

Sex and aging signatures of proteomics in human cerebrospinal fluid identify distinct clusters linked to neurodegeneration.

Dahun Seo^{1,2*}, Cheolmin Matthew Lee^{1,3*}, Catherine Apio^{1,4*}, Gyujiin Heo^{1,5}, Jigyasha Timsina^{1,5}, Pat Kohlfeld^{1,5}, Merce Boada⁶, Adelina Orellana⁶, Maria Victoria Fernandez⁶, Agustin Ruiz^{6,7,8}, John C. Morris^{9,10}, Suzanne E. Schindler^{9,10}, Taesung Park^{2,4}, Carlos Cruchaga^{1,5,9,11,¶}, Yun Ju Sung^{1,5,12,¶}

1 NeuroGenomics and Informatics Center, Washington University School of Medicine, St. Louis, MO, USA

2 Department of Statistics, Seoul National University, Seoul, Korea

3 Health Sciences Integrated Program, Northwestern University, Chicago, IL, USA

4 Interdisciplinary programs in Bioinformatics, Seoul National University, Seoul 08826, Korea

5 Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA

6 Research Center and Memory Clinic, ACE Alzheimer Center Barcelona, Universitat Internacional de Catalunya, Barcelona, Spain

7 Networking Research Center on Neurodegenerative Diseases (CIBERNED), Instituto de Salud Carlos III, Madrid, Spain.

8 Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases. University of Texas Health Sciences Center, Tx, United States.

9 The Charles F. and Joanne Knight Alzheimer Disease Research Center, Washington University School of Medicine, St. Louis, MO 63108, USA

10 Department of Neurology, Washington University School of Medicine, St Louis, MO 63108, USA

11 Hope Center, Washington University School of Medicine, St. Louis, MO, USA

12 Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, USA

*These authors contributed equally

¶These authors contributed equally

Corresponding Author: Yun Ju Sung, PhD
Washington University School of Medicine
4444 Forest Park Ave
Room 5501, Box 8134
St. Louis, MO 63108
Email: yunju@wustl.edu

Abstract

Sex and age are major risk factors for chronic diseases. Recent studies examining age-related molecular changes in plasma provided insights into age-related disease biology. Cerebrospinal fluid (CSF) proteomics can provide additional insights into brain aging and neurodegeneration. By comprehensively examining 7,006 aptamers targeting 6,139 proteins in CSF obtained from 660 healthy individuals aged from 43 to 91 years old, we subsequently identified significant sex and aging effects on 5,097 aptamers in CSF. Many of these effects on CSF proteins had different magnitude or even opposite direction as those on plasma proteins, indicating distinctive CSF-specific signatures. Network analysis of these CSF proteins revealed not only modules associated with healthy aging but also modules showing sex differences. Through subsequent analyses, several modules were highlighted for their proteins implicated in specific diseases. Module 2 and 6 were enriched for many aging diseases including those in the circulatory systems, immune mechanisms, and neurodegeneration. Together, our findings fill a gap of current aging research and provide mechanistic understanding of proteomic changes in CSF during a healthy lifespan and insights for brain aging and diseases.

Introduction

Aging is a complex and gradual process involving both physiological and pathological changes characterized by progressive functional decline, resulting in decreased fitness, increased susceptibility, and vulnerability to diseases, and eventually death¹. Many chronic diseases, including obesity, diabetes, cardiovascular disease (CVD)², stroke, and neurodegenerative brain diseases,³ occur more frequently as aging, and multiple of these diseases often co-occur. These age-related diseases are the leading causes of mortality and increasing healthcare costs. With increased life expectancy, these costs are continually rising and represent an ever-growing challenge. In addition to aging, clear sex differences exist in many of these diseases and continue to present across the lifespan. Parkinson's disease (PD)⁴ and cardiovascular diseases⁵ are more prevalent in males, whereas Alzheimer's disease (AD) is more prevalent in females⁶. Other diseases including osteoporosis⁷ and multiple sclerosis⁸ also show the disparity between males and females. Therefore, understanding the mechanisms of aging and sex differences and their association with age-related diseases and resulting comorbidity conditions is essential for developing therapies that mitigate age-related disease and prolong healthy aging and longevity.⁹⁻¹²

Several studies using omics data including methylations, transcriptomics, and proteomics¹³ examined age-related changes, identifying a broad spectrum of biomarkers associated with the aging process and providing new insights into the biology of chronic diseases¹⁴. Compared to methylation and transcriptomic levels, proteins are final effectors of most physiological pathways, often dysregulated in diseases, and important drug targets. Most human proteomic studies have considered plasma due to easy accessibility^{13,15}. For example, Lehallier et al. identified dynamic changes in several plasma proteomes throughout the life span, demonstrating biological aging is a better predictor for morbidity and mortality than chronological aging^{16,17}. Koichiro et al. identified sex differences in serum protein composition¹⁸. Couillard et al. found that females have three times higher plasma leptin and weight-losing hormones than males¹⁹. These and several other plasma proteomic studies have successfully utilized plasma proteome profiling, elucidated the aging process, and provided quantitative and functional characterization of pathological conditions²⁰. Compared to plasma, cerebrospinal fluid (CSF) is adjacent to the brain and can better capture the aging process of brain and central nervous system. For most neurodegenerative diseases, CSF proteins are gold standard biomarkers and provide critical pathological changes occurring in the brain. As certain age-associated diseases including neurodegeneration demonstrate their pathological evidence through CSF, studying CSF aging signatures has immense potential in uncovering idiopathic conditions that cannot be discovered using other specimens. However, due to the perceived invasiveness of obtaining CSF samples via lumbar puncture, relatively few studies collect CSF samples from healthy individuals.

