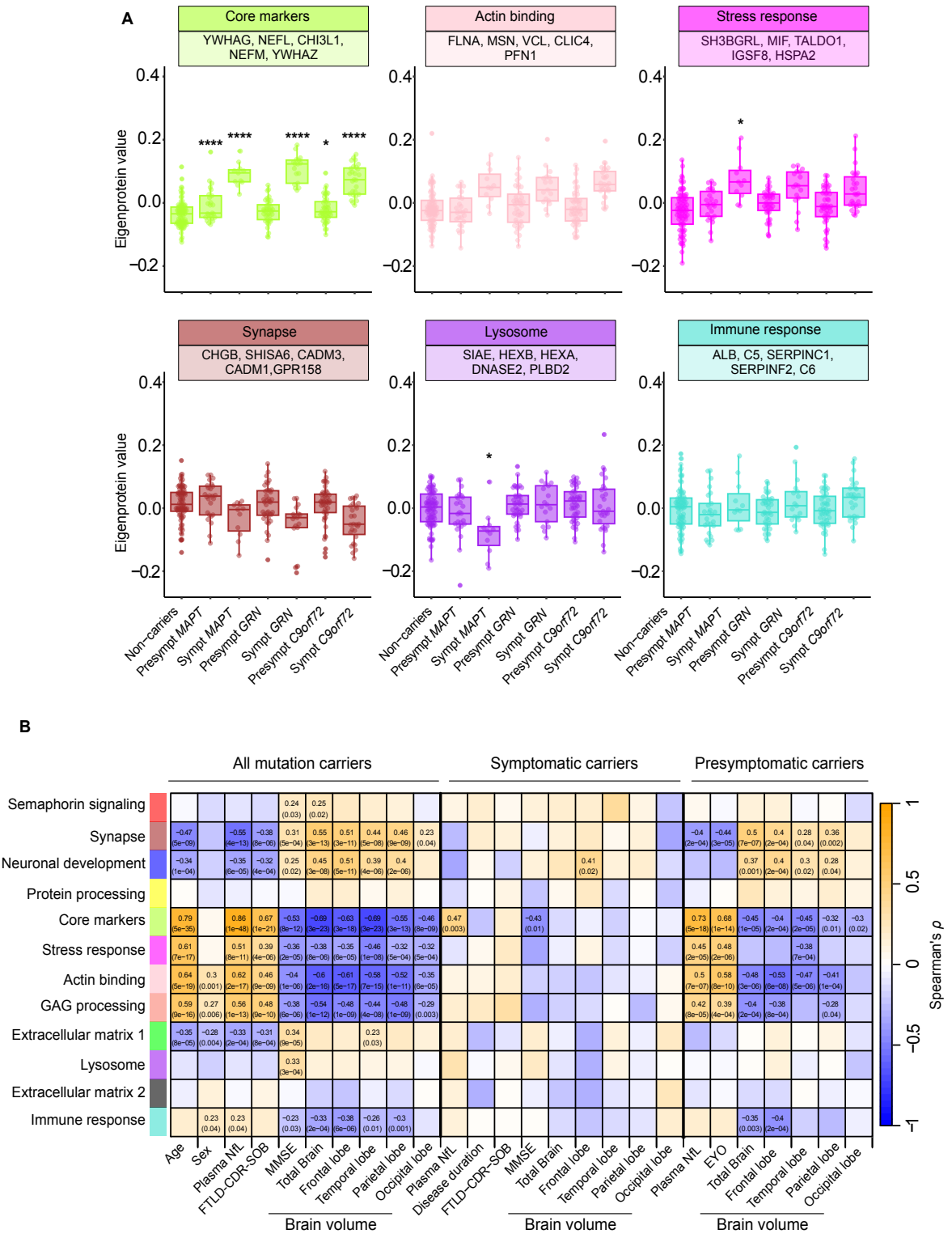


Fig. 4. Identification of mutation-associated proteins. (A to C) Left: Forest plots of the top five proteins most strongly associated with *C9orf72* (A), *GRN* (B), and *MAPT* (C) mutation. For the identification of mutation-associated proteins, linear models were fitted testing the effect of mutation carrier group on protein abundance, including affectation (presence/absence of symptoms) as well as age and sex as covariates. Non-carriers served as reference group. Coefficients with an adjusted $P<0.05$ are depicted as coloured points and 95% CI were added. Standardised β estimates including corresponding Benjamini-Hochberg-adjusted P-values for each association and mutation group were reported. Right: Boxplots of two manually selected proteins across the entire cohort. The dotted line denotes the median $\log_2$-transformed protein abundance value of the non-carrier group. Abbreviations: CI, confidence interval.



1206

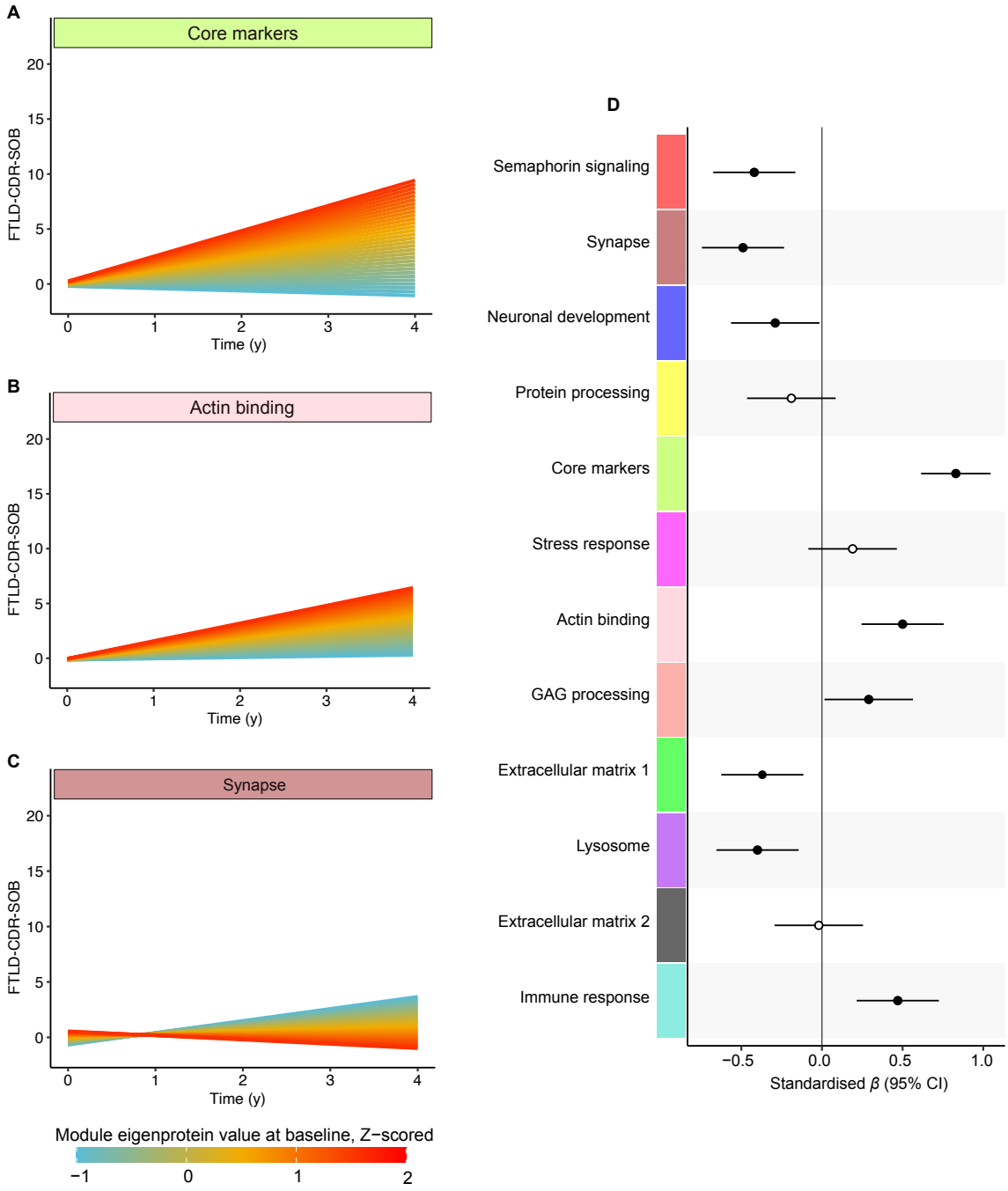
1207 **Fig.5. Weighted gene co-expression network modules show mutation/pathology-specific changes**

1208 **and correlate with relevant clinical parameters.**

1209 **(A)** Employing Weighted gene co-expression network analysis (WGCNA), we identified 14 distinct

1210 highly correlated modules of proteins. For this figure, six modules of particular interest were selected

and their Eigenprotein values were plotted across the entire cohort: 'Core markers', 'Actin binding', 'Stress response', 'Synapse', 'Lysosome', 'Immune response'. Modules were named in accordance with gene ontology (GO) terms mapped to their constituent proteins. Framed boxes contain the names of the top five hub proteins of each module, as determined by having the highest module membership value (kME). *P*-values for respective group comparisons vs. non-carriers are derived from linear regression analyses with post hoc Tukey's honestly significant difference (HSD) to adjust for multiplicity. Boxplots of the remaining modules can be found in Fig. S24. * $P<0.05$, ** $P<0.01$, *** $P<0.001$, **** $P<0.0001$. **(B)** Heatmap of correlation parameters of module Eigenproteins with different clinical measures. Spearman's rho values are colour-coded, and the corresponding Bonferroni-corrected *P*-values are included in parentheses for each tile. To evaluate the association of protein modules with clinical parameters at different time points of the disease continuum, correlations were performed in an indicated subset of individuals only. Abbreviations: FTLD-CDR-SOB, frontotemporal lobar degeneration clinical dementia rating sum of boxes; EYO, estimated years to disease onset; GAG, Glycosaminoglycan; MMSE, Mini Mental State Examination.



1239

1240 **Fig 6. Weighted gene co-expression network modules predict cognitive change in genetic FTD.**
1241 The plots show estimates of the fixed effect (standardised module Eigenprotein value*time in years
1242 from baseline) of linear mixed effects models with FTLD-CDR-SOB as dependent variable in mutation
1243 carriers (n=146). The models included standardised Eigenprotein values*time, age, sex, years of
1244 education and affectation at baseline (presymptomatic/symptomatic) as independent variables. Panels
1245 (A-C) Estimates of ME values for 'Core biomarkers' (A), 'Actin binding' (B), and 'Synapse' (C)

modules from separate models. The colours visualise the estimated cognitive trajectories at Z-scored baseline Eigenprotein values ranging from -1 to 2 SD from the mean. (D) Forest plot of the standardised β estimates and 95% CI of each module Eigenprotein value*time. The filled points denote statistically significant (P<0.05) interaction terms.

Supplementary Materials for

Proteomic analysis reveals distinct cerebrospinal fluid signatures across genetic frontotemporal dementia subtypes

Sogorb-Esteve A, Weiner S and Simrén J *et al.*

The PDF file includes:

Supplementary methods

Supplementary Figures S1-S25

References 53-56

Other Supplementary Material for this manuscript includes the following:

Supplementary data file – Tables S1-S10

Reproducibility checklist

Supplementary Methods

Fluid, imaging and cognitive biomarker collection and processing.

Participants also had plasma samples collected as part of the GENFI protocol. Plasma was collected, processed, and stored in aliquots at -80°C according to standardised procedures. Plasma neurofilament light chain (NfL) levels were correlated (Spearman's correlation) with TMT relative reporter ion intensities of CSF NfL (n=163). Plasma NfL concentration was measured with Single molecule array (Simoa) technology using the Neurology 4-Plex A kit (Quanterix, Billerica, USA) on an HD-X Analyzer following the manufacturer's instructions (Quanterix, Billerica, USA). Measurements were completed in duplicate (all CVs below 15%) over a total of 3 batches, each with an 8-point calibration curve tested in triplicate and 2 controls tested in duplicate, as reported before *(7)*.

In addition, TMT tryptic peptide measurements of the synaptic proteins 14-3-3 epsilon ([R].IISSIEQK.[E], n=119), neuronal pentraxin 2 ([K].VAELEDEK.[S], n=181) and neuronal pentraxin receptor ([R].NNYMYAR.[V], n=169) were correlated with corresponding tryptic peptide measurements as performed in the publication Sogorb-Esteve et al, 2020 *(26)*. In brief, to 100 µL of CSF, a mixture of stable-isotope-labeled peptides (internal standard) was added (25 µL, 0.032 pmol/µL, JPT Peptide Technologies, Berlin, Germany; SpikeTides L). This was then followed by a stepwise protocol of reduction, alkylation, and tryptic digestion, and lastly solid-phase extraction for purification purposes. LC-MS/MS analysis was performed using a microflow HPLC, equipped with a Hypersil Gold reversed-phase column (100 × 2.1 mm, particle size 1.9 µm, Thermo Fisher Scientific), and a Triple Quadrupole mass spectrometer (6495 Triple Quadrupole LC/MS system, Agilent Technologies). To monitor the performance of the assay over time, quality control (QC) sample replicates were injected at regular intervals during runs. The panel of synaptic markers included: AP-2 complex subunit beta, complexin-2, beta-synuclein, gamma-

synuclein, 14-3-3 proteins (eta, epsilon, zeta/delta), neurogranin, Rab GDP dissociation inhibitor alpha (Rab GDI alpha), syntaxin-1B, syntaxin-7, phosphatidylethanolamine-binding protein 1 (PEBP-1), neuronal pentraxin receptor (NPTXR), neuronal pentraxin 1 (NPTX1), and neuronal pentraxin 2 (NPTX2).

