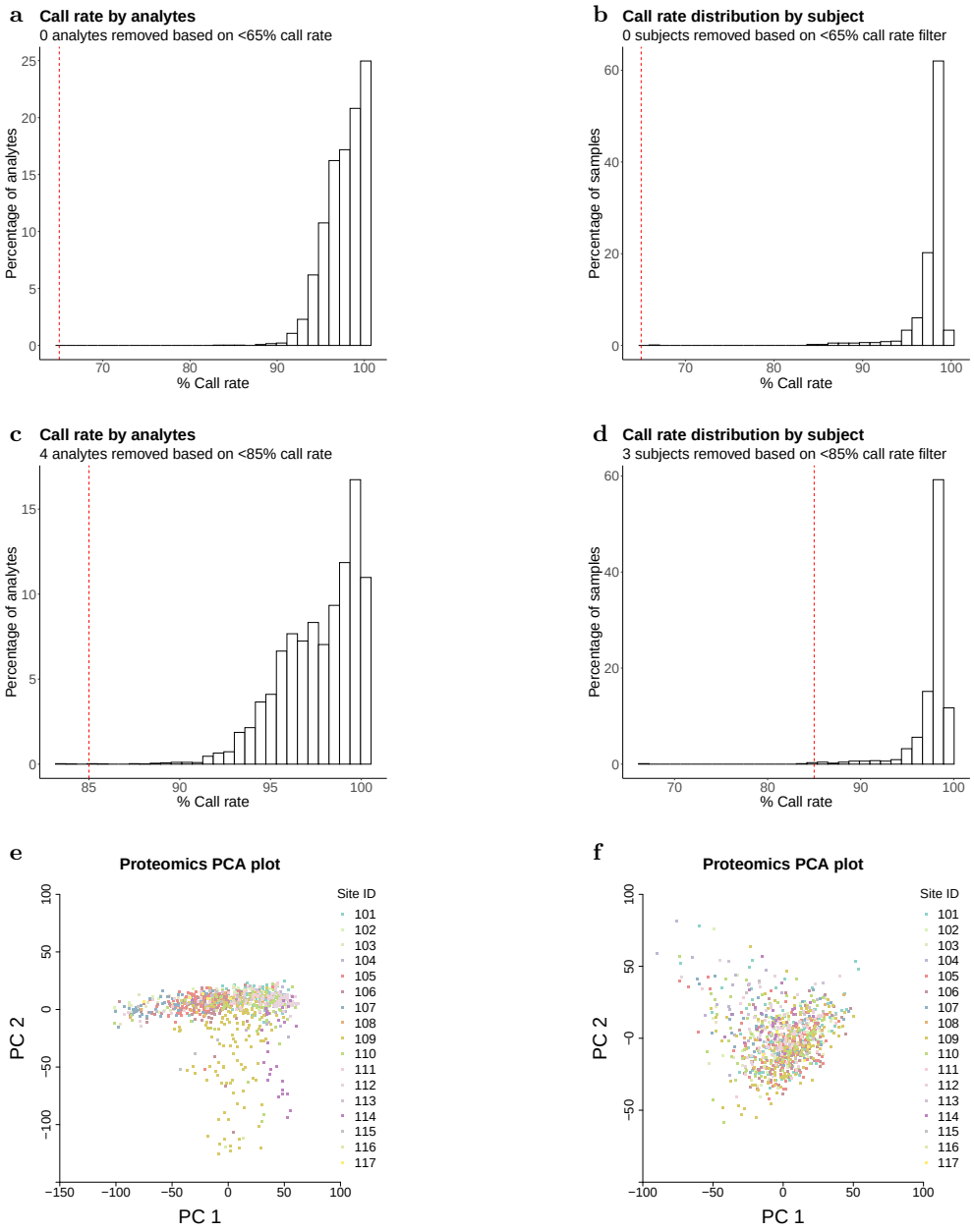
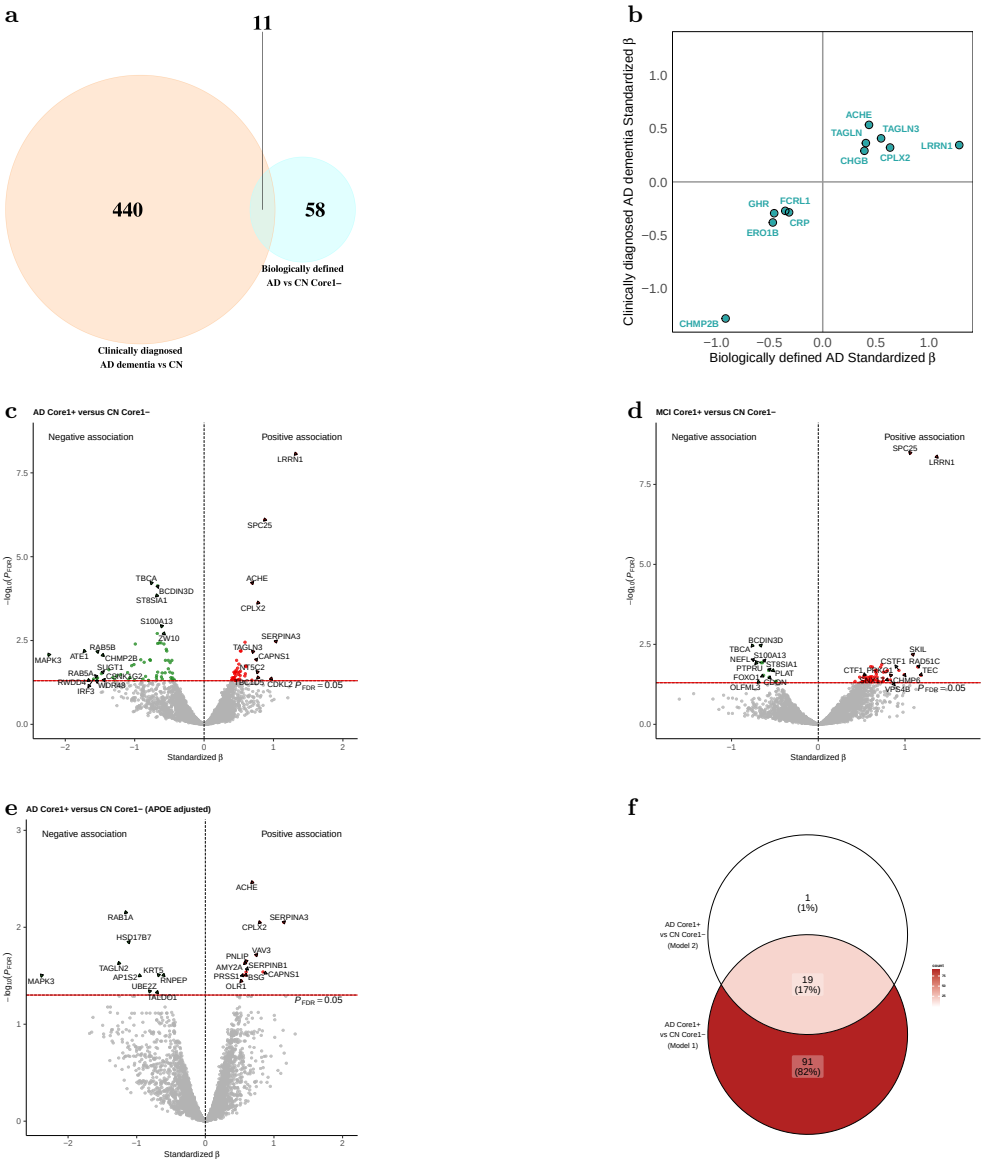
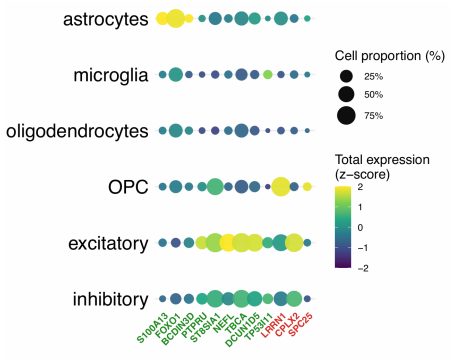