We hypothesize that the CSF proteome has different sex and aging signatures when compared to the plasma proteome, which in turn can uncover biological aging processes that may be specific to brain aging. To test this hypothesis, we generated 7,006 aptamers targeting 6,162 proteins in CSF from 994 cognitively normal individuals across three large cohorts (Table 1). We pursued an unbiased and hypothesis-free approach to examine the impact of aging and sex differences on CSF proteome in the discovery cohort and validated them in the two independent cohorts. We performed clustering analysis to group CSF proteins into modules, each with similar sex and aging trajectories. Because aging plays a significant role in the pathophysiology of age-related diseases, we subsequently examined these modules for disease and gene

ontology. Our findings show distinctive aging trajectories and sex differences in CSF proteome and their connection to common chronic diseases.

Results

Sex and aging effects of CSF proteomics are distinctive, compared to plasma

To examine the impact of aging and sex differences on CSF proteome, we collected CSF samples from 660 cognitively healthy individuals in the Knight Alzheimer Disease Research Center (Knight ADRC) and subsequently quantified proteomic data through aptamer-based SOMAscan platform (SOMAscan7k)²¹. After a rigorous data processing and quality control, 7,006 aptamers targeting 6,162 human proteins remained and were utilized in the study. By performing regression analysis in this cohort, we identified 4,682 proteins significantly associated with age (including NEFL, GDF15, and MYL4; Fig.1a; Supplementary Table 1) and 1,587 proteins significantly associated with sex at FDR < 0.05 (including PUDP, MYL4, PZP, and APCS; Fig. 1b; Supplementary Table 2). There were 1,172 proteins affected by both age and sex (Fig. 1c). Among the age associated proteins, 3,037 proteins (64.9%) including NEFL, GDF15, and CFD increased with age, while 1,645 proteins (35.1%) including NEU1 and ID2 decreased with age (Fig. 1d). Among the sex associated proteins, 831 proteins (52.4%; including MYL4 and APCS) had higher abundance levels in males, while 756 proteins (47.6%; including PUDP and PZP) had higher abundance levels in females (Fig. 1d). The remaining 1,909 proteins such as BCHE and GSTT1 were not affected by either age or sex (Fig. 1d).

To validate these age and sex effects on CSF proteomics, we obtained the same SOMAscan7k proteomic data in CSF from 334 individuals in the Alzheimer's Disease Neuroimaging Initiative (ADNI) and Fundació ACE Alzheimer Center Barcelona (FACE) cohorts, which were examined for aging effects and sex differences (Supplementary Tables 3-4). Out of the 4,682 age associated proteins in Knight ADRC, 1,942 proteins (41.5%) were nominally significant in this validation data at $P < 0.05$. Of the 1,587 sex associated proteins in Knight ADRC, 939 proteins (59.2%) were nominally significant in this data. When checked with all 7,006 analytes, correlation between the Knight ADRC and the validation data was very high (Pearson correlation $\rho=0.85$, $P < 2.0 \times 10^{-16}$, for aging effect, Fig 1c; $\rho=0.86$, $P < 2.0 \times 10^{-16}$, for sex difference; Fig. 1d), indicating the robustness of these identified aging effects and sex differences in CSF proteins.

To examine whether these findings would be specific to CSF or shared with plasma, we also obtained the SOMAscan7k proteomic data in plasma from 1,384 healthy individuals in Knight ADRC. In plasma data, we found 3,296 proteins associated with age (Supplementary Fig.1a; Supplementary Table 5) and 3,873 proteins associated with sex (Supplementary Fig.1b; Supplementary Table 6) at FDR < 0.05. There were 1,803 proteins associated with both age and sex (Supplementary Fig.1c). These age and sex effects on plasma proteome was also validated with the external UKB-PPP ($\rho=0.44$ for aging effect, Fig 1c; $\rho=0.52$ for sex difference; Supplementary Fig. 1d-e). These moderate correlation values were somewhat expected as the different proteomic platforms were used (SOMAscan in Knight ADRC vs. OLINK in UKB-PPP).

When these effects on plasma proteomics were compared to those on CSF proteomics, their correlations were much reduced ($\rho=0.39$ for aging effect, Fig 1c; $\rho=0.23$ for sex difference; Fig. 1g-h). The aging effects of 1,824 (55.3%) proteins were not consistently identified in CSF and plasma, with 693 (21.0%) showing the opposite directions between CSF and plasma. While 1,472 (44.7%) proteins (including GDF15; $\beta_A=0.011$ in CSF vs 0.0099 in plasma; Fig. 1g;

Supplementary Fig. 1f) had the same directions in both CSF and plasma, several of these proteins had pronounced magnitude differences in their aging effect. For example, PTN increased with age 13 times faster in plasma than in CSF ($\beta_A=0.0008$ in CSF vs 0.0105 in plasma; Fig. 1g; Supplementary Fig. 1f), while MYL4 increased with age seven times faster in CSF than in plasma ($\beta_A=0.0034$ in CSF vs 0.0005 in plasma). With sex effects, only 502 (13.0%) proteins (including CGA|FSHB; $\beta_S=0.389$ in CSF vs 0.713 in plasma; Supplementary Fig. 1f) had the same direction in CSF and plasma. For the remaining 3,371 (87.0%) proteins, sex effects were not consistent in CSF and plasma. For example, C3 in CSF had higher abundance in males, but C3 in plasma had higher abundance in females ($\beta_S=-0.125$ in CSF vs 0.112 in plasma; Supplementary Fig. 1f). In contrasts, LANCL1 in CSF had higher abundance in females, but LANCL1 in plasma had higher abundance in males ($\beta_S=0.038$ in CSF vs -0.054 in plasma; Fig. 1h; Supplementary Fig. 1f). These findings point that CSF proteome show distinctive age and sex differences compared to plasma proteome.