Volumetric T1-weighted MRI scans were bias field corrected and parcellated using the geodesic information flow algorithm *(53)*. From this parcellation, the volumes of the bilateral frontal, temporal, parietal ad occipital cortices and of the whole brain were extracted and expressed as a percentage of the total intracranial volume, which was computed with SMP12 v6470 (Statistical Parametric Mapping, Wellcome Trust Centre for Neuroimaging, London, UK) running under Matlab R201b (Math Works, Natick, MA, USA) *(54)*.

The standardized GENFI clinical assessment included a history, examination, cognitive assessment (including Mini-Mental State Examination [MMSE]), FRS, and the CDR plus NACC FTLD rating scale. Mutation carriers were classified into asymptomatic, prodromal, or symptomatic if they scored 0, 0.5, or ≥ 1, respectively, on the CDR plus NACC FTLD global score. As part of the GENFI clinical assessment, the CDR plus NACC FTLD was administered as per standard protocol (interviewing both the participant and an informant separately) including the core cognitive and functional domain items from the CDR (memory, orientation, judgment and problem solving, community affairs, hobbies, personal care), and the two-clinician judgment (global) scores from the NACC FTLD for behavior and language.

Data processing and normalisation

All RAW files were processed using Proteome Discoverer Version 2.5.0.400 (Thermo Scientific). The most confident centroid integration method with an integration tolerance of 20 ppm was employed to perform peak integration for reporter ion quantification. Peptides

were identified searching against the UniProtKB Swiss-Prot (TaxID = 9606, Homo sapiens) database utilising the SequestHT search engine with search parameters specified as follows: precursor Δm tolerance = 5 ppm, fragment Δm tolerance = 0.02 Da, missed cleavages = 2, min. peptide length = 6, fixed modifications = carbamidomethyl, TMTpro (peptide N-terminus, K residues). Percolator was used for peptide scoring, filtering peptide spectral matches and peptides to a false discovery rate (FDR) of <1%. Peptides were then assembled into proteins based on their uniqueness (unique peptides). In the event of redundancy, peptides were assigned to a protein sequence in accordance with the principle of parsimony (razor peptides).

For data normalisation, individual protein abundances were divided by their corresponding set-wise global internal standard (GIS) protein measurement. Each obtained protein ratio was then additionally divided by the respective sample median, accounting for aberrant differences in total protein amount. Finally, all data was transformed into a $\log_2$-space. Potential batch effects and sample outliers were assessed by performing a principal component (PCA) analysis (Fig. S25) and hierarchical clustering considering all sample-wise protein abundances.

Weighted gene co-expression network analysis (WGCNA) and correlation with clinical parameters

Weighted gene co-expression networks were constructed using the R package *WGCNA (55)*. Due to the limited sample sizes of individual diagnostic groups, a network including all samples of the present cohort was built. Following the removal of proteins with missing values in >50% of all study participants, the optimal soft threshold power was chosen as the power at which scale free topology R^2 approached an asymptote at around 0.9 and the mean and median connectivity were <100. A signed network was built using the

WGCNA::blockwiseModules function with the following settings: soft threshold power = 14, deepSplit = 4, corType = bicor, minModuleSize = 10, mergeCutHeight = 0.2, pamRespectsDendro TRUE, pamStage TRUE, maxPOutliers $p < 0.05$, reassignThreshold = 0.05. In brief, a robust correlation metric insensitive to outliers (bicor) is used to compute the correlation between all pairs of proteins. Next, the resulting correlation matrix is transformed into an adjacency matrix raising the co-expression similarities to the determined soft threshold power. The adjacency matrix is then used to construct a topological overlap matrix (TOM), reflecting the relative interconnectedness of each protein. Finally, hierarchical protein clustering is performed on the corresponding topological overlap dissimilarity measure (1-TOM), resulting in module construction via dynamic tree cutting. A total of 14 modules could be identified, including a grey module (645 proteins) containing proteins that could not be assigned to any of the modules and a module containing contaminants from the laboratory environment (tan module). In a next step, module Eigenproteins corresponding to the module's first principal component were identified. Protein module membership kME was determined by performing Pearson correlation of each protein with each module Eigenprotein. Proteins with a $kME > 0.7$ were considered as the module's respective hub proteins. Module Eigenproteins of different subsets of the cohort (presymptomatic and/or symptomatic mutation carriers) were correlated (Spearman rank-order correlation) with clinical parameters. Significance levels were adjusted with Bonferroni correction to account for multiple testing. To investigate the prognostic properties of module Eigenprotein values in mutation carriers, separate linear mixed effects models for each module Eigenprotein value were fitted including only mutation carriers (both symptomatic and presymptomatic carriers) with cognitive score (FTLD-NACC+CDR-SOB) as dependent variable. Fixed effects included the interaction of module Eigenprotein value*time (years since baseline), with age, sex, years of education and cognitive status (symptomatic or presymptomatic) as covariates.

All models included random intercepts and slopes for each participant. To enable comparability between models, module Eigenprotein values were standardised. These analyses were performed using the *lme4* package in R.

Gene ontology analysis of WGCNA modules

Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis *(56)*. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher's one-tailed test, to calculate the significance of functional terms in the input protein list. Calculated *P*-values represent the probability of randomly drawing n or more proteins in a subset of proteins (WGCNA modules) annotating to a specific GO term from the total number of proteins identified in the study. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05 to apply a less stringent approach for obtaining corrected *P*-values. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms, as described in *(56)*. Terms best representing the proteins in a respective module were chosen for module annotation.

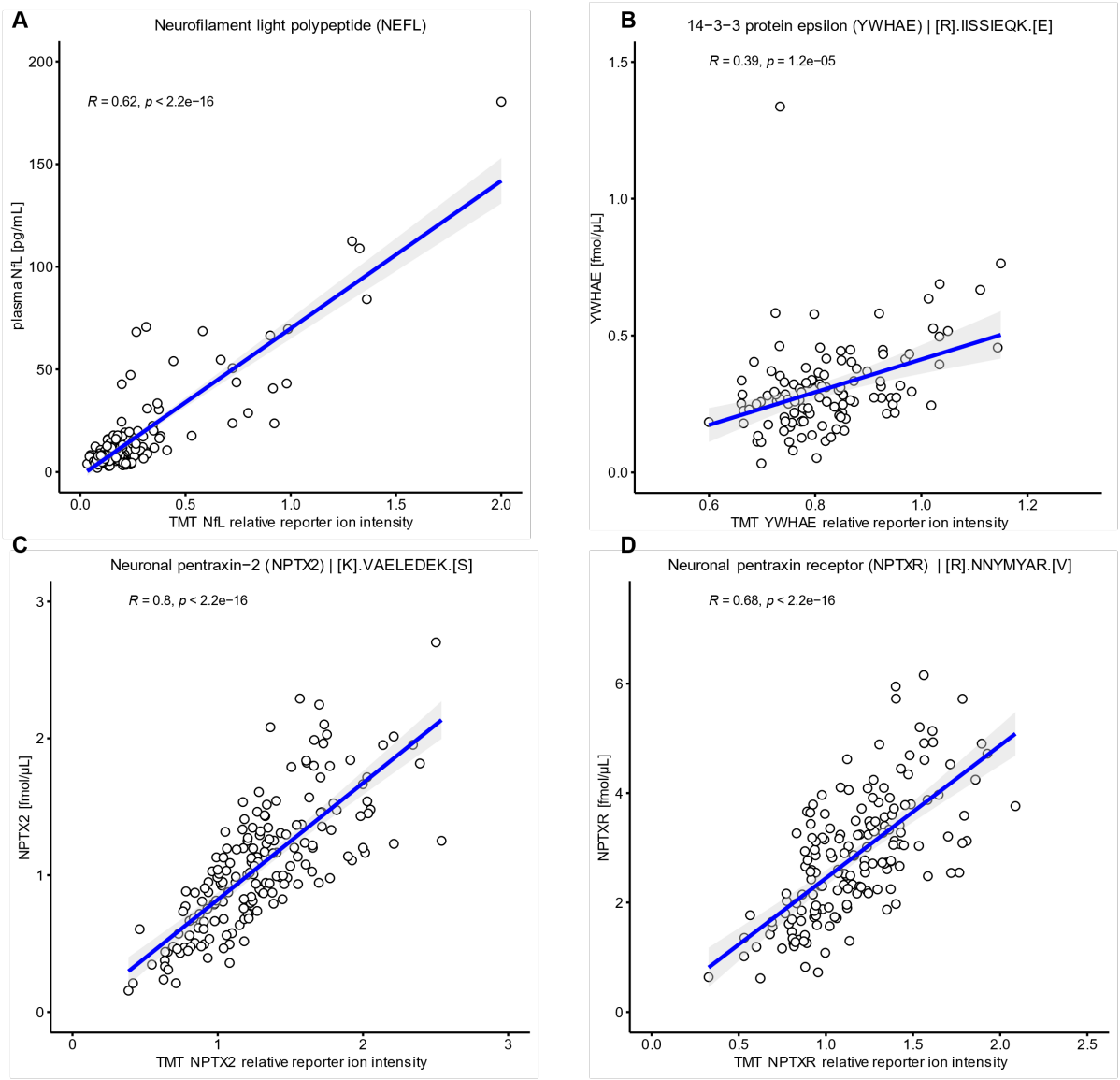


Figure S1. Correlation between biomarkers when measured with TMT-based proteome-wide quantification and targeted techniques. (A to D) Correlations between Neurofilament light (A), 14-3-3 protein epsilon (B), Neuronal pentraxin-2 (C) and Neuronal pentraxin receptor (D) abundances measured in the same samples with tandem mass tag (TMT)-based quantification on the x-axis and single molecule array (Simoa) (A) or multiple reaction monitoring (MRM)-based (B-D) quantification on the y-axis. All measurements except for Simoa Neurofilament light, which was carried out in plasma, were made in cerebrospinal fluid using mass spectrometric methods.

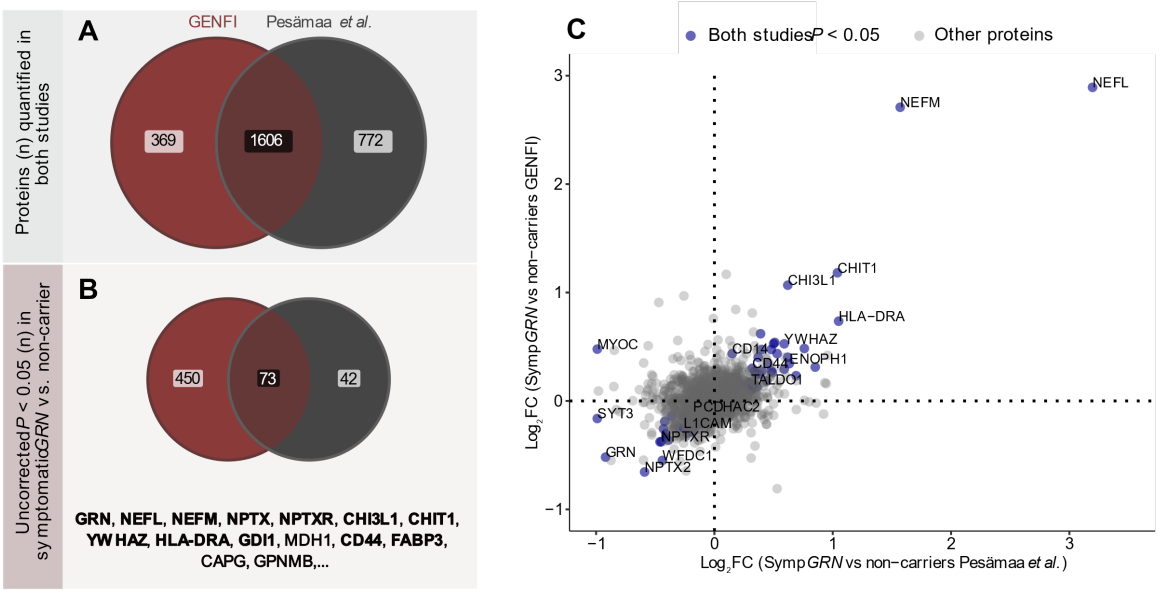


Figure S2. (A) Cross-cohort comparisons of CSF proteomic changes in symptomatic *GRN* carriers. Venn diagram showing the number of proteins quantified in the current study (red), a published label-free mass spectrometry dataset of symptomatic *GRN* carriers and non-carriers from the ALLFTD cohort (grey), with the number of proteins quantified in both studies being shown in grey/red. (B) Number of proteins significantly changed in unadjusted ANOVA analysis in the same studies. At the bottom of the panel, a selection of proteins changed in both studies are shown. Gene names highlighted in bold denote proteins with false discovery rate (FDR)-corrected *P*-values <0.05 in linear regression analyses with age and sex as covariates in the GENFI cohort. (C) Correlation between $\log_2$ fold changes between symptomatic *GRN* carriers of proteins in GENFI and ALLFTD cohorts (Spearman $R = 0.87$, $P<0.0001$). Proteins with ANOVA *P*-values <0.05 in both studies are shown in blue, whereas proteins not matching this criterion are shown in grey. Spearman correlation was performed in this subset of proteins. Abbreviations; GENetic Frontotemporal dementia Initiative, GENFI; European Medical Information Framework.