Extended Data Fig.1 Assessment of proteomics quality control steps. a-d Histograms of call rate distributions with thresholds used at different proteomics QC steps: 65% call rate (red vertical line) by proteins (**a**), 65% call rate by participants (**b**), 85% call rate by proteins (**c**), and 85% call rate by participants (**d**). **e,f,** Scatter plots of the first two proteomic principal components (PC1 and PC2) before (**e**) and after (**f**) adjustment for technical effects. Colours indicate recruitment site (17 sites) in the Bio-Hermes Study.

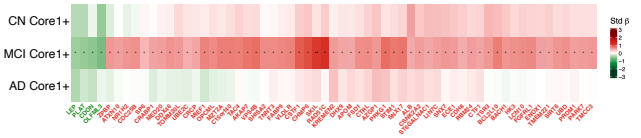


Extended Data Fig.2 Proteins with *APOE-$\epsilon 4$*-dependent associations with AD. **a,** Venn diagram showing the overlap of significant dysregulated proteins between clinically diagnosed AD dementia and biologically defined AD. **b,** Comparison of the standardised beta estimates from linear regression models for clinically diagnosed AD dementia and biologically defined AD in the Bio-Hermes Study. Protein associations with Benjamini–Hochberg corrected $P_{\mathrm{FDR}} < 0.05$ are shown in green. **c-e,** Volcano plots showing the standardised beta (x axis) and $-\log_{10}(P_{\mathrm{FDR}})$ (y axis) for AD dementia Core1+ versus CN Core1− in model 1 (**c**), MCI Core1+ versus CN Core1− for model 1 (**d**), and AD dementia Core1+ versus CN Core1− in model 2 (**e**). **f,** Venn diagram showing the overlap of significant dysregulated proteins from AD dementia Core1+ versus CN Core1− in models 1 and 2.

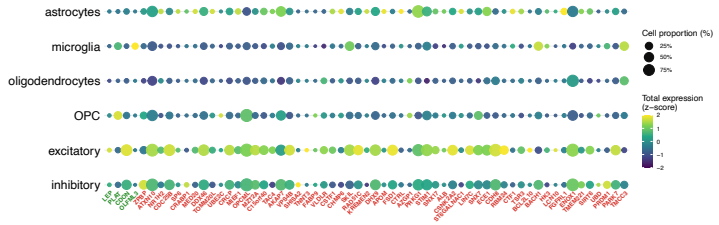
a



b

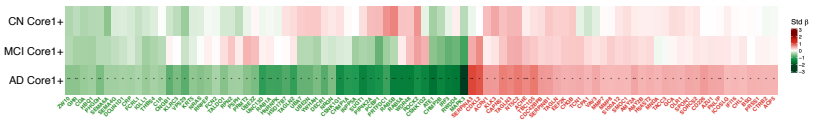


c

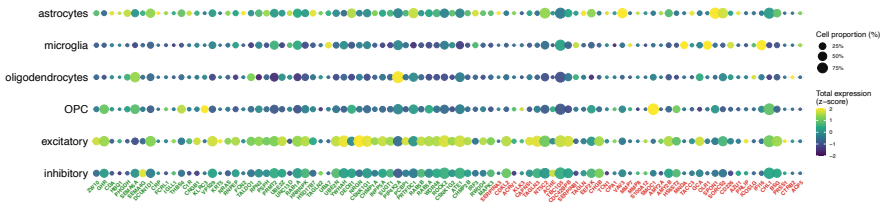


Extended Data Fig.3 (continued on next page)