CSF biological age is consistent to plasma biological age

To examine the biological age using proteomics in CSF and in plasma, we constructed a proteomic clock based on the 569 Knight ADRC samples that had proteomic data in both CSF and plasma. Since we found that proteins showed sex differences, we constructed sex-stratified biological clocks. We split the data into testing vs training (3/4 vs 1/4) and subsequently selected the model based on four-fold cross-validation. In CSF from males, the optimal proteomic aging model included 120 proteins (Fig. 2a; Supplementary Table 7), which provided the biological age highly consistent with the chronological age (correlation $\rho=0.97$ in training data; $\rho=0.75$ in testing data; Fig. 2b). Similarly, in CSF from females, the biological age predicted based on 134 proteins was highly consistent with the chronological age ($\rho=0.97$ in training data; $\rho=0.87$ in testing data; Fig. 2b). In the independent ADNI and FACE cohorts, these sex-specific biological clocks based on CSF proteomics provided comparable performance (in ADNI, $\rho=0.72$ for males and $\rho=0.61$ for females; in FACE, $\rho=0.75$ for males and $\rho=0.61$ for females; Supplementary Fig. 2a). Compared to the CSF aging model, the proteomic aging model in plasma included fewer proteins (62 proteins in males and 89 proteins in females; Fig. 2a; Supplementary Table 7). The biological aging in plasma performed equally well ($\rho=0.92$ in training males; $\rho=0.86$ in testing males; $\rho=0.94$ in training females; $\rho=0.83$ in testing females; Fig. 2c). These plasma clocks were validated well in an additional set of 815 Knight-ADRC samples ($\rho=0.83$ for males and $\rho=0.81$ for females; Supplementary Fig. 2b).

To compare the predicted biological age based on CSF or plasma clocks, we examined 143 Knight ADRC samples in which plasma and CSF were collected within 6 months. Across all samples, there was no statistically significant difference in the predicted biological ages based on CSF or plasma clocks (paired t-test $P=0.915$). The correlations between predicted biological age based on CSF or plasma clocks was moderate ($\rho=0.61$; Fig. 2d). When the correlation was examined separately for each sex, males showed somewhat higher correlation than females ($\rho=0.67$ in males; $\rho=0.58$ in females).

To examine whether predicted biological age based on CSF or plasma clocks had a similar relationship with chronological age (i.e., delayed, or accelerated aging), we computed the delta (the departure of each predicted biological age from the chronological age) in CSF and plasma. The overall correlation of the difference between predicted age based on CSF and plasma clocks and chronological age was $\rho=0.620$ with almost no difference in males and females ($\rho=0.624$ in males; $\rho=0.632$ in females; Fig. 2e). For 103 individuals (72%), the differences between age predicted by both CSF and plasma clocks and chronological age were in the same direction (kappa statistic=0.44). Fifty-three individuals (23 males; 30 females) showed delayed

aging based on both the CSF and plasma clock, where predicted biological age was younger than the chronological age, while 50 samples (20 males; 30 females) showed accelerated aging with their biological age older than the chronological age. In the remaining 40 samples (28%; 17 males; 23 females), the differences between age predicted by both CSF and plasma clocks and chronological age were in different directions (e.g., the individual's CSF clock provided accelerated aging, while his/her plasma clock provided delayed aging). Together, these results show that CSF proteomic data generally provide consistent information as the plasma proteomic data in terms of individual's biological age.

Network analyses of CSF proteomics identify modules with distinctive age and sex signatures

To reduce the complexity of the proteome, we performed weighted gene co-expression network analysis (WGCNA) with 7,006 aptamers in the Knight ADRC CSF samples. The network analysis provided 16 modules, which collectively included a total of 5,131 aptamers targeting 4,917 proteins (Fig 3a; Supplementary Table 8). The remaining 1,875 aptamers were not assigned into any of the 16 modules. When we visually examined each module with locally estimated scatterplot smoothing (LOESS) in males and females separately, each module included proteins with similar aging or sex signatures (Fig. 3b; Supplementary Fig. 3). Among them, several modules exhibited non-linear age trajectories.

When each module was tested for aging effects, we found 13 modules significantly correlated with age (at $P < 0.05$; Fig. 3a; Supplementary Table 9). Correlations within each sex (in males and females separately) were like those in the combined data (Supplementary Table 9). For some modules (M3, M4, and M6), aging trajectories in males and females were parallel with similar patterns (for example, showing protein levels increasing or decreasing around the same age). Several modules (including M10 and M15) had complex aging trajectories. For example, females increasing their protein abundances around the age when males decreasing their proteins. Seven modules (M1, M3, M5, M6, M7, M13, and M14) were positively correlated with age, showing increasing protein abundances as aging (Fig. 3a). Among them, M7 (including *KCNE5*, *HSPB6*, and *MYL9*) had the highest positive correlation ($\rho = 0.50$, $P = 3.3 \times 10^{-42}$). Proteins in M7 showed a stronger linear aging trajectory in males, whereas proteins showed non-linear aging trajectory with increasing abundance until 75 years old and decreasing afterward. In M5, the male trajectory increased until 65 years old and remained constant afterward, whereas its female trajectory stayed constant until 65 years old and changed afterward. With strong non-linear trajectory, the correlation with age for M5 ($\rho = 0.082$; $P = 3.45 \times 10^{-02}$) did not capture full aging effect. Five modules (M2, M4, M9, M10, and M15) were negatively correlated with age, showing decreasing protein abundances as aging. M2 (including *NEU1* and *ID2*) had the lowest correlation, showing most decreases ($\rho = -0.37$, $P = 7.2 \times 10^{-23}$).

When each module was tested for sex differences, nine modules showed statistically significant sex differences at $P < 0.05$ (Fig. 3). In three modules (M3, M6, and M7), males had higher protein levels. Among them, M6 (including *APCS*, *ELMO1*, *PI3*, *PLG*) showed the most noticeable sex difference with the largest Cohen's distance (-0.70 , $P = 1.8 \times 10^{-17}$), whose aging trajectories in males and females were parallel. In six modules (M2, M4, M5, M10 and M11), females had higher protein levels. Among them, M2 (including *ENPP2*, *FIBP*, and *TNNI1*) showed the largest sex differences (Cohen's $d = 0.25$, $P = 1.51 \times 10^{-03}$), with non-linear aging trajectories parallel in males and females.