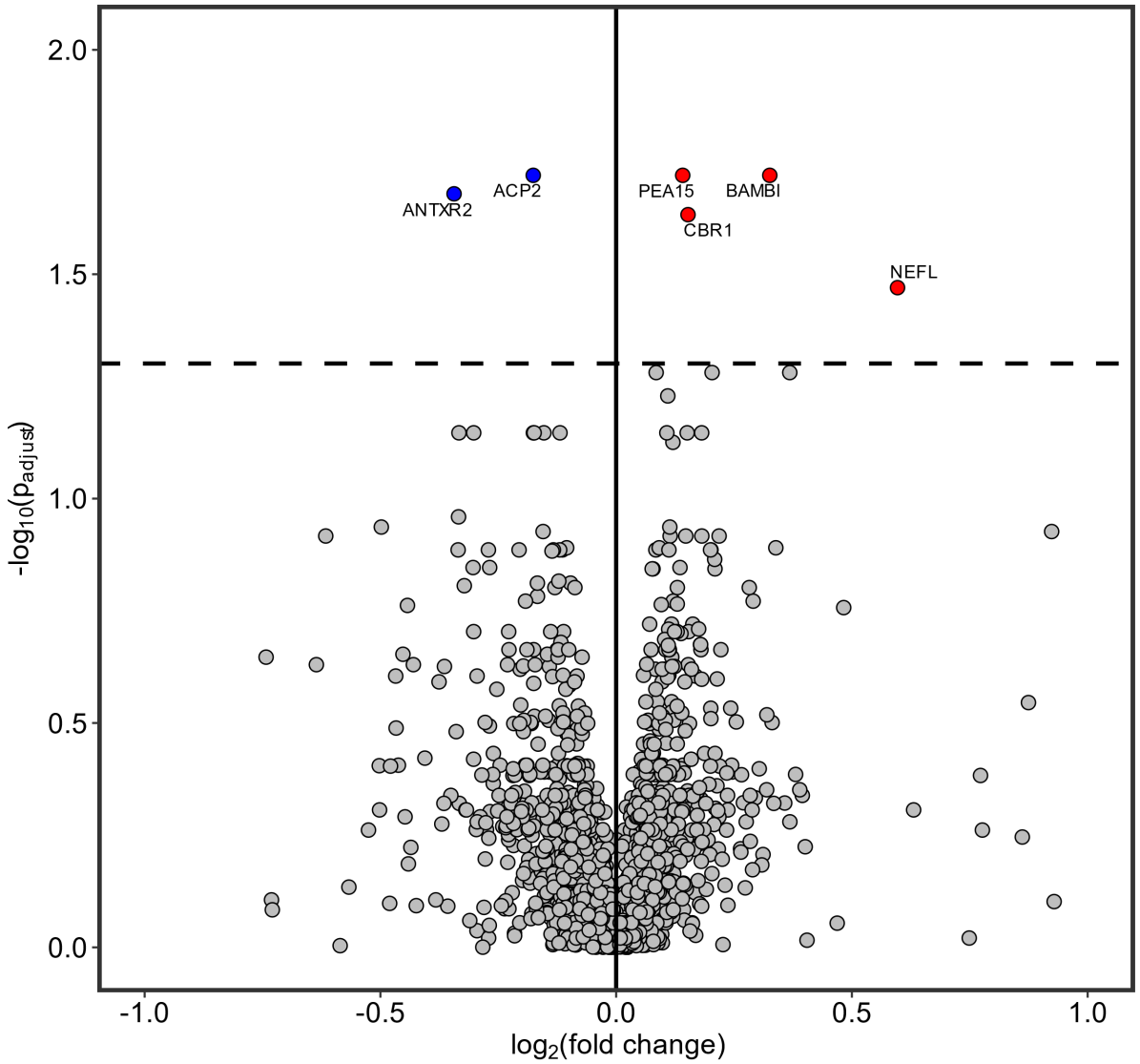


Figure S3. Volcano plot showing proteins changed in presymptomatic *MAPT* carriers vs. non-carriers. The volcano plot displays an overview of the altered proteins in presymptomatic *MAPT* mutation carriers when compared with non-carriers based on an analysis of covariance (ANCOVA) with age and sex as covariates. *P*-values were corrected for multiple testing according to the Benjamini-Hochberg method. P_{adjust} cut-off: 0.05. Abbreviations: neurofilament light, NEFL; ACP2, Acid Phosphatase 2, Lysosomal; PEA15, Astrocytic phosphoprotein PEA-15; ANTXR2, ANTXR Cell Adhesion Molecule 2; CBR1, Carbonyl reductase 1; BAMBI, BMP And Activin Membrane Bound Inhibitor.

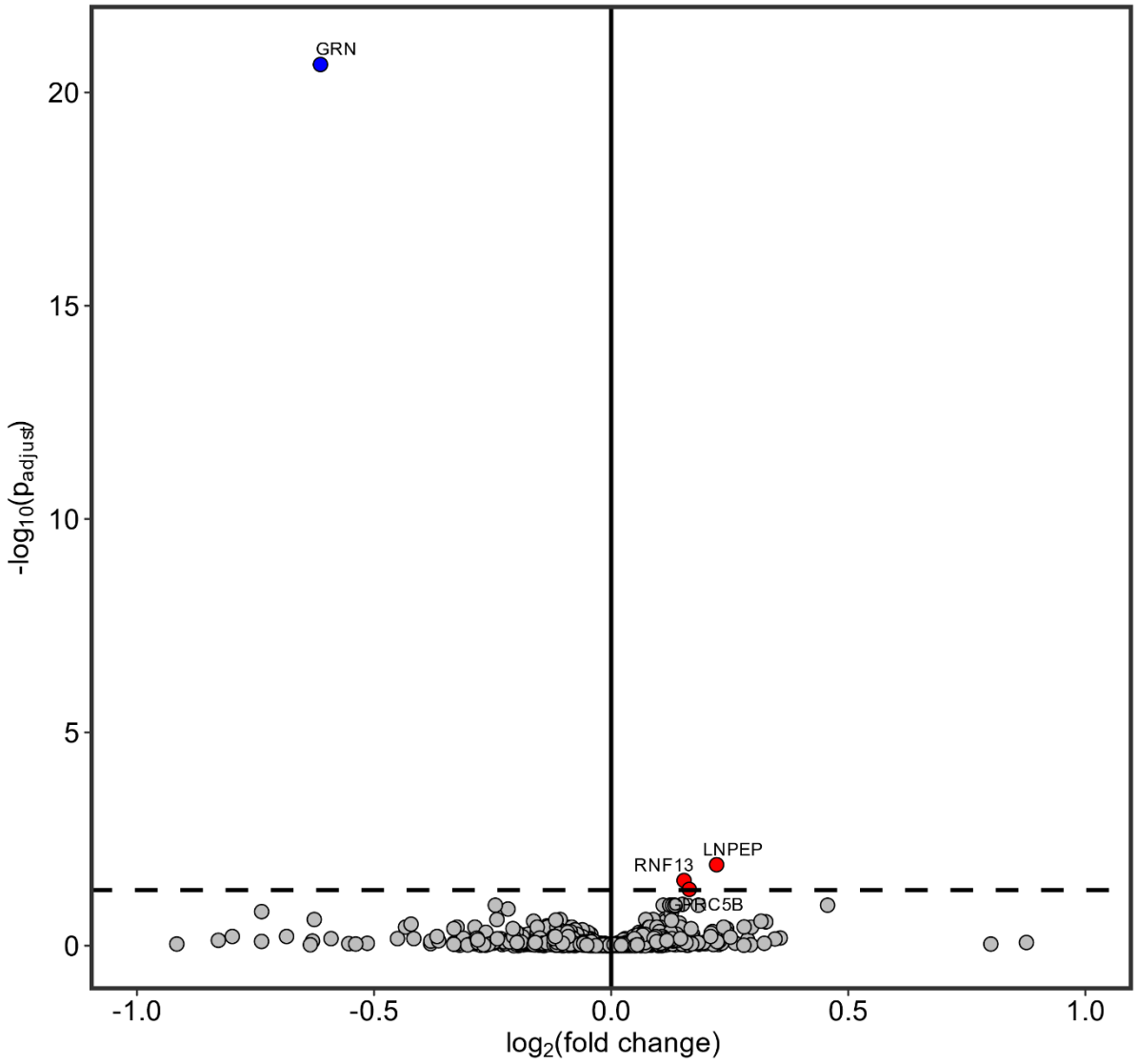


Figure S4. Volcano plot showing proteins changed in presymptomatic *GRN* carriers vs. non-carriers. The volcano plot displays an overview of the altered proteins in presymptomatic *GRN* mutation carriers when compared with non-carriers based on an analysis of covariance (ANCOVA) with age and sex as covariates. *P*-values were corrected for multiple testing according to the Benjamini-Hochberg method. P_{adjust}cut-off: 0.05. Abbreviations: Progranulin, GRN; LNPEP, Leucyl and Cystinyl Aminopeptidase; RNF13, Ring Finger Protein 13; GPRC5B, G Protein-Coupled Receptor Class C Group 5 Member B.

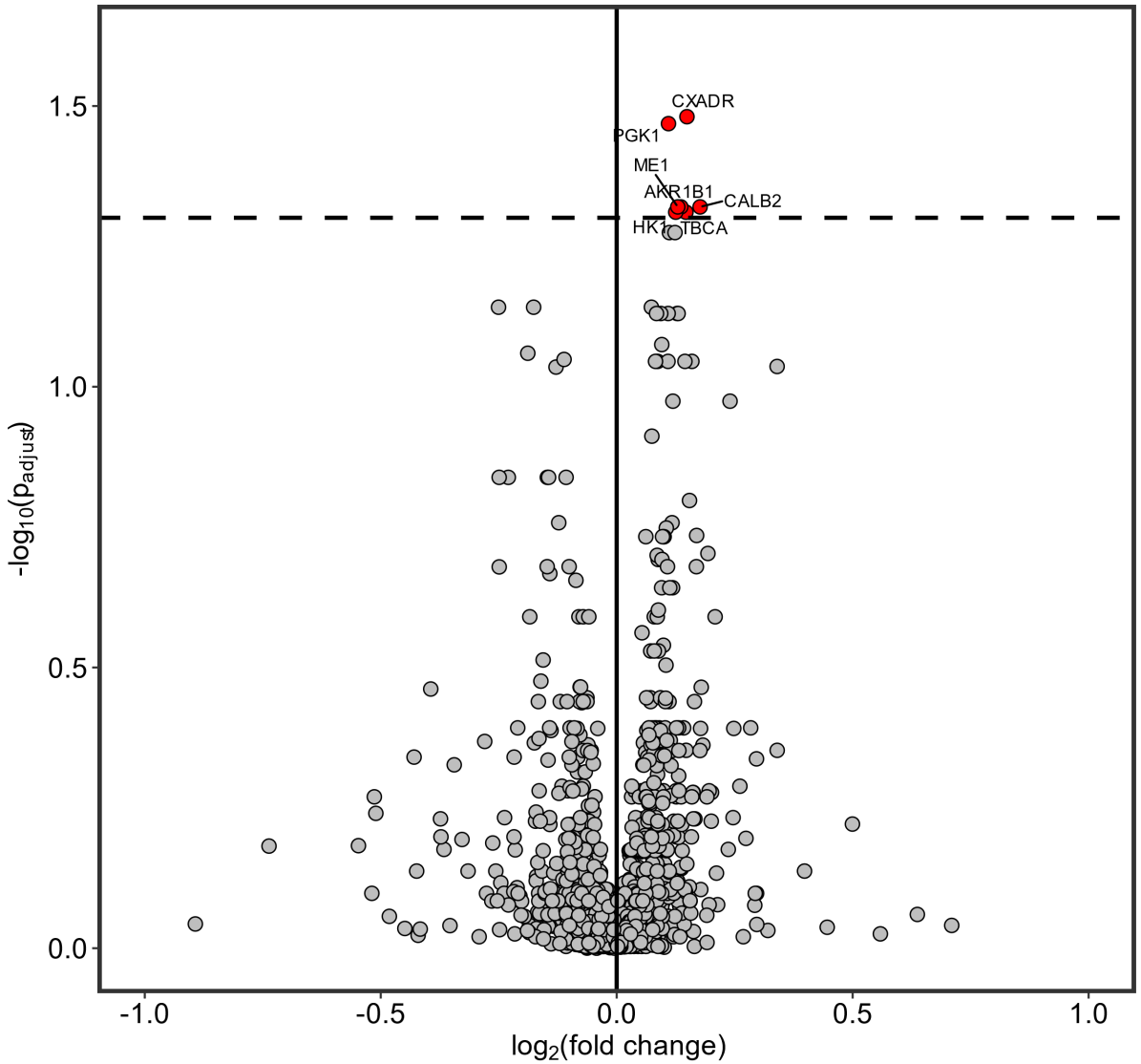


Figure S5. Volcano plot showing proteins changed in presymptomatic *C9orf72* carriers vs. non-carriers. The volcano plot displays an overview of the altered proteins in presymptomatic *C9orf72* mutation carriers when compared with non-carriers based on an analysis of covariance (ANCOVA) with age and sex as covariates *P*-values were corrected for multiple testing according to the Benjamini-Hochberg method. P_{adjust} cut-off: 0.05. Abbreviations: CXADR Ig-Like Cell Adhesion Molecule, CXADR; Phosphoglycerate Kinase 1, PGK1; Malic enzyme 1, ME1; Calretinin, CALB2; Aldo-keto reductase family 1, AKR1B1; Tubulin-specific chaperone A, TBCA; Hexokinase 1; HK1.