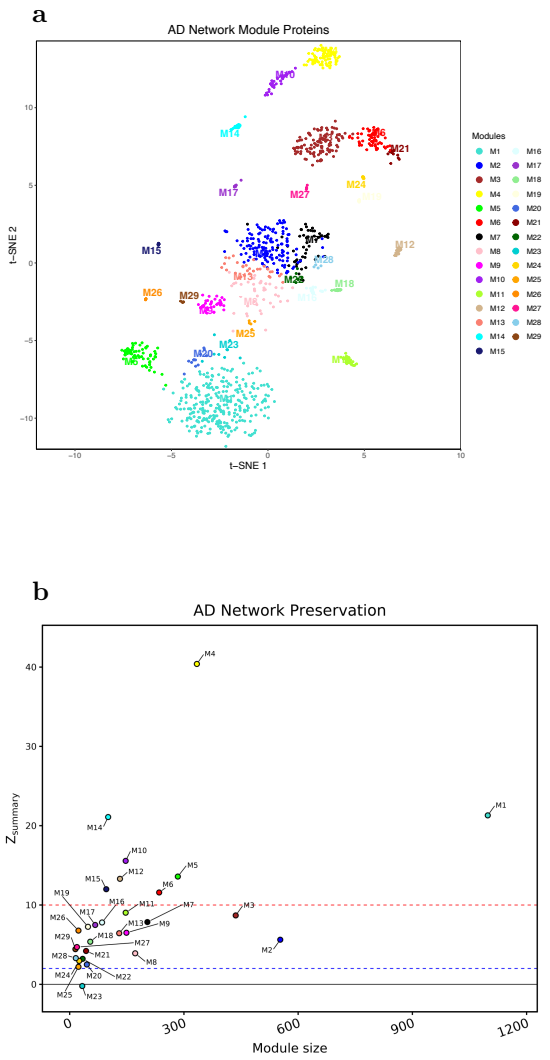
d



e

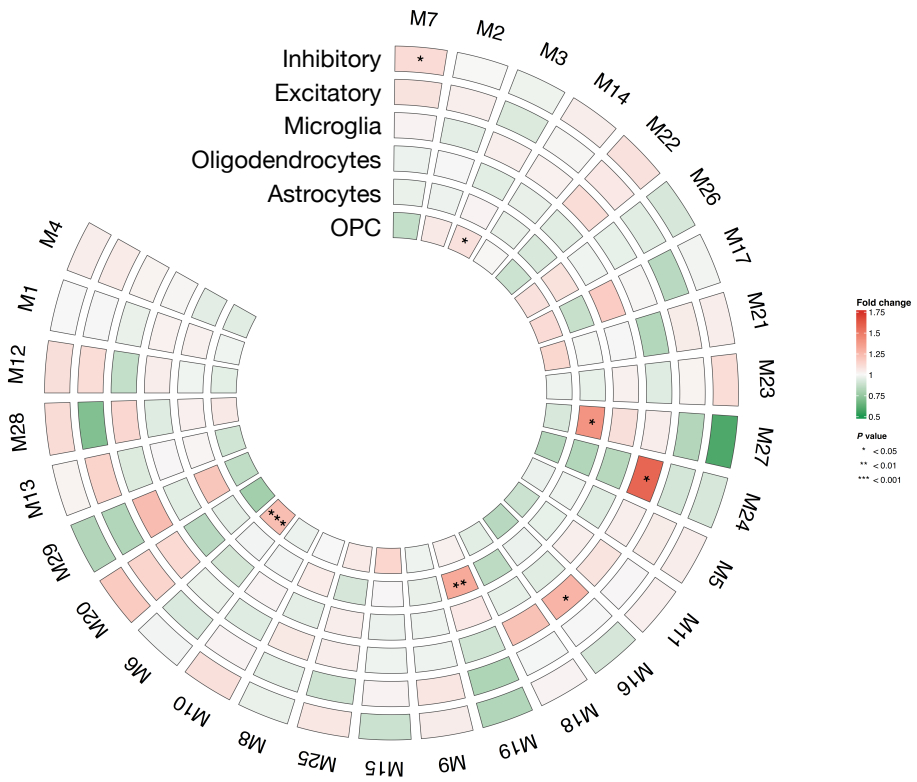


Extended Data Fig.3 Characterisation of DAPs. a, Proportion of expression by cell type based on single-cell transcriptomic data from the middle temporal gyrus for the overlapping proteins shown in Fig. 2d. **b,d,** Heat maps showing changes in uniquely dysregulated proteins in MCI Core1+ (b) and AD dementia Core1+ (d) phases. Linear regression models were adjusted for age, sex, and mean overall protein level. Colours represent standardised β values, and asterisks indicate significance after Benjamini–Hochberg FDR correction ($^{\text{no symbol}}P \geq 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). **c,e,** Proportion of expression by cell type based on single-cell transcriptomic data from the middle temporal gyrus for uniquely dysregulated proteins in the MCI Core1+ group (c) and the AD dementia Core1+ phase (e).