Enrichment analysis highlights modules linked to multiple aging diseases

To identify the protein modules relevant to common aging diseases, disease ontology enrichment was performed by examining enrichment for the diseases classified in the international classification of diseases' 10th revision (ICD-10). For the 1,145 diseases that were examined with each of 16 modules, 106 diseases were significantly enriched with eight modules (M2, M3, M5, M6, M7, M9, M13, and M15) at FDR < 0.05 (Fig. 4a; Supplementary Table 10). To uncover mechanistic insights into their relevance with diseases, we subsequently performed gene ontology (GO) enrichment for these eight modules. Six modules (M2, M3, M5, M6, M7, and M13) showed enrichments for GO biological processes (BP) at FDR < 0.05 (Supplementary Table 11). To examine cellular context of these modules and individual proteins, we also performed enrichment analysis using cell type expression data from human brain cells²² (Supplementary Table 12).

Six modules (M3, M5, M7, M9, M13 and M15) showed enrichment for three or fewer ICD-10 chapters. M3 showed enrichment for the complement deficiency disease in the blood and immune disease chapter (FDR = 2.42×10^{-02}), involving 11 proteins including *C7*, *C4A*, *C1S*, and *C3* (Supplementary Fig. 4a), mainly expressed in the microglia and astrocytes. M3 showed enrichment for 8 GO pathways including extracellular matrix organization (FDR = 2.55×10^{-02}) and complement activation (FDR = 2.55×10^{-02}), explaining its link to complement deficiency disease (Supplementary Fig. 4b). M5 was enriched for lysosomal storage diseases, mucopolysaccharidoses (FDR = 6.89×10^{-03}), involving seven proteins including *IDUA*, *NT5E*, *GNS* and *EGFR* (Supplementary Fig. 4c). They were enriched for 25 GO pathways including synapse assembly (FDR = 7.09×10^{-03}) and the lysosomal protein catabolic process (FDR = 2.67×10^{-02} ; Supplementary Fig. 4d). M7 was enriched for the keloid skin disease (FDR = 8.79×10^{-03}) involving 15 proteins including *TAGLN*, *SERPINE1*, and *FLNA* (Supplementary Fig. 4e). They showed enrichment for GO pathways in actin cytoskeleton organization (FDR = 6.88×10^{-06}), which regulates several essential wound-healing components²³, thereby explaining its link to skin diseases (Supplementary Fig. 4f). M9 was enriched for the musculoskeletal system (spondylitis, FDR = 4.45×10^{-03}), the digestive system (recurrent aphthous ulcer, FDR = 0.02), and eye diseases (keratoconjunctivitis, FDR = 0.026), involving multiple proteins including *IFNG*, *PLG* and *IL23R* (Supplementary Fig. 4g). M13 was enriched for enterovirus infections (FDR = 1.55×10^{-02}) involving eight proteins (*VIM*, *STAT1*, and *G6PD*; Supplementary Fig. 4h). They were enriched for GO pathways involved in translation (FDR = 2.84×10^{-02}) and stress granule assembly (FDR = 4.99×10^{-02} ; Supplementary Fig. 4i). Stress granules formed from mRNAs modulates the stress response and viral infection²⁴, explaining its link to infections and parasitic diseases. M15 was enriched for dilated cardiomyopathy (FDR = 1.05×10^{-02}), with ten proteins including *CASP8*, *GCH1*, *TNNT2*, and *ADCYAP1*, mainly expressed in the microglia and neuron (Supplementary Fig. 4j).

Two modules are linked to neurodegeneration and multiple aging diseases

We found that modules M2 and M6 were enriched for multiple ICD-10 chapters, suggesting their relevance to comorbidity conditions across multiple aging diseases (Supplementary Table 10).

M2 showed significant enrichment for 32 diseases spanning 13 ICD-10 chapters at FDR < 0.05 (Fig. 4a). 43 proteins including *APOE*, *APP*, and *IL6* were involved with mental disorders (multi-infarct dementia, FDR = 0.029) and diseases in the nervous system (transient ischemic attack, FDR = 2.96×10^{-02} ; sleep apnea syndrome, FDR = 3.44×10^{-02} ; Fig 4b). Enrichment for blood and immune mechanism was found with 52 proteins including *JAK2*, *CCL5*, and *IL1B*; eosinophilia, FDR = 0.029; blood coagulation disorders, FDR = 4.20×10^{-02} ; Fig; 5c). Enrichment for circulatory system was notable (81 proteins including *APOE*, *AGT*, *SERPINA3*; subarachnoid

hemorrhage, FDR = 2.96×10^{-02} ; cerebral amyloid angiopathy, FDR = 2.96×10^{-02} ; Supplementary Fig. 5a). M2 was the only module that was enriched for neoplasms (acute megakaryocytic leukemias, FDR = 8.74×10^{-03} ; prostate cancer, FDR = 1.52×10^{-02}), involving 259 proteins including CD36 and ANGPTL4 (Supplementary Fig. 5b). Proteins involved with the additional six ICD-10 chapters were presented in Supplementary Fig. 6. Proteins in M2 were enriched for microglia and macrophage (57 proteins including NFKB, C3, and IL1B, enrichment P = 1.34×10^{-02} , Fig. 4e).

For 384 GO pathways, belonging to the 25 higher-level representative terms, M2 showed enrichment (Fig. 4f). Pathways related to neurodegeneration and immune mechanisms include cellular response to peptide (response to stress, FDR = 2.63×10^{-04} ; neuron apoptotic process, FDR = 2.19×10^{-02}) and negative regulation of apoptotic process (FDR = 3.16×10^{-03}). Several pathways related to neoplasm were found, including apoptosis, leukocyte activation (FDR = 2.54×10^{-03}), cell cycles and cell death (necrosis), RNA splicing (FDR = 1.08×10^{-07}), and RNA localization (FDR = 1.20×10^{-02}). Cancer cells exploit RNA splicing to promote tumor growth²⁵⁻²⁸. When tumor cells avoid apoptosis, regulated cell death is activated during tumor cell proliferation in defense tumor progression and migration^{29,30}. Adhesion cascade leads to localization of inflammatory cells to tumor sites including the recruitment of effector leukocytes³¹. In addition, pathways involved leukocyte activation explain this module's relevance to peripheral T-cell lymphoma and other types of leukemia.