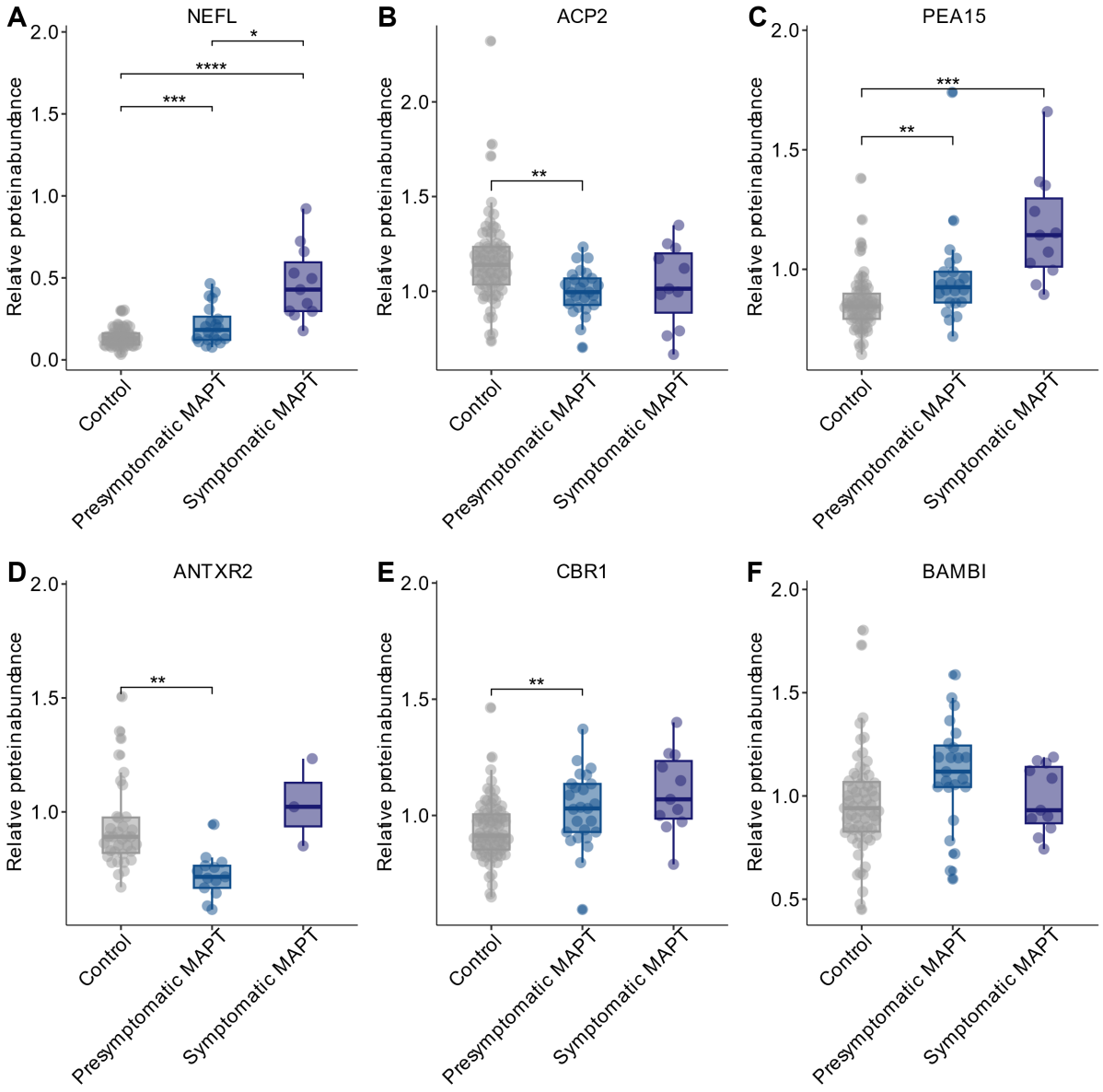


Figure S6. Proteins changed in presymptomatic *MAPT* carriers vs. non-carriers extracted from Figure S3. Overview of the altered proteins in presymptomatic *MAPT* mutation carriers when compared with controls based on an analysis of covariance (ANCOVA) with age and sex as covariates and post hoc Tukey's honest significant difference (HSD). (A) NEFL; (B) ACP2; (C) PEA15; (D) ANTXR2; (E) CBR1, and (F) BAMBI in non-carriers, presymptomatic and symptomatic *MAPT* carriers. Abbreviations: neurofilament light, NEFL; ACP2, Acid Phosphatase 2, Lysosomal; PEA15, Astrocytic phosphoprotein PEA-15; ANTXR2, ANTXR Cell Adhesion Molecule 2; CBR1, Carbonyl reductase 1; BAMBI, BMP And Activin Membrane Bound Inhibitor. * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001. Note: Non-significant comparison in tile F is due to the presence of outliers which were removed for analyses performed for Figure S3.

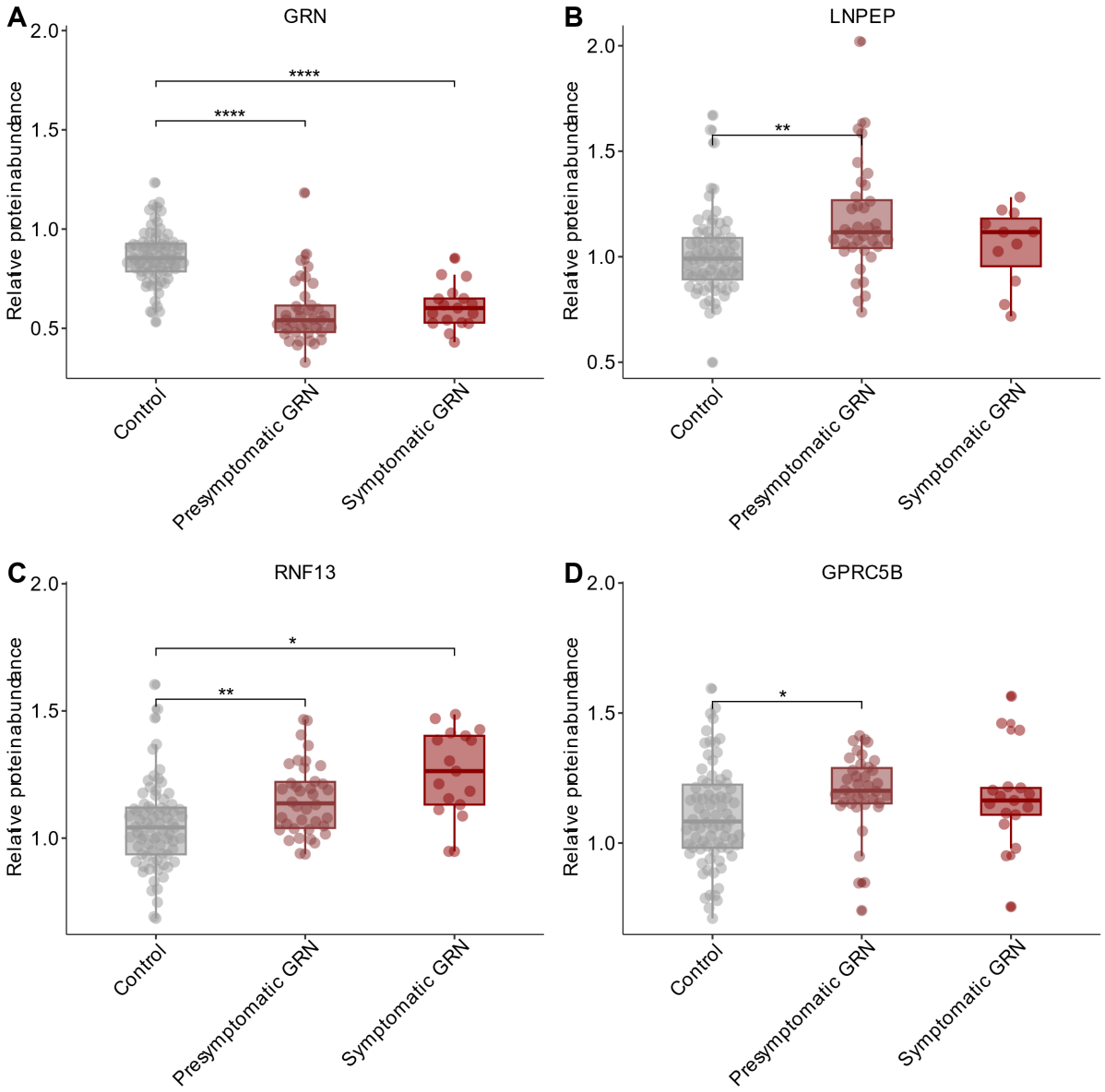


Figure S7. Proteins changed in presymptomatic *GRN* carriers vs. non-carriers extracted from Figure S4. Overview of the altered proteins in presymptomatic *GRN* mutation carriers when compared with controls based on an analysis of covariance (ANCOVA) with age and sex as covariates and post hoc Tukey's honest significant difference (HSD). Panels show (A) GRN, (B) LNPEP, (C) RNF13, and (D) GPRC5B in non-carriers, presymptomatic and symptomatic *GRN* carriers. Abbreviations: Progranulin, GRN; LNPEP, Leucyl and Cystinyl Aminopeptidase; RNF13, Ring Finger Protein 13; GPRC5B, G Protein-Coupled Receptor Class C Group 5 Member B. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$.

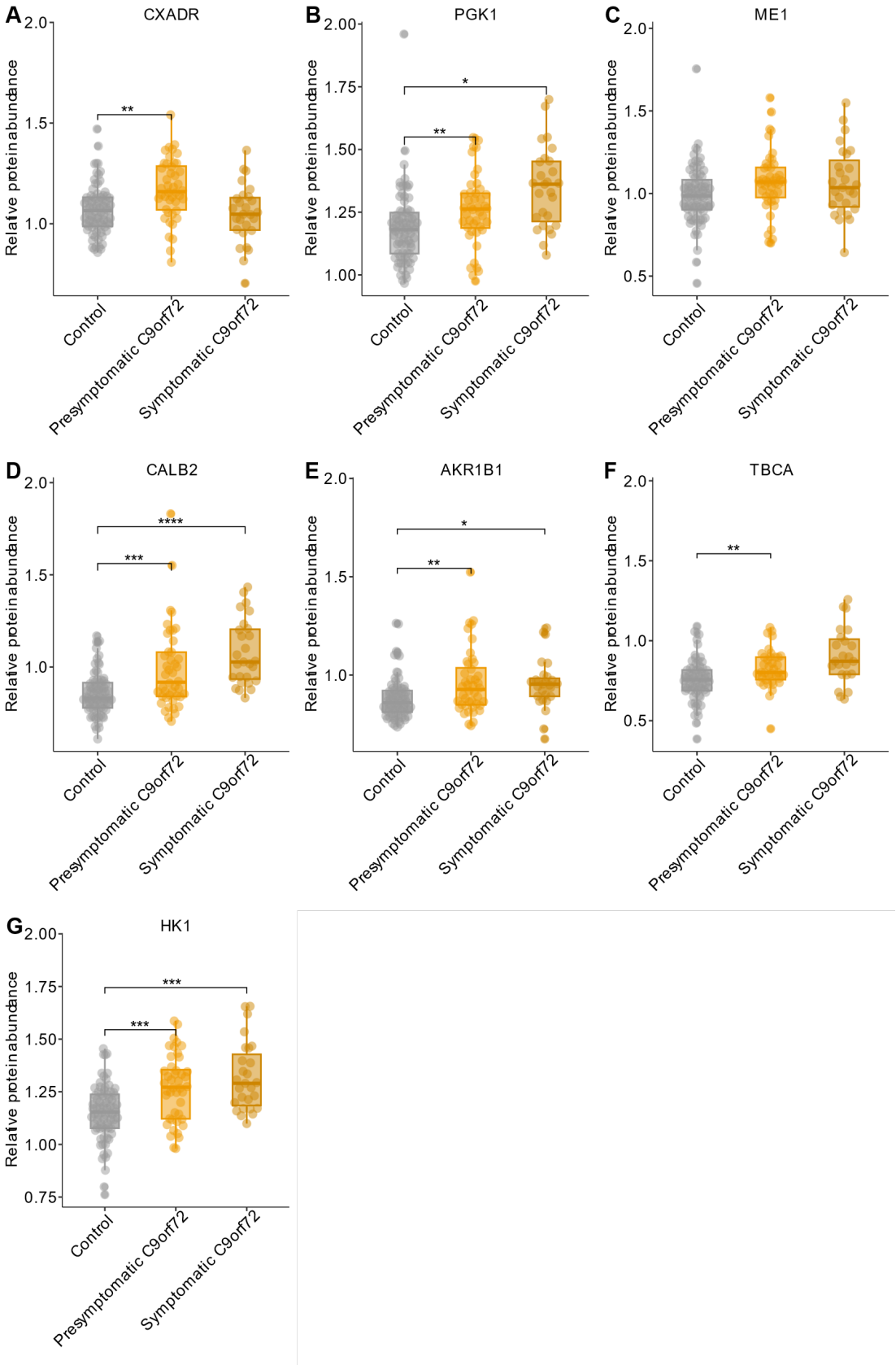


Figure S8. Proteins changed in presymptomatic *C9orf72* carriers vs. non-carriers extracted from Figure S5. Overview of the altered proteins in presymptomatic *C9orf72* mutation carriers when compared with controls based on an analysis of covariance (ANCOVA) with

age and sex as covariates and post hoc Tukey's honest significant difference (HSD). Panels show (A) CXADR; (B) PGK1, (C) ME1, (D) CALB2, (E) AKR1B1, (F) TBCA and (G) HK1 in non-carriers, presymptomatic and symptomatic *C9orf72* carriers. Abbreviations: CXADR Ig-Like Cell Adhesion Molecule, CXADR; Phosphoglycerate Kinase 1, PGK1; Malic enzyme 1, ME1; Calretinin, CALB2; Aldo-keto reductase family 1, AKR1B1; Tubulin-specific chaperone A, TBCA; Hexokinase 1; HK1. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$. Note: Non-significant comparison in tile C is due to the presence of outliers which were removed for analyses performed for Figure S5.