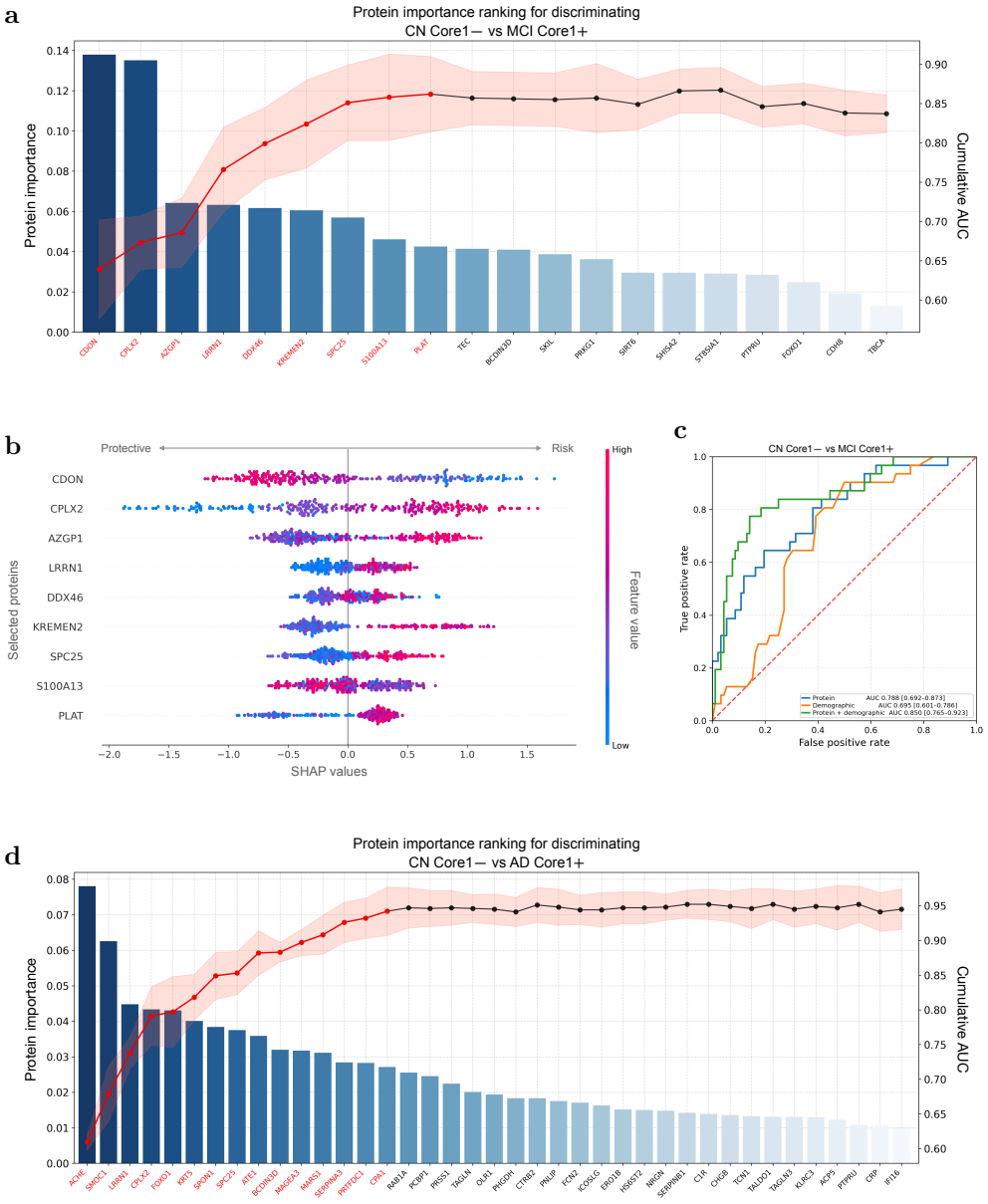


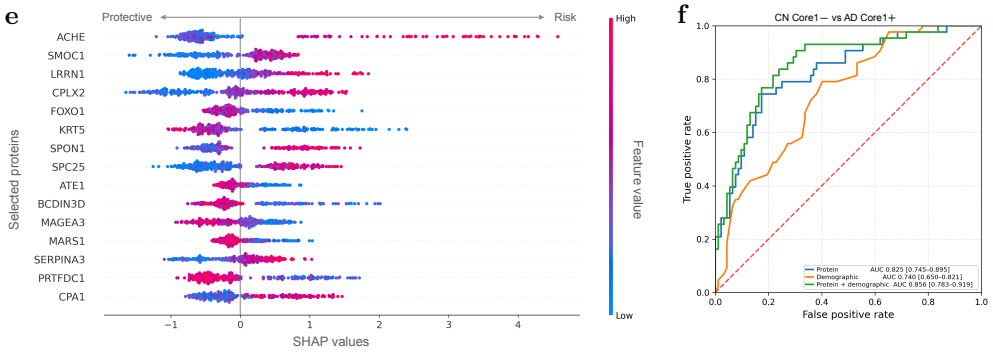
Extended Data Fig.4 AD protein network is preserved using an independent dimensionality reduction method and in an external cohort. a, Dimensionality reduction and visualization by t-SNE was applied to proteins in the top 25% by kME value within each AD network module. Proteins are colour-coded according to WGCNA module membership in Fig. 3a. **b,** Preservation AD network modules from Bio-Hermes in the GNCP validation cohort. Modules with $Z_{summary} \geq 1.96$ (blue dotted line) were considered preserved,; modules that had a $Z_{summary}$ score ≥ 10 (red dotted line) were considered highly preserved.

Extended Data Fig.5 AD protein network enriched GO biological pathways. Summary of biological pathways enrichment analyses using GO databases for each protein module identified