Among all 16 modules, M6 showed enrichment for the largest number of diseases (70 belonging to 14 ICD-10 chapters) with the highest statistical significance (Fig. 4a; Fig. 5a). In M6, 61 proteins including LRRK2, SERPINA3, and C3 were involved with mental disorders (mild cognitive disorder, FDR = 8.37×10^{-03} ; schizophrenia, FDR = 2.51×10^{-02}) and nervous system (multiple sclerosis, FDR = 1.81×10^{-02} ; late onset AD, FDR = 5.02×10^{-02} ; Fig. 5b). The most significant enrichment was found for blood and immune mechanism (complement deficiency disease, FDR = 2.14×10^{-09} ; blood coagulation disorders, FDR = 7.96×10^{-08}), involving 68 proteins including PLG, SELE, SERPINA3 (Fig. 5c). In addition, the circulatory system (cardiovascular diseases, FDR = 1.16×10^{-08} ; atherosclerosis, FDR = 4.81×10^{-07} ; Supplementary Fig. 7a), infections (septicemia, FDR = 1.65×10^{-06} ; Supplementary Fig. 7b), metabolic diseases (diabetes mellitus, FDR = 1.04×10^{-05} ; Supplementary Fig. 7c) were strongly enriched for this module. Proteins involved with the additional five ICD-10 chapters were presented in Supplementary Fig. 8.

Among all 16 modules, M6 also had the highest statistical significance for GO pathways, showing enrichment for 126 GO biological processes belonging to the 18 higher-level representative terms at FDR < 0.05 (Fig. 5d). The strongest enrichment was involved with humoral immune response (FDR = 2.14×10^{-19} ; complement activation, FDR = 2.14×10^{-19}). In particular, complement activation plays a defensive role in our body through proteolytic cascades and opsonization³², explaining a direct link to complement deficiency disease. The second strongest enrichment involved blood coagulation (FDR = 5.64×10^{-10} ; coagulation, FDR = 1.189×10^{-10} ; defense response, FDR = 2.47×10^{-08}), which explains the link to blood coagulation disorders of the blood and immune system. Other pathways included phagocytosis recognition (FDR = 6.13×10^{-05}), zymogen activation (FDR = 5.73×10^{-07}), and regulation of coagulation (FDR = 5.89×10^{-08}). Proteins in M6 were enriched for endothelial cells (15 proteins including APOL1 and SELE; enrichment P = 1.41×10^{-02} , Fig. 5e).

Discussion

In this study, we systematically investigated the aging effects and sex differences on CSF proteomics in a large scale. CSF protein levels can be influenced only by factors or diseases that influence the brain and the central nervous system, showing more direct relevance to brain aging and diseases. By comprehensively examining 7,006 aptamers targeting 6,139 proteins in CSF obtained from 660 healthy individuals aged from 43 to 91 years old, we subsequently identified significant sex and aging effects on 5,097 aptamers in CSF. They were confirmed by two additional independent cohorts. We found that many of the sex and aging effects on CSF proteins had different magnitude or even opposite direction, when compared to their corresponding effects on plasma proteins, indicating distinctive CSF-specific signatures. Clustering of these CSF proteins revealed not only the modules with clear aging trajectories but also the modules with sex differences in their protein abundance. We subsequently identified the relevance for complex diseases for the eight modules, out of which two modules (M2 and M6) showed striking connection to neuropsychiatric diseases and many aging diseases.

In this study, we identified significant age-associated increase in GDF15, NEFH, NEFL, and IL6. The growth differentiation factor 15 (GDF15) is one of the transforming growth factor-beta cytokine superfamily essential in regulating cellular response to stress signals. Elevated GDF15 levels are reported to be associated with pathological conditions involving inflammation, myocardial ischemia, and cancer^{33,34}. Increase of GDF15 with aging has been reported in several studies¹⁶, including their link to mitochondrial function decline with aging³⁵. Neurofilament heavy polypeptide (NEFH) is one of type IV intermediate cytoplasmic filament proteins in neurons³⁶, which provide structural support for the axon and regulate axon diameter, thus maintaining the transmission of electrical impulses along axons³⁷. Neurofilament light polypeptide (NEFL) is a neuronal cytoplasmic protein from the member of the intermediate filament protein family, highly expressed in myelinated axons. It increases in CSF and blood along with axonal damage in neurological disorders including inflammatory, neurodegenerative, traumatic and cerebrovascular diseases³⁸⁻⁴⁰. Age-associated increase of NEFL has been previously reported⁴¹⁻⁴³. Interleukin 6 (IL6), a multifunctional cytokine, is a major contributor to acute phase inflammatory response. While it is normally present in low levels, IL6 is increased in aging and with chronic age-related diseases including neurodegeneration, diabetes, rheumatoid arthritis, cancer, atherosclerosis, and infections⁴⁴⁻⁴⁶. In addition, this study confirmed several of 82 age-associated proteins (including PTN, MB, and CST3) identified by Baird et al⁴⁷. Johnson et al in his systemic review compiled 32 age-associated proteins that were consistently reported across multiple studies¹⁶. Out of these 32 proteins, our study confirmed the significant aging effects of 23 proteins including GDF15, ANXA1, C3, C4A, and EGFR^{16,17,48-52} in both CSF and plasma.