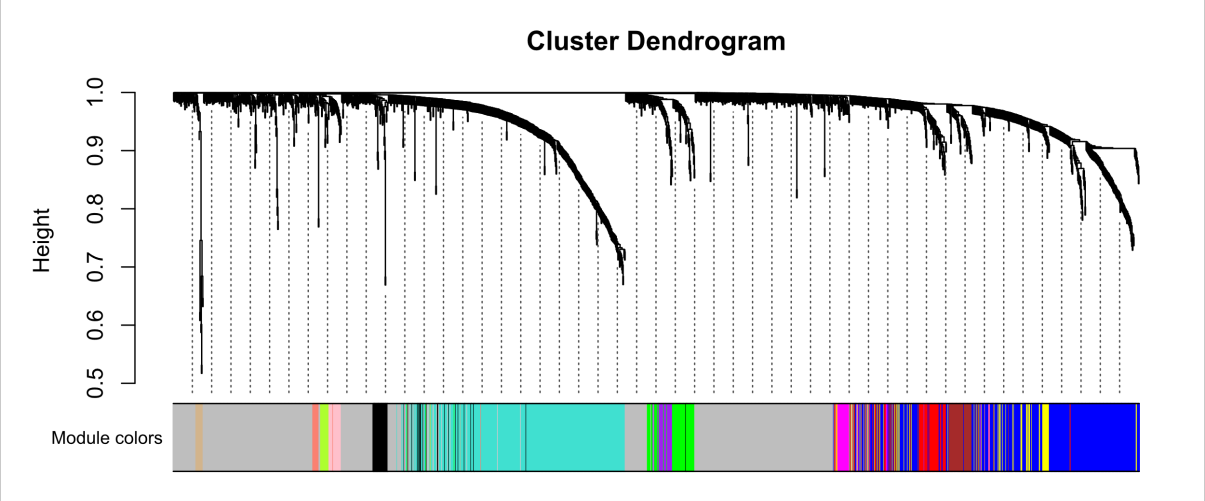
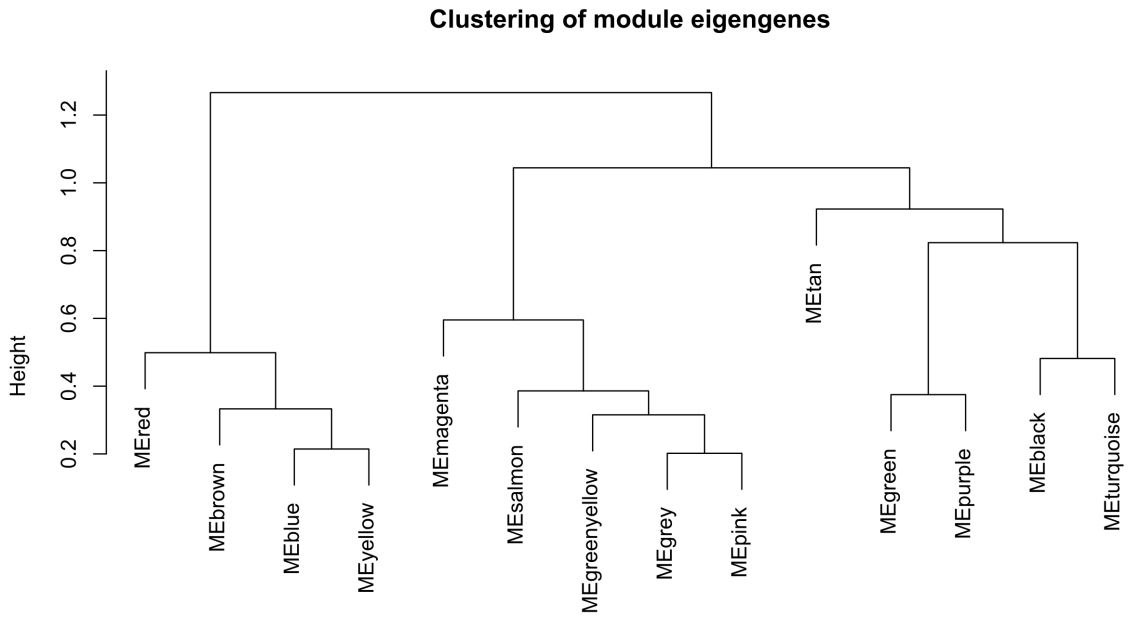


Figure S9. Cluster dendrogram of the weighted gene co-expression network analysis (WGCNA). Cluster dendrogram and color representation of the network modules produced by average linkage hierarchical clustering of proteins based on their topological overlaps.



1543

1544 **Figure S10.** Clustering of module Eigenproteins of the weighted gene co-expression network
1545 analysis (WGCNA). Hierarchical clustering of module Eigenproteins identified in the
1546 WGCNA.

1547

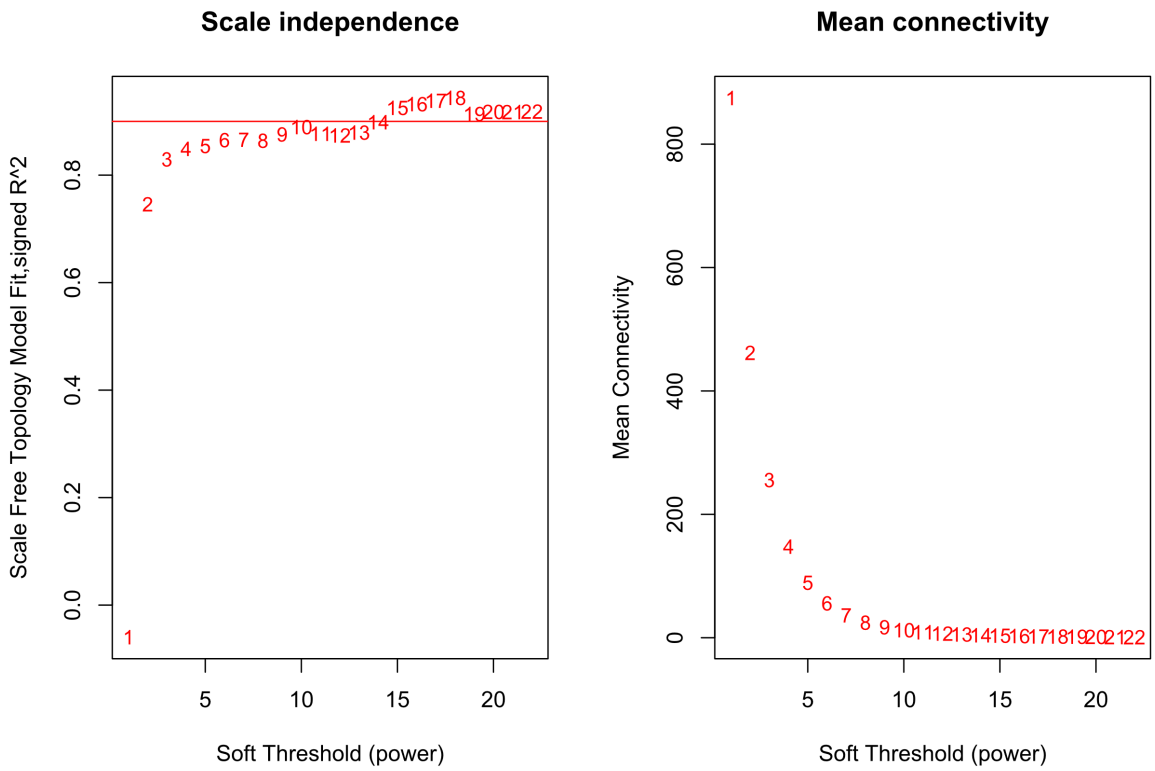
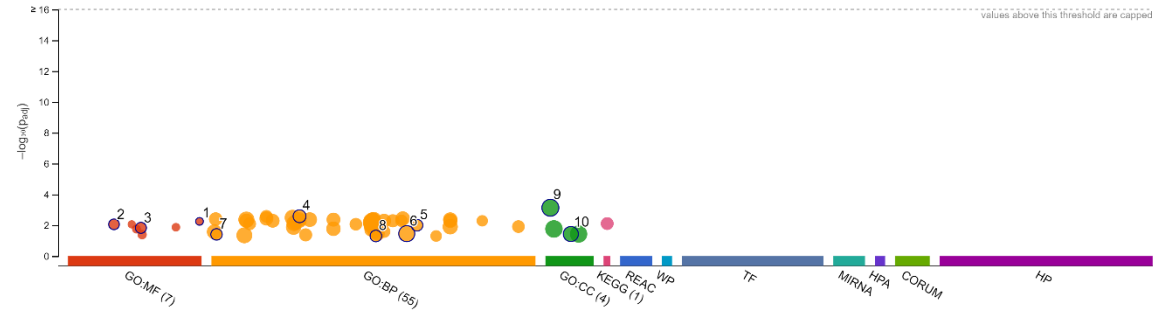


Figure S11. Determination of soft-threshold power in weighted gene co-expression network analysis (WGCNA). This figure shows the scale-free topology index (A) and mean connectivity (B) for each power value between 1 and 22. The R^2 cut-off was drawn at 0.875.

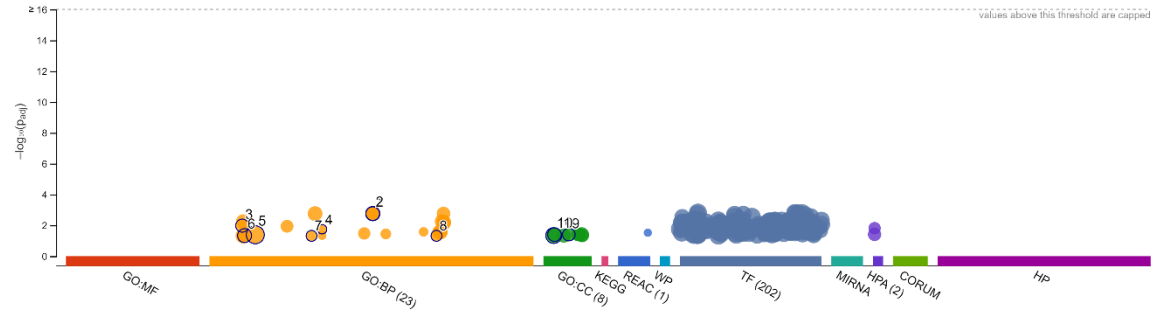


ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:1905573	ganglioside GM1 binding	5.635×10^{-3}
2	GO:MF	GO:0030215	semaphorin receptor binding	8.772×10^{-3}
3	GO:MF	GO:0045499	chemorepellent activity	1.481×10^{-2}
4	GO:BP	GO:0031345	negative regulation of cell projection organization	2.606×10^{-3}
5	GO:BP	GO:0071526	semaphorin-plexin signaling pathway	1.002×10^{-2}
6	GO:BP	GO:0065008	regulation of biological quality	3.529×10^{-2}
7	GO:BP	GO:0001755	neural crest cell migration	3.932×10^{-2}
8	GO:BP	GO:0050919	negative chemotaxis	4.992×10^{-2}
9	GO:CC	GO:0005886	plasma membrane	7.371×10^{-1}
10	GO:CC	GO:0045202	synapse	3.724×10^{-2}

version	e111_eg58_p18_30541362
date	24/05/2024, 14:39:34
organism	hsapiens

g:Profiler

Figure S12. Gene ontology terms associated with the ‘Semaphorin signaling’ (red) module. Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher’s one-tailed test, to calculate the significance of functional terms in the input protein list. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05. The y-axis of the plot displays the negative decadic logarithm of the adjusted *p*-value while the x-axis highlights the category of all corresponding GO terms. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms (displayed in the list below the plot).

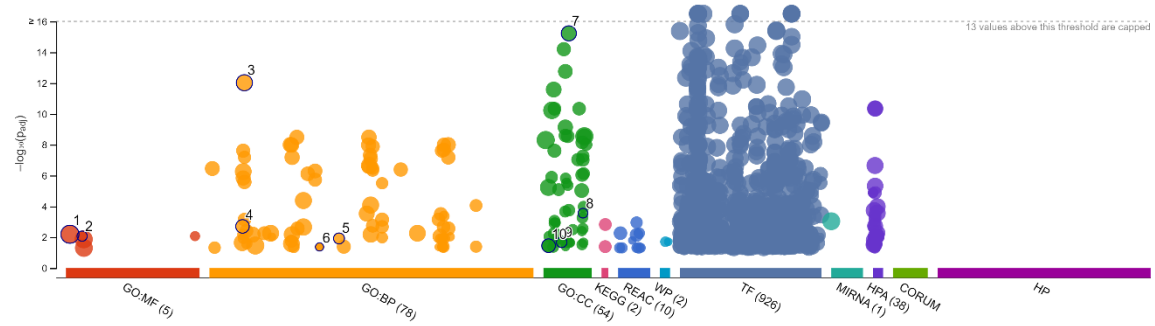


ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:BP	GO:0050804	modulation of chemical synaptic transmission	1.733×10^{-3}
2	GO:BP	GO:0050808	synapse organization	1.733×10^{-3}
3	GO:BP	GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	1.046×10^{-2}
4	GO:BP	GO:0035640	exploration behavior	1.791×10^{-2}
5	GO:BP	GO:0009987	cellular process	4.075×10^{-2}
6	GO:BP	GO:0007420	brain development	4.611×10^{-2}
7	GO:BP	GO:0033555	multicellular organismal response to stress	4.774×10^{-2}
8	GO:BP	GO:0097120	receptor localization to synapse	4.774×10^{-2}
9	GO:CC	GO:0045211	postsynaptic membrane	3.947×10^{-2}
10	GO:CC	GO:0030425	dendrite	3.947×10^{-2}
11	GO:CC	GO:0030054	cell junction	4.582×10^{-2}

version	e111_eg58_p18_30541362
date	24/05/2024, 15:40:09
organism	hsapiens

g:Profiler

Figure S13. Gene ontology terms associated with the 'Synapse' (brown) module. Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher's one-tailed test, to calculate the significance of functional terms in the input protein list. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05. The y-axis of the plot displays the negative decadic logarithm of the adjusted *p*-value while the x-axis highlights the category of all corresponding GO terms. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms (displayed in the list below the plot).