in the AD network. For enrichment analyses the SomaScan-measured proteins were used as background. Shown are the top 10 enriched GO biological pathways with adjusted significance levels of $P_{\mathrm{FDR}} < 0.05$ after redundancy reduction. Colour shade corresponds to statistical significance, represented as $-\log_{10}(P_{\mathrm{FDR}})$.

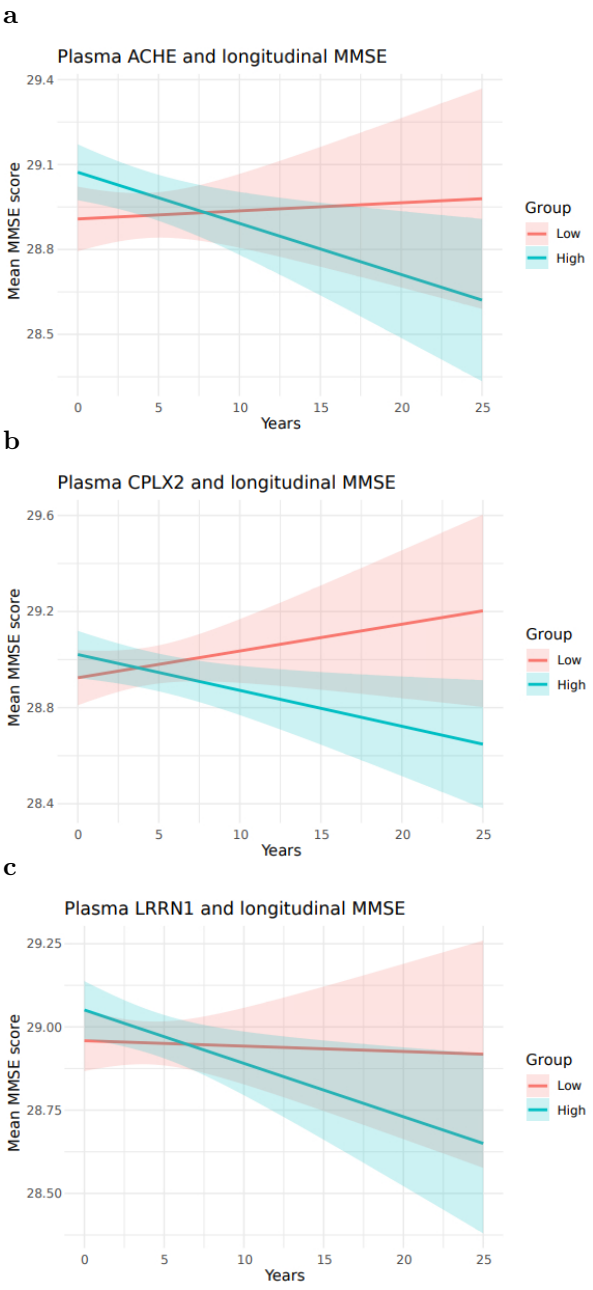


Extended Data Fig.6 AD protein network cell-type enrichment Cell-type enrichment analyses based on single-cell transcriptomics data from the middle temporal gyrus using proteins from 29 co-expression network modules. Bootstrap enrichment tests were performed. Colour intensities represent fold change, and asterisks indicate statistical significance after Benjamini–Hochberg FDR correction ($^{\text{no symbol}} P \geq 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$)