This study found the strong sex difference in APOE, SERPINA3, C3, PLG, SERFINF2, SELE, APCS, which are known to be critical for multiple diseases including neurodegeneration, cardiovascular disease (CVD), kidney failures, and diabetes. Apolipoprotein E (APOE), a carrier protein for cholesterol and lipids, is produced in abundance in the brain and serves as the principal lipid transport vehicle in CSF⁵³. It is critically involved in the pathogenesis of AD and CVD, through induction at high concentration in peripheral nerve injury and repair by redistribution of lipids to regenerating axons and to Schwann cells during remyelination. The APOE gene alleles modulate human aging⁵³⁻⁵⁶. SERPINA3, a part of protease inhibitors, is critically involved in the anti-inflammatory response and antiviral responses⁵⁷, and identified as a specific biomarker of delirium and AD^{58,59}. Complement component 3 (C3) is the central point of the three-cascade activation pathway of the complement system. The complement plays a role in inflammatory processes, metabolism, apoptosis, mitochondrial function, Wnt signaling

pathways, and plays a significant role in aging-related diseases, including AD, macular degeneration, and osteoarthritis^{60,61}. Plasminogen (PLG) is the zymogen of plasmin, a broad specificity serine protease whose activity contributes to pathological conditions including cardiovascular diseases, infections, metabolic disorders, asthma, kidney disorders, and musculoskeletal diseases⁶²⁻⁶⁸. SERPINF2 is a serine protease inhibitor of plasmin, modulates insulin sensitivity, and is associated with CVD, blood coagulation disorders, rheumatoid arthritis, asthma, and diabetes⁶⁹⁻⁷¹. Soluble endothelial leukocyte adhesion molecule-1 (SELE), a part of cell adhesion molecules expressed only on endothelial cells by cytokines, is a pro-inflammatory protein important in cell-cell/cell-matrix interactions during immune responses and inflammatory process⁷². Elevated levels are associated with CVDs, diabetes, cerebrovascular diseases and rheumatoid arthritis⁷³⁻⁷⁶. Serum amyloid P component (APCS) is a soluble pattern-recognition protein, binding to damaged membranes, nuclear autoantigens, and determinants on microbial pathogens and recruiting other elements of host defense⁷⁷. APCS is involved in the regulation of matrix formation, regulation of complement activation and an acute phase response protein in response to infection and inflammatory cytokines during the innate immune response⁷⁸.

Multiple studies including Lehallier et al. clustered plasma proteins into modules and found non-linear changes throughout the lifespan, which may be important to fully appreciate the undulating changes¹⁷. In this study, we examined the clustering of the CSF proteins into modules and identified not only the modules with often non-linear aging trajectories but also the modules with clear sex differences. Module 6 showed the strongest sex differences and the most extensive enrichment for multiple aging diseases including neurodegeneration. Several studies demonstrated the immune system's role in contribution to diverse pathological conditions, which aligns with our findings. Immune system functions have been associated with neurodegeneration including AD. Gate et al. revealed the adaptive immune response in AD⁷⁹ and demonstrated the T-cell involvement in Lewy body dementia⁸⁰. In addition, we found strong connection of two module (M2 and M6) to complement deficiency disease. Complement, a group of over 30 proteins involved in the complement pathway or cascade, is critical in the inflammation response of the innate immune system against micro-organisms and infections⁸¹⁻⁸⁶. The complement system orchestrates opsonization, facilitate cytotoxic destruction, formulate membrane attack complexes (MAC), and the liberation of peptides that promote inflammatory responses⁸⁶. Complement deficiencies arise when any of these proteins are missing from the cascade or fail to function properly, which results in increase prevalence in infections by pyogenic organisms via phagocytosis^{81,86}. Both modules showed enrichment of the related biological processes involved in translation, phagocytosis, stress granule assembly, and cell-cell adhesion⁸⁷⁻⁹⁶.

Several clocks based on transcriptomics, methylations, or proteomics in plasma provided insights for biological aging and their relevance to the diseases. These studies demonstrated that biological aging is profoundly involved in the pathogenesis of many diseases and age-related health conditions. In this study, we constructed sex-specific proteomic clock in both CSF and plasma. While an individual protein may change differently in CSF and plasma as aging, the performance of the CSF or plasma-based clocks was similar in terms of the correlation with the chronological age. Furthermore, they also provided similar inference on whether an individual has delayed or accelerated aging. These results indicate that brain aging may be like the overall biological aging based on circulating plasma proteomics.

Our study has some limitations. First, while this study has the largest sample size to examine aging and sex differences in CSF proteomics from healthy individuals as of today, it is based on a cross-sectional study. Therefore, a longitudinal follow-up study may be needed to validate the findings from this study. Second, while the aging and sex effects of CSF proteins were often

different from those of plasma proteins, the sex specific proteomic clock in these biofluid provided similar performance. This may be because this study examined only healthy individuals, who often exhibit a less pronounced degree of accelerated aging, when compared to individuals with diseases. An additional follow-up will be needed to distinguish the organ-specific biological aging. Third, to examine the relevance of CSF proteins to biological processes and diseases, we employed statistical and computational approaches. Additional follow-up in the select proteins will be needed for their clinical utility.

In summary, we comprehensively examined age effects and sex differences in over 6,000 CSF proteins, identifying 4,682 age-associated and 1,587 sex-different proteins. These age effects and sex differences were distinctive of CSF, when compared to plasma. Many of these proteins were involved in multiple diseases including neurodegeneration and cardiovascular diseases. Our findings based on CSF proteomics fill a gap of current aging research and help understanding the role of age and sex in CSF protein regulation during a healthy lifespan. They may help developing personalized treatment for aging interventions and clinically translatable interventions for brain aging and diseases.

Methods

Study populations

This study examined 1,807 cognitively normal individuals from the Knight Alzheimer Disease Research Center (Knight-ADRC)⁹⁷⁻⁹⁹, Alzheimer's Disease Neuroimaging Initiative (ADNI)¹⁰⁰, and Fundació ACE (FACE)¹⁰¹ studies (Table 1). The age of these individuals spanned from 43 to 91 years old, representing middle and elderly adults. The Knight-ADRC at Washington University in St. Louis aims to advance AD research with the goal of treatment or prevention of AD. The ADNI is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD. Headquartered in Barcelona, FACE has collected and analyzed almost 18,000 genetic samples, diagnosed over 8,000 patients. This study was approved by the Institutional Review Boards of the Washington University School of Medicine in St. Louis, and the research was performed in accordance with the approved protocols.