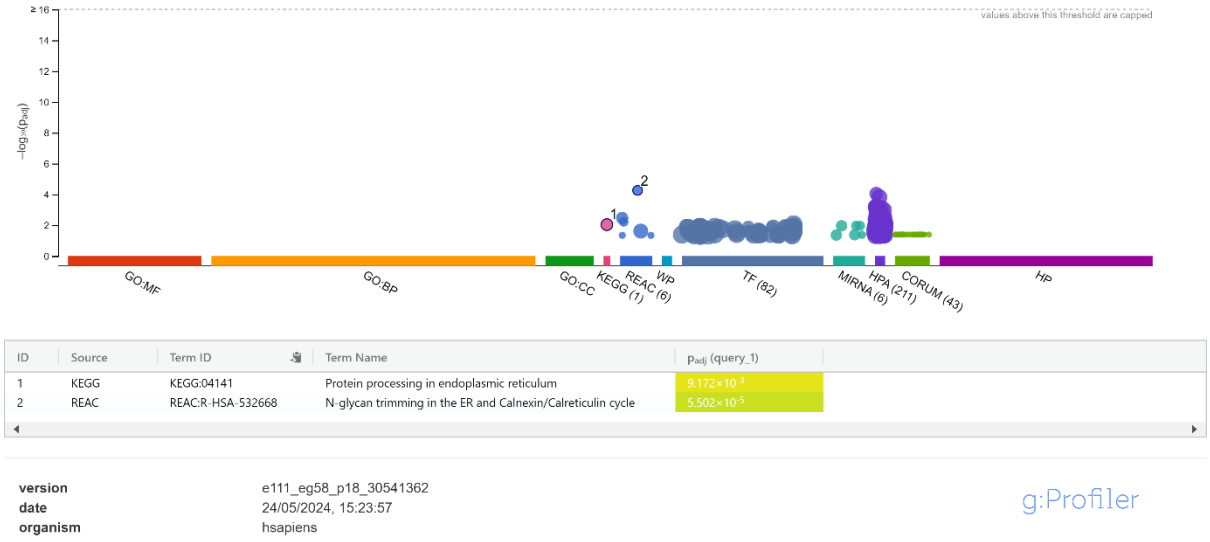


ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0003674	molecular_function	6.419×10^{-3}
2	GO:MF	GO:0005184	neuropeptide hormone activity	8.504×10^{-3}
3	GO:BP	GO:0007399	nervous system development	9.474×10^{-13}
4	GO:BP	GO:0007186	G protein-coupled receptor signaling pathway	1.979×10^{-3}
5	GO:BP	GO:0043087	regulation of GTPase activity	1.198×10^{-2}
6	GO:BP	GO:0035268	protein mannosylation	4.280×10^{-2}
7	GO:CC	GO:0045202	synapse	5.917×10^{-16}
8	GO:CC	GO:0098992	neuronal dense core vesicle	2.752×10^{-4}
9	GO:CC	GO:0034703	cation channel complex	2.143×10^{-2}
10	GO:CC	GO:0005911	cell-cell junction	3.548×10^{-2}

version	e111_eg58_p18_30541362
date	24/05/2024, 15:41:24
organism	hsapiens

g:Profiler

Figure S14. Gene ontology terms associated with the 'Neuronal development' (blue) module. Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher's one-tailed test, to calculate the significance of functional terms in the input protein list. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05. The y-axis of the plot displays the negative decadic logarithm of the adjusted *p*-value while the x-axis highlights the category of all corresponding GO terms. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms (displayed in the list below the plot).

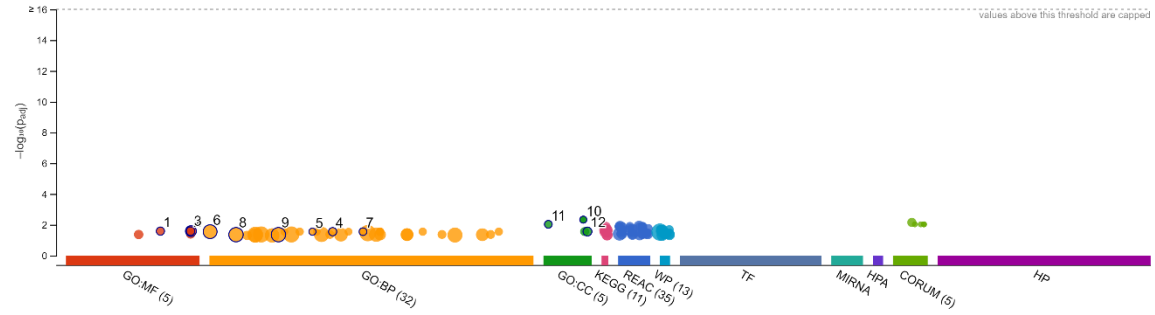


ID	Source	Term ID	Term Name	P_{adj} (query_1)
1	KEGG	KEGG:04141	Protein processing in endoplasmic reticulum	9.172×10^{-3}
2	REAC	REAC:R-HSA-532668	N-glycan trimming in the ER and Calnexin/Calreticulin cycle	5.502×10^{-5}

version e111_eg58_p18_30541362
date 24/05/2024, 15:23:57
organism hsapiens

g:Profiler

Figure S15. Gene ontology terms associated with the 'Protein processing' (yellow) module. Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher's one-tailed test, to calculate the significance of functional terms in the input protein list. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05. The y-axis of the plot displays the negative decadic logarithm of the adjusted *p*-value while the x-axis highlights the category of all corresponding GO terms. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms (displayed in the list below the plot).

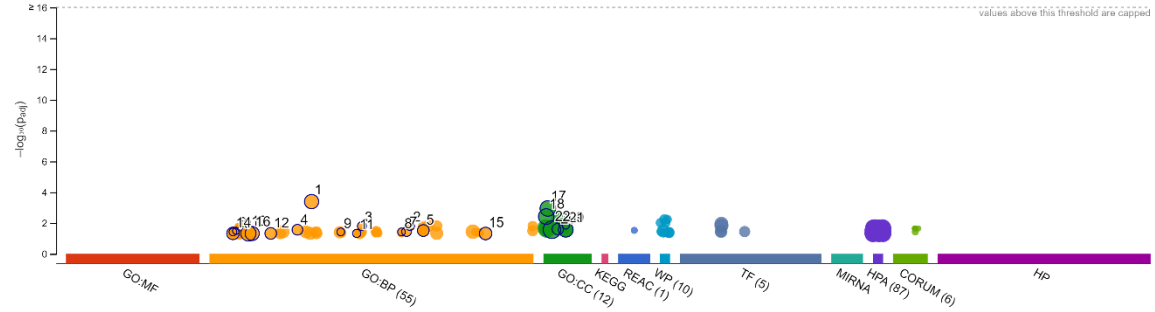


ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0050815	phosphoserine residue binding	2.600×10^{-2}
2	GO:MF	GO:0140311	protein sequestering activity	2.600×10^{-2}
3	GO:MF	GO:0140375	immune receptor activity	2.600×10^{-2}
4	GO:BP	GO:0042149	cellular response to glucose starvation	2.819×10^{-2}
5	GO:BP	GO:0033693	neurofilament bundle assembly	2.819×10^{-2}
6	GO:BP	GO:0000165	MAPK cascade	2.819×10^{-2}
7	GO:BP	GO:0047497	mitochondrion transport along microtubule	2.819×10^{-2}
8	GO:BP	GO:0006468	protein phosphorylation	4.419×10^{-2}
9	GO:BP	GO:0019220	regulation of phosphate metabolic process	4.419×10^{-2}
10	GO:CC	GO:0099160	postsynaptic intermediate filament cytoskeleton	4.698×10^{-3}
11	GO:CC	GO:0005883	neurofilament	9.354×10^{-3}
12	GO:CC	GO:1904115	axon cytoplasm	2.768×10^{-2}

version	e111_eg58_p18_30541362
date	24/05/2024, 15:42:44
organism	hsapiens

g:Profiler

Figure S16. Gene ontology terms associated with the 'Core markers' (greenyellow) module. Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher's one-tailed test, to calculate the significance of functional terms in the input protein list. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05. The y-axis of the plot displays the negative decadic logarithm of the adjusted *p*-value while the x-axis highlights the category of all corresponding GO terms. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms (displayed in the list below the plot).

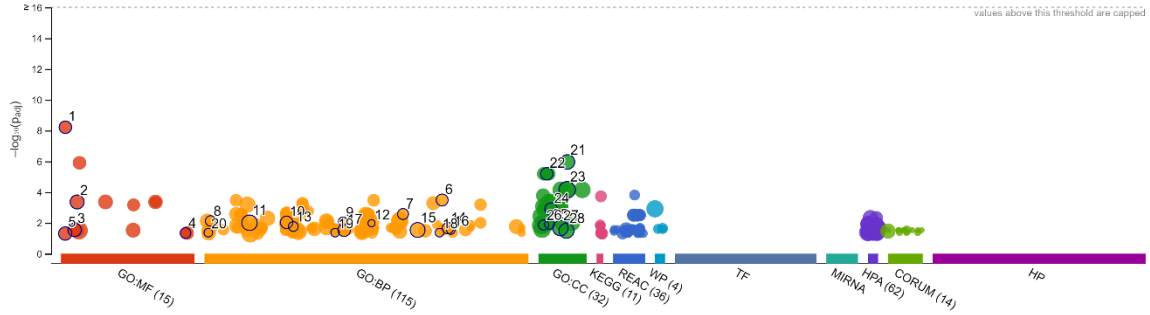


ID	Source	Term ID	Term Name	p_{adj} (query 1)
1	GO:BP	GO:0033554	cellular response to stress	4.067×10^{-4}
2	GO:BP	GO:0070841	inclusion body assembly	1.558×10^{-2}
3	GO:BP	GO:0046825	regulation of protein export from nucleus	1.558×10^{-2}
4	GO:BP	GO:0031333	negative regulation of protein-containing complex assembly	2.736×10^{-2}
5	GO:BP	GO:0072524	pyridine-containing compound metabolic process	3.072×10^{-2}
6	GO:BP	GO:0006473	protein acetylation	3.206×10^{-2}
7	GO:BP	GO:0070301	cellular response to hydrogen peroxide	3.578×10^{-2}
8	GO:BP	GO:0061691	detoxification of hydrogen peroxide	3.775×10^{-2}
9	GO:BP	GO:0043388	positive regulation of DNA binding	3.775×10^{-2}
10	GO:BP	GO:0006089	lactate metabolic process	3.775×10^{-2}
11	GO:BP	GO:0046185	aldehyde catabolic process	4.715×10^{-2}
12	GO:BP	GO:0016052	carbohydrate catabolic process	4.715×10^{-2}
13	GO:BP	GO:0009056	catabolic process	4.715×10^{-2}
14	GO:BP	GO:0006091	generation of precursor metabolites and energy	4.715×10^{-2}
15	GO:BP	GO:1902903	regulation of supramolecular fiber organization	4.782×10^{-2}
16	GO:BP	GO:0009628	response to abiotic stimulus	4.782×10^{-2}
17	GO:CC	GO:0005829	cytosol	1.133×10^{-3}
18	GO:CC	GO:0005634	nucleus	3.785×10^{-3}
19	GO:CC	GO:0043220	Schmidt-Lanterman incisure	1.472×10^{-2}
20	GO:CC	GO:0031965	nuclear membrane	2.464×10^{-2}
21	GO:CC	GO:0043232	intracellular non-membrane-bounded organelle	2.729×10^{-2}
22	GO:CC	GO:0016020	membrane	2.729×10^{-2}

version	e111_eg58_p18_30541362
date	24/05/2024, 15:43:51
organism	hsapiens

g:Profiler

Figure S17. Gene ontology terms associated with the 'Stress response' (magenta) module. Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher's one-tailed test, to calculate the significance of functional terms in the input protein list. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05. The y-axis of the plot displays the negative decadic logarithm of the adjusted *p*-value while the x-axis highlights the category of all corresponding GO terms. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms (displayed in the list below the plot).