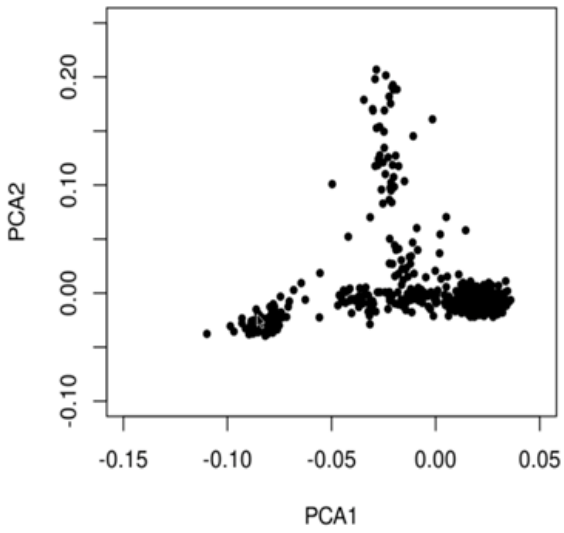

a
Protein importance ranking for discriminating
CN Core1− vs MCI Core1+
Protein importance
Cumulative AUC
CDON
CPLX2
AZGP1
LRRN1
DDX46
KREMEN2
SPC25
S100A13
PLAT
TEC
BCDIN3D
SKIL
PRKG1
SIRT6
SH3GL2
ST8SIA1
PTPRU
FOXO1
CDH8
TBCA
b
Protective
Risk
High
Low
Feature value
Selected proteins
SHAP values
c
CN Core1− vs MCI Core1+
True positive rate
False positive rate
Protein AUC 0.788 [0.692–0.873]
Demographic AUC 0.695 [0.601–0.786]
Protein + demographic AUC 0.850 [0.765–0.923]
d
Protein importance ranking for discriminating
CN Core1− vs AD Core1+
Protein importance
Cumulative AUC



Extended Data Fig.7 Protein selection, feature importance, and ROC curves for distinguishing MCI Core1+ and AD dementia Core1+ from controls. a,d, Protein panel selection from the derivation set. Bar plots demonstrate ranked importance of proteins according to contribution to distinguishing CN Core1- from MCI Core1+(**a**) and AD Core1+(**d**), assessed by information gain. Line plots illustrate the cumulative AUC (right axis) obtained by sequentially adding proteins one at a time in each iteration. Shaded areas represent standard errors estimated using 5-fold cross-validation within the derivation set. Proteins retained in final panels are highlighted in red. Protein selection was terminated when no further AUC increment $\geq$0.005 was observed over two consecutive iterations. **b,e,** SHAP plots illustrating the contributions of selected proteins to model predictions. The x axis indicates impact on model output, with positive values (right) corresponding to the higher probability of MCI Core1+(**b**) or AD Core1+(**e**) and negative values (left) corresponding to cognitive normal status. Colour of the horizontal bars indicates the magnitude of contribution, shown as a gradient. **c,f,** ROC curves derived from the internal replication cohort showing discriminatory performance (AUC with 95% CIs) of selected protein panels alone and combined with demographic variables (age and sex) for distinguishing CN Core1- from MCI Core1+(**c**) or AD Core1+(**f**). AUCs and CIs were estimated using bootstrap resampling (2,000 iterations).



Extended Data Fig.8 Longitudinal cognitive trajectories among cognitively normal individuals in the GNPC, stratified by baseline levels of top three plasma proteins. a-c, Trajectories were derived from linear mixed-effects models with the interaction between each protein and time as a predictor, adjusted for age at baseline, sex, the first two proteomic PCs, and baseline MMSE score.



Extended Data Fig.9 First two principal components for European and near-European population in the Bio-Hermes cohort.