Proteomics data

Cerebrospinal fluid (CSF) samples from the Knight-ADRC, ADNI, and FACE were collected using lumbar puncture (LP) in the morning after an overnight fast. All samples underwent the identical protocols for preparation and processing, and stored at -80 °C. All samples were randomized across plates to avoid batch effects and were sent together to SomaLogic. Protein levels were quantified using the SOMAscan platform based on a multiplexed, single-stranded DNA aptamer assay. The quantitative levels of 7,584 aptamers were reported as relative fluorescence unit (RFU). Initial data normalization was performed by SomaLogic using hybridization controls for intra-plate and median signal to account for inter-plate variances¹⁰². Normalization against an external reference to control for biological variances was also performed by SomaLogic¹⁰³. In-house quality control (QC) was also performed to exclude outliers. More details of the QC process were published elsewhere^{103,104}. After processing, CSF proteomic data consisted of 7,006 aptamers targeting 6,139 proteins in CSF from 994 individuals. Plasma proteomic data consisted of 6,905 aptamers targeting 6,106 proteins in plasma from 1,372 samples.

Age effects and sex differences in CSF and plasma proteomes

To examine the effect of age and sex on CSF protein abundance, we performed regression analysis using the following model:

$$\log_{10}(\text{Protein abundance}) = \beta_0 + \beta_A \text{Age} + \beta_S \text{Sex} + \beta_1 \text{PlateID} + \beta_2 \text{Cohort}$$

Our primary interest is the age effect β_A and the sex effect β_S on \log_{10} -transformed protein abundance. Age is an individual's age at LP; sex is coded as 1 for a female and 0 for a male; PlateID is a plate identifier for proteomic sample; and Cohort is a cohort information (Knight-ADRC, ADNI, FACE). Analysis was performed first using the Knight-ADRC data, where a cohort was excluded in the model. Additional analysis was performed using the combined set of ADNI and FACE data. To control for multiple testing, the false discovery rate (FDR) < 0.05 based on the FDR-BH method was used.¹⁰⁵ In addition, to compare the effect of sex and age on CSF proteomics with those of plasma proteome, the same analysis was performed using the Knight-ADRC plasma proteomic data.

Proteomics based biological clock

We constructed sex-stratified predictive models for biological age using proteomic measurements with the least absolute shrinkage and selection operator (LASSO) regression.

Four models were constructed: aging clocks for males and females, each separately with CSF proteomics and plasma proteomics. To compare the predictive models between CSF and plasma proteomic data, we considered 569 individuals that had both CSF and plasma samples. We used log transformed RFU values of 6,738 analytes in both CSF and plasma proteomic data. We split the data into training data (3/4 with 177 males and 249 females) and testing data (the remaining 1/4 with 60 males and 83 females). The R package `glmnet`¹⁰⁶ was used for LASSO regression analysis. The shrinkage parameter that controls the model complexity was determined using 4-fold cross-validation on the training data. To examine model performance, we examined Pearson correlation coefficient and root mean squared error (RMSE) between the predicted ages and the actual chronological age.

Network analyses of CSF proteomics

To cluster CSF proteins into modules, we performed the weighted gene co-expression network analysis (WGCNA)¹⁰⁷. Specifically, we utilized the *blockwiseModules* function with the following parameters: `power=7`, `corType="bicor"`, `networkType="signed"`, `deepSplit=4`, `reassignThreshold=0.05`, and `mergeCutHeight=0.15`. After adjusting for covariates, the soft threshold power was determined as 7 by considering both the scale-free topology model fit ($R^2 = 0.9595$) and mean connectivity. We used `bicor`, a median-based measure of similarity that is more robust to outliers than the Pearson or Spearman correlation coefficient¹⁰⁸. A signed network was considered to account for both positive and negative correlations. The topological overlap matrix (TOM) was first calculated to determine the interconnection within the network structure. Hierarchical clustering analysis was conducted based on the calculated 1-TOM values to classify the proteins into distinct modules. The final modules were defined by grouping and rearranging similar groups using the dynamic tree cut method¹⁰⁹. For each module, we calculated the first eigenvector, known as the eigen protein. For the aging effect in each module, we considered the Pearson correlation coefficient between the eigen protein and age. For the sex differences in each module, we considered Cohen's distance¹¹⁰ between male eigen proteins and female eigen proteins.

Sex-specific aging trajectory

To visualize aging trajectories of each protein, we used the locally estimated scatterplot smoothing (LOESS)¹¹¹ method in R. To examine sex differences, we drew LOESS lines for males and females, separately. In addition, to examine the sex-stratified aging trajectories in each module, we obtained the average line of LOESS lines in males and females, separately, of all the proteins belonging to that module. In addition, to determine whether these sex-stratified aging trajectories in the Knight-ADRC data were reproducible, we examined proteomic data from two independent data (ADNI and FACE) in the same manner. Specifically, we created a grid point for age and calculated the Pearson correlation coefficients of the LOESS values at these points between the two datasets. We performed a one-sided correlation test at the FDR < 0.05 using the FDR-BH method¹⁰⁵. Only those validated proteins were included in the subsequent enrichment analysis.

Enrichment analysis

For disease ontology, we performed enrichment analysis using the `enrichDGN` function in the R package `DOSE`¹¹². The Entrez gene names for the validated proteins in each module were considered. To convert the CUI¹¹³ codes used by DisGeNET into the international classification of disease, 10th revision (ICD-10) codes developed by World Health Organization, we used the Python package `Owlready2` (ref¹¹⁴). ICD-10 codes¹¹⁵ are available at <https://icd.who.int/browse10/2019/en>. Using the first three digits of ICD-10 codes, 1,145

diseases were grouped into 22 chapters. Cases without a matching ICD code were excluded from the analysis.

For Gene Ontology (GO) biological processes, we performed enrichment analysis using enrichGO function in the R package clusterProfiler¹¹⁶. Significant GO terms were visualized into the hierarchically related GO pathways based on their similarity to distill them into common higher-level representative terms using REVIGO¹¹⁷. GO IDs were restricted to Homo sapiens species and visualization was created with treemaps.

To obtain cellular context of these modules and individual proteins, we also performed enrichment analysis using cell type expression data from human brain cells²² (Supplementary Table 12). Based on the cell type data from gene expression data from human astrocytes, neurons, oligodendrocytes, microglia/macrophages, and endothelial cells, we obtained the degree of specificity to relevant cell types for the proteins. For 5,750 proteins, these cell-type expression data were available. Fisher's exact test was used to examine cell-type enrichment of each module against the cell-types for 5,750 proteins.