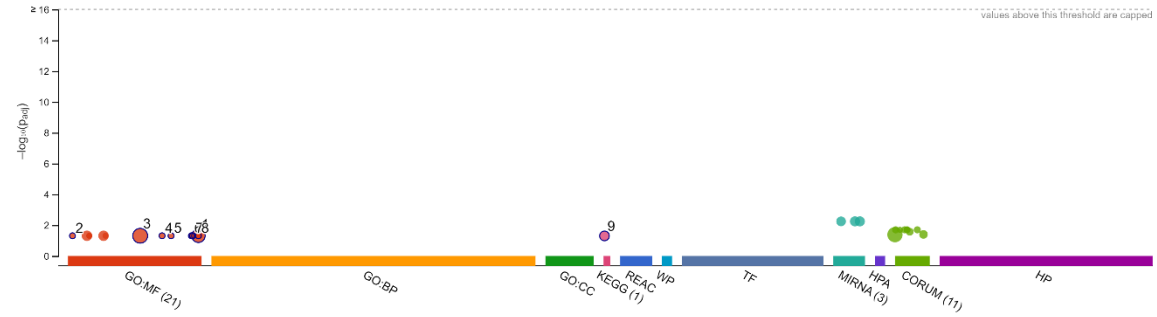


ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0003779	actin binding	6.058×10^{-9}
2	GO:MF	GO:0005198	structural molecule activity	4.316×10^{-4}
3	GO:MF	GO:0004857	enzyme inhibitor activity	2.823×10^{-2}
4	GO:MF	GO:0140297	DNA-binding transcription factor binding	4.513×10^{-2}
5	GO:MF	GO:0003723	RNA binding	4.886×10^{-2}
6	GO:BP	GO:0110053	regulation of actin filament organization	3.241×10^{-4}
7	GO:BP	GO:0070527	platelet aggregation	2.628×10^{-3}
8	GO:BP	GO:0001885	endothelial cell development	7.384×10^{-3}
9	GO:BP	GO:0045022	early endosome to late endosome transport	8.047×10^{-3}
10	GO:BP	GO:0030031	cell projection assembly	9.127×10^{-3}
11	GO:BP	GO:0009888	tissue development	9.691×10^{-3}
12	GO:BP	GO:0051220	cytoplasmic sequestering of protein	1.036×10^{-2}
13	GO:BP	GO:0031529	ruffle organization	1.704×10^{-2}
14	GO:BP	GO:0140694	non-membrane-bounded organelle assembly	2.175×10^{-2}
15	GO:BP	GO:0072359	circulatory system development	2.864×10^{-2}
16	GO:BP	GO:0150105	protein localization to cell-cell junction	2.941×10^{-2}
17	GO:BP	GO:0045216	cell-cell junction organization	2.978×10^{-2}
18	GO:BP	GO:0099623	regulation of cardiac muscle cell membrane repolarization	4.261×10^{-2}
19	GO:BP	GO:0043297	apical junction assembly	4.261×10^{-2}
20	GO:BP	GO:0001502	cartilage condensation	4.261×10^{-2}
21	GO:CC	GO:0070161	anchoring junction	1.089×10^{-6}
22	GO:CC	GO:0015629	actin cytoskeleton	6.391×10^{-6}
23	GO:CC	GO:0070062	extracellular exosome	7.034×10^{-5}
24	GO:CC	GO:0031252	cell leading edge	1.278×10^{-3}
25	GO:CC	GO:0030496	midbody	1.191×10^{-2}
26	GO:CC	GO:0005902	microvillus	1.352×10^{-2}
27	GO:CC	GO:0042995	cell projection	2.142×10^{-2}
28	GO:CC	GO:0062023	collagen-containing extracellular matrix	3.192×10^{-2}

version	e111_eg58_p18_30541362
date	24/05/2024, 15:45:08
organism	hsapiens

g:Profiler

Figure S18. Gene ontology terms associated with the 'Actin binding' (pink) module. Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher's one-tailed test, to calculate the significance of functional terms in the input protein list. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05. The y-axis of the plot displays the negative decadic logarithm of the adjusted *p*-value while the x-axis highlights the category of all corresponding GO terms. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms (displayed in the list below the plot).



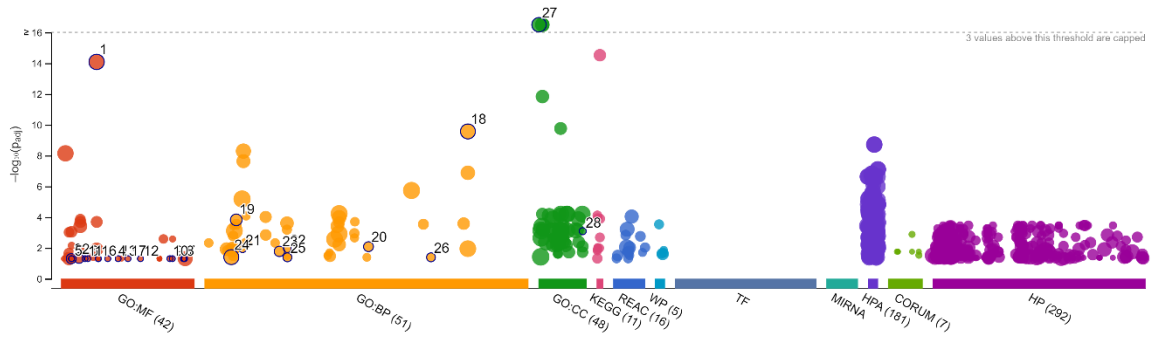
ID	Source	Term ID	Term Name	P_{adj} (query_1)
1	GO:MF	GO:1901681	sulfur compound binding	4.874×10^{-2}
2	GO:MF	GO:0003827	alpha-1,3-mannosylglycoprotein 2-beta-N-acetylglucosaminyltransferase activity	4.874×10^{-2}
3	GO:MF	GO:0044877	protein-containing complex binding	4.874×10^{-2}
4	GO:MF	GO:0050659	N-acetylgalactosamine 4-sulfate 6-O-sulfotransferase activity	4.874×10^{-2}
5	GO:MF	GO:0070300	phosphatidic acid binding	4.874×10^{-2}
6	GO:MF	GO:0120019	phosphatidylcholine transfer activity	4.874×10^{-2}
7	GO:MF	GO:0140337	diacylglyceride transfer activity	4.874×10^{-2}
8	GO:MF	GO:1901611	phosphatidylglycerol binding	4.874×10^{-2}
9	KEGG	KEGG:00532	Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	4.970×10^{-2}

version	e111_eg58_p18_30541362
date	24/05/2024, 14:42:43
organism	hsapiens

g:Profiler

Figure S19. Gene ontology terms associated with the ‘GAG processing’ (salmon) module. Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher’s one-tailed test, to calculate the significance of functional terms in the input protein list. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05. The y-axis of the plot displays the negative decadic logarithm of the adjusted *p*-value while the x-axis highlights the category of all corresponding GO terms. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms (displayed in the list below the plot).

Figure S20. Gene ontology terms associated with the 'Extracellular matrix 1' (green) module. Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher's one-tailed test, to calculate the significance of functional terms in the input protein list. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05. The y-axis of the plot displays the negative decadic logarithm of the adjusted *p*-value while the x-axis highlights the category of all corresponding GO terms. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms (displayed in the list below the plot).

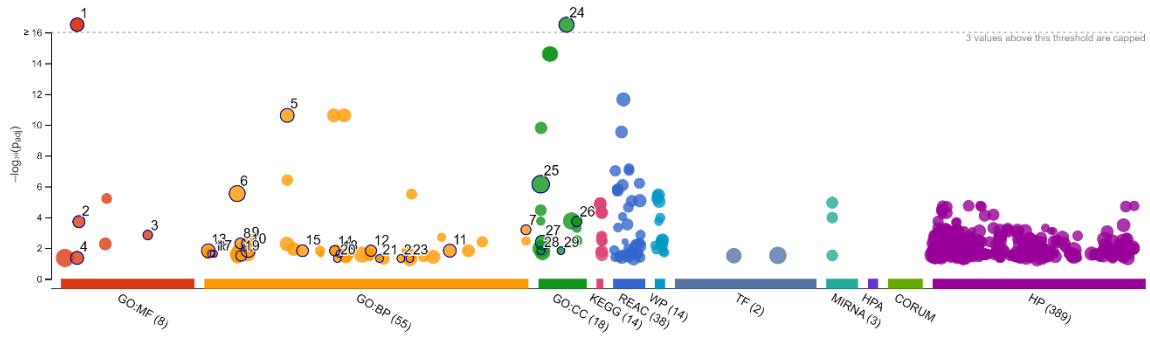


ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0016787	hydrolase activity	8.227×10^{-15}
2	GO:MF	GO:0005537	mannose binding	4.921×10^{-2}
3	GO:MF	GO:0004519	endonuclease activity	4.942×10^{-2}
4	GO:MF	GO:0004531	deoxyribonuclease II activity	4.942×10^{-2}
5	GO:MF	GO:0004567	beta-mannosidase activity	4.942×10^{-2}
6	GO:MF	GO:0010465	nerve growth factor receptor activity	4.942×10^{-2}
7	GO:MF	GO:0120146	sulfatide binding	4.942×10^{-2}
8	GO:MF	GO:0106330	sialate 9-O-acetylesterase activity	4.942×10^{-2}
9	GO:MF	GO:0102121	ceramidase activity	4.942×10^{-2}
10	GO:MF	GO:0097655	serpin family protein binding	4.942×10^{-2}
11	GO:MF	GO:0008889	glycerophosphodiester phosphodiesterase activity	4.942×10^{-2}
12	GO:MF	GO:0047390	glycerophosphocholine cholinephosphodiesterase activity	4.942×10^{-2}
13	GO:MF	GO:0033897	ribonuclease T2 activity	4.942×10^{-2}
14	GO:MF	GO:0030379	neurotensin receptor activity, non-G protein-coupled	4.942×10^{-2}
15	GO:MF	GO:0017064	fatty acid amide hydrolase activity	4.942×10^{-2}
16	GO:MF	GO:0017040	N-acylsphingosine amidohydrolase activity	4.942×10^{-2}
17	GO:MF	GO:0042393	histone binding	4.942×10^{-2}
18	GO:BP	GO:1901575	organic substance catabolic process	2.694×10^{-10}
19	GO:BP	GO:0007033	vacuole organization	1.490×10^{-4}
20	GO:BP	GO:0050885	neuromuscular process controlling balance	8.219×10^{-3}
21	GO:BP	GO:0008333	endosome to lysosome transport	1.113×10^{-2}
22	GO:BP	GO:0030209	dermatan sulfate catabolic process	1.113×10^{-2}
23	GO:BP	GO:0019915	lipid storage	1.668×10^{-2}
24	GO:BP	GO:0006508	proteolysis	3.938×10^{-2}
25	GO:BP	GO:0030214	hyaluronan catabolic process	4.047×10^{-2}
26	GO:BP	GO:0097067	cellular response to thyroid hormone stimulus	4.047×10^{-2}
27	GO:CC	GO:0000323	lytic vacuole	4.746×10^{-19}
28	GO:CC	GO:1905379	beta-N-acetylhexosaminidase complex	8.124×10^{-4}

version	e111_eg58_p18_30541362
date	24/05/2024, 15:46:15
organism	hsapiens

g:Profiler

Figure S21. Gene ontology terms associated with the ‘Lysosome’ (purple) module. Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher’s one-tailed test, to calculate the significance of functional terms in the input protein list. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05. The y-axis of the plot displays the negative decadic logarithm of the adjusted *p*-value while the x-axis highlights the category of all corresponding GO terms. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms (displayed in the list below the plot).

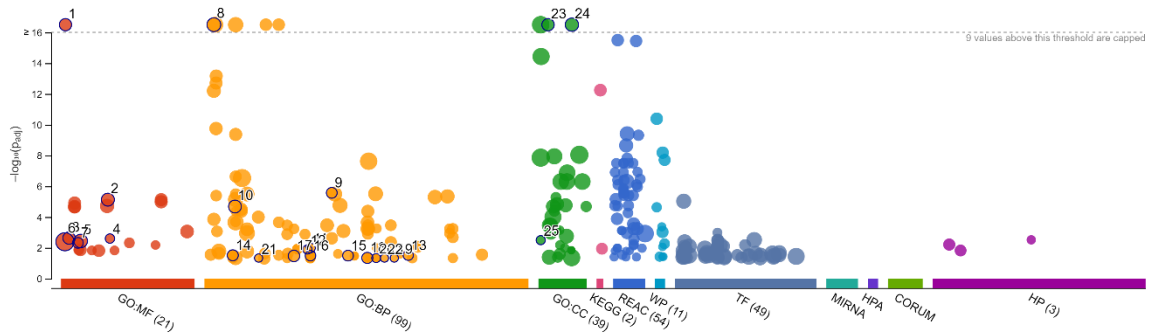


ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0005201	extracellular matrix structural constituent	5.640×10^{-20}
2	GO:MF	GO:0005518	collagen binding	1.935×10^{-4}
3	GO:MF	GO:0048407	platelet-derived growth factor binding	1.416×10^{-3}
4	GO:MF	GO:0005178	integrin binding	4.386×10^{-2}
5	GO:BP	GO:0030198	extracellular matrix organization	2.434×10^{-11}
6	GO:BP	GO:0007155	cell adhesion	2.900×10^{-6}
7	GO:BP	GO:2001044	regulation of integrin-mediated signaling pathway	6.530×10^{-4}
8	GO:BP	GO:0007492	endoderm development	4.968×10^{-3}
9	GO:BP	GO:0009790	embryo development	5.378×10^{-3}
10	GO:BP	GO:0009725	response to hormone	1.448×10^{-2}
11	GO:BP	GO:0141091	transforming growth factor beta receptor superfamily signaling pathway	1.481×10^{-2}
12	GO:BP	GO:0051147	regulation of muscle cell differentiation	1.482×10^{-2}
13	GO:BP	GO:0001501	skeletal system development	1.482×10^{-2}
14	GO:BP	GO:0043200	response to amino acid	1.482×10^{-2}
15	GO:BP	GO:0032963	collagen metabolic process	1.482×10^{-2}
16	GO:BP	GO:0001957	intramembranous ossification	2.363×10^{-2}
17	GO:BP	GO:0002248	connective tissue replacement involved in inflammatory response wound healing	2.363×10^{-2}
18	GO:BP	GO:0044341	sodium-dependent phosphate transport	2.363×10^{-2}
19	GO:BP	GO:0007601	visual perception	3.140×10^{-2}
20	GO:BP	GO:0043589	skin morphogenesis	4.738×10^{-2}
21	GO:BP	GO:0055062	phosphate ion homeostasis	4.738×10^{-2}
22	GO:BP	GO:0070206	protein trimerization	4.738×10^{-2}
23	GO:BP	GO:0071492	cellular response to UV-A	4.738×10^{-2}
24	GO:CC	GO:0062023	collagen-containing extracellular matrix	4.664×10^{-17}
25	GO:CC	GO:0005576	extracellular region	7.176×10^{-7}
26	GO:CC	GO:0098637	protein complex involved in cell-matrix adhesion	1.869×10^{-4}
27	GO:CC	GO:0005788	endoplasmic reticulum lumen	4.222×10^{-3}
28	GO:CC	GO:0005606	laminin-1 complex	1.475×10^{-2}
29	GO:CC	GO:0043259	laminin-10 complex	1.475×10^{-2}

version	e111_eg58_p18_30541362
date	24/05/2024, 15:47:06
organism	hsapiens

g:Profiler

Figure S22. Gene ontology terms associated with the ‘Extracellular matrix 2’ (black) module. Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher’s one-tailed test, to calculate the significance of functional terms in the input protein list. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05. The y-axis of the plot displays the negative decadic logarithm of the adjusted *p*-value while the x-axis highlights the category of all corresponding GO terms. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms (displayed in the list below the plot).



ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0003823	antigen binding	2.451×10^{-17}
2	GO:MF	GO:0030414	peptidase inhibitor activity	7.414×10^{-6}
3	GO:MF	GO:0004252	serine-type endopeptidase activity	2.328×10^{-3}
4	GO:MF	GO:0031210	phosphatidylcholine binding	2.508×10^{-3}
5	GO:MF	GO:0008289	lipid binding	3.636×10^{-3}
6	GO:MF	GO:0003674	molecular_function	4.049×10^{-3}
7	GO:MF	GO:0005319	lipid transporter activity	4.780×10^{-3}
8	GO:BP	GO:0002250	adaptive immune response	1.408×10^{-24}
9	GO:BP	GO:0042730	fibrinolysis	2.682×10^{-6}
10	GO:BP	GO:0006869	lipid transport	2.074×10^{-5}
11	GO:BP	GO:0034377	plasma lipoprotein particle assembly	1.040×10^{-2}
12	GO:BP	GO:0033700	phospholipid efflux	1.298×10^{-2}
13	GO:BP	GO:0071347	cellular response to interleukin-1	3.071×10^{-2}
14	GO:BP	GO:0006639	acylglycerol metabolic process	3.071×10^{-2}
15	GO:BP	GO:0045834	positive regulation of lipid metabolic process	3.246×10^{-2}
16	GO:BP	GO:0034381	plasma lipoprotein particle clearance	3.246×10^{-2}
17	GO:BP	GO:0031638	zymogen activation	3.386×10^{-2}
18	GO:BP	GO:0050764	regulation of phagocytosis	4.543×10^{-2}
19	GO:BP	GO:0061365	positive regulation of triglyceride lipase activity	4.655×10^{-2}
20	GO:BP	GO:0051838	cytolysis by host of symbiont cells	4.655×10^{-2}
21	GO:BP	GO:0010901	regulation of very-low-density lipoprotein particle remodeling	4.655×10^{-2}
22	GO:BP	GO:0060438	trachea development	4.655×10^{-2}
23	GO:CC	GO:0019814	immunoglobulin complex	2.285×10^{-30}
24	GO:CC	GO:0072562	blood microparticle	1.431×10^{-23}
25	GO:CC	GO:0005579	membrane attack complex	3.190×10^{-3}

version	e111_eg58_p18_30541362
date	24/05/2024, 15:48:06
organism	hsapiens

g:Profiler

Figure S23. Gene ontology terms associated with the 'Immune response' (turquouise) module. Gene ontology (GO) analysis of WGCNA modules was conducted with g:Profiler, a web server for functional enrichment analysis. g:Profiler performs statistical overrepresentation analysis utilising cumulative hypergeometric probability, also known as Fisher's one-tailed test, to calculate the significance of functional terms in the input protein list. Multiple testing correction was performed with the method of Benjamini and Hochberg with a threshold of <0.05. The y-axis of the plot displays the negative decadic logarithm of the adjusted *p*-value while the x-axis highlights the category of all corresponding GO terms. GO results were then filtered to reduce redundancy and highlight driver terms, i.e. representative GO terms for a larger group of terms (displayed in the list below the plot).

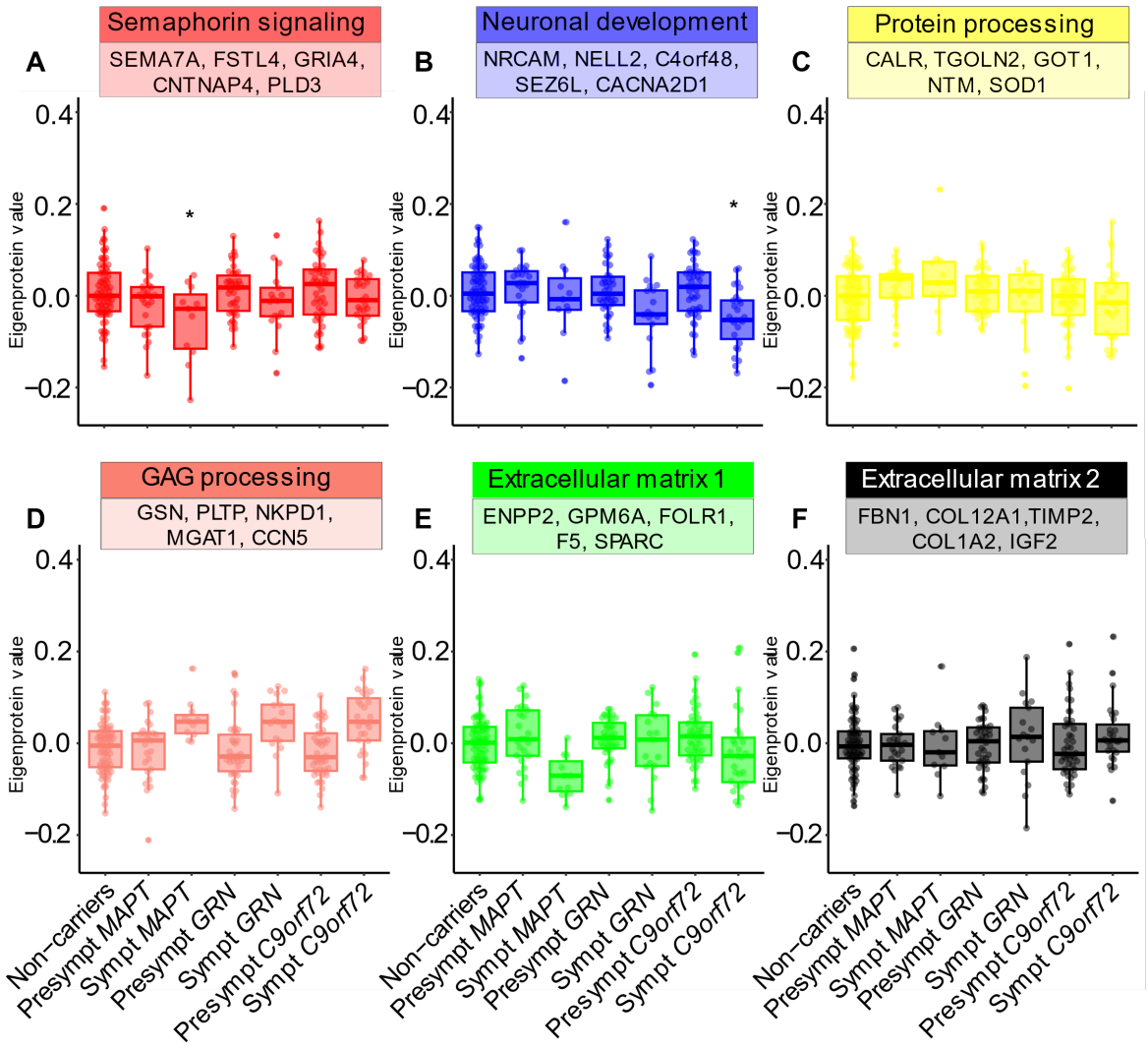


Figure S24. Remaining WGCNA modules that are not depicted in the main body of the manuscript. Eigenprotein values obtained through weighted gene co-expression network analysis (WGCNA), plotted across the continuum of controls as well as presymptomatic and symptomatic mutation carriers. The plots represent Eigenprotein values for (A) 'Semaphorin signaling', (B) 'Neuronal development', (C) 'Protein processing', (D) 'GAG processing', (E) 'Extracellular matrix 1', and (F) 'Extracellular matrix 2' modules. Hub proteins are displayed in boxes. *$P < 0.05$ compared with non-carriers.

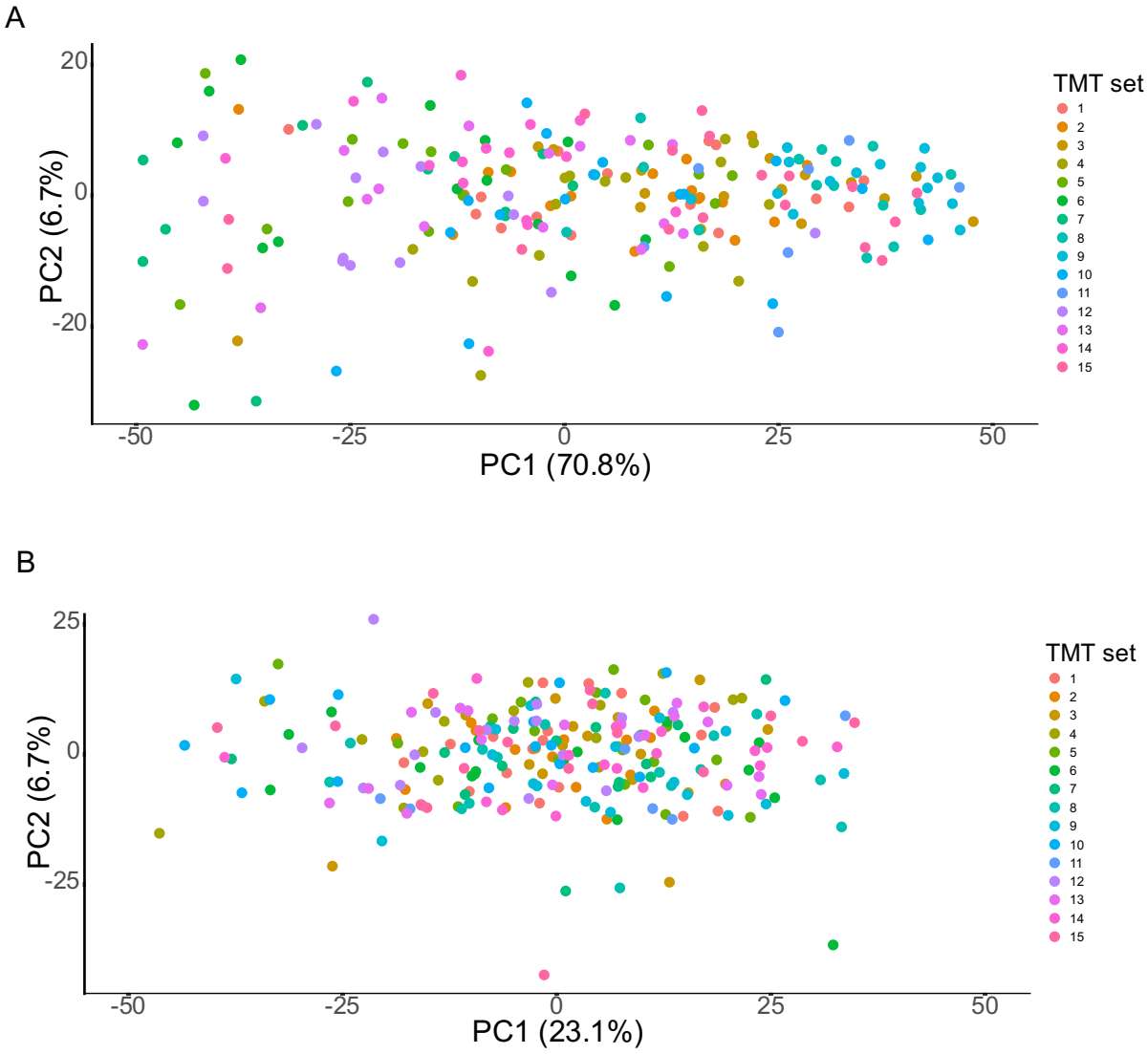


Figure S25. Evaluation of TMT batch effect and sample outliers prior to (A) and after normalization (B) via PCA. PCA was performed on all participants (n=238) of the GENFI cohort including proteins without missingness. (A) Before normalisation, inter-sample variance was high: Most samples separated along PC 1, which accounted for 70% of the total variance. Moreover, batches clustered together. (B) Upon normalisation, overall sample variance (PC1=23% and PC2=6%) and clustering of batches were reduced.