Declarations

Ethics approval and consent to participate.

Ethics approval for every individual cohort was obtained from the respective Institutional Review Boards and research was carried out in accordance with the approved protocols (WUSTL IRB approval 201109148). Written informed consent was obtained from participants or their family members and all participating institutions approved the study design.

Consent for publication

All authors have approved the contents of this manuscript and provided consent for publication.

Availability of data and materials

All data generated in this study are included in this published article.

Competing interests

CC has received research support from: GSK and Eisai. The funders of the study had no role in the collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. CC is a member of the advisory board of Circular Genomics and owns stocks.

Funding and Acknowledgements

We thank all the participants and their families, as well as the involved cohorts, institutions, and their staff.

This work was supported by grants from the National Institutes of Health, RF1 AG074007 (YJS), R01 AG044546 (CC), P01 AG003991 (CC), RF1 AG053303 (CC), RF1 AG058501 (CC), U01 AG058922 (CC), P30 AG066444 (JCM), P01 AG003991 (JCM), P01 AG026276 (JCM), and the Chan Zuckerberg Initiative (CZI), the Michael J. Fox Foundation (CC), and the Alzheimer's Association Zenith Fellows Award (ZEN-22-848604, awarded to CC).

ADNI acknowledgement: Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in

Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Author's contributions

DS, CML and CA performed analysis, interpreted the results, and drafted the manuscript. GH contributed to the manuscript preparation. JT performed proteomics data processing and quality control. PK acquired Knight ADRC samples and data. MB, AO, MVF and AG acquired phenotypes and CSF samples in FACE. JCM and SES obtained funding, recruited Knight ADRC cohort, and curated phenotype data. TSP obtained funding, supervised the work, and interpreted the results. YJS and CC obtained funding to generate proteomic data, designed the study, supervised the work, interpreted the results, and drafted the manuscript. All authors read and approved the manuscript.

Table 1. The sex and aging distribution of the individuals included in this study.

Healthy individuals	CSF			Plasma
	Knight-ADRC	ADNI	FACE	Knight-ADRC
N	660	166	168	1,382
Males	280 (42%)	89 (54%)	83 (49%)	565 (41%)
Females	380 (58%)	77 (46%)	85 (51%)	817 (59%)
Age (mean, range)	70 (43, 91)	75 (60, 90)	66 (48, 86)	73 (44, 91)

References

- 1 Zierer, J. *et al.* Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model. *Sci Rep* **6**, 37646 (2016). <https://doi.org/10.1038/srep37646>
- 2 Camici, G. G. & Liberale, L. Aging: the next cardiovascular disease? *Eur Heart J* **38**, 1621-1623 (2017). <https://doi.org/10.1093/eurheartj/ehx239>
- 3 Xia, X., Jiang, Q., McDermott, J. & Han, J. J. Aging and Alzheimer's disease: Comparison and associations from molecular to system level. *Aging Cell* **17**, e12802 (2018). <https://doi.org/10.1111/accel.12802>
- 4 Wooten, G. F., Currie, L. J., Bovbjerg, V. E., Lee, J. K. & Patrie, J. Are men at greater risk for Parkinson's disease than women? *J Neurol Neurosurg Psychiatry* **75**, 637-639 (2004). <https://doi.org/10.1136/jnnp.2003.020982>
- 5 Kander, M. C., Cui, Y. & Liu, Z. Gender difference in oxidative stress: a new look at the mechanisms for cardiovascular diseases. *J Cell Mol Med* **21**, 1024-1032 (2017). <https://doi.org/10.1111/jcmm.13038>
- 6 Li, R. & Singh, M. Sex differences in cognitive impairment and Alzheimer's disease. *Front Neuroendocrinol* **35**, 385-403 (2014). <https://doi.org/10.1016/j.yfrne.2014.01.002>
- 7 Kenny, A. M. & Prestwood, K. M. Osteoporosis. Pathogenesis, diagnosis, and treatment in older adults. *Rheum Dis Clin North Am* **26**, 569-591 (2000). [https://doi.org/10.1016/s0889-857x\(05\)70157-5](https://doi.org/10.1016/s0889-857x(05)70157-5)
- 8 Orton, S. M. *et al.* Sex ratio of multiple sclerosis in Canada: a longitudinal study. *Lancet Neurol* **5**, 932-936 (2006). [https://doi.org/10.1016/S1474-4422\(06\)70581-6](https://doi.org/10.1016/S1474-4422(06)70581-6)
- 9 Benetos, A. *et al.* Influence of age, risk factors, and cardiovascular and renal disease on arterial stiffness: clinical applications. *Am J Hypertens* **15**, 1101-1108 (2002). [https://doi.org/10.1016/s0895-7061\(02\)03029-7](https://doi.org/10.1016/s0895-7061(02)03029-7)
- 10 Owsley, C. Aging and vision. *Vision Res* **51**, 1610-1622 (2011). <https://doi.org/10.1016/j.visres.2010.10.020>
- 11 Harman, D. The aging process: major risk factor for disease and death. *Proc Natl Acad Sci U S A* **88**, 5360-5363 (1991). <https://doi.org/10.1073/pnas.88.12.5360>
- 12 Harman, D. Aging: overview. *Ann N Y Acad Sci* **928**, 1-21 (2001). <https://doi.org/10.1111/j.1749-6632.2001.tb05631.x>
- 13 Moaddel, R. *et al.* Proteomics in aging research: A roadmap to clinical, translational research. *Aging Cell* **20**, e13325 (2021). <https://doi.org/10.1111/accel.13325>
- 14 Chambers, G., Lawrie, L., Cash, P. & Murray, G. I. Proteomics: a new approach to the study of disease. *J Pathol* **192**, 280-288 (2000). [https://doi.org/10.1002/1096-9896\(200011\)192:3<280::AID-PATH748>3.0.CO;2-L](https://doi.org/10.1002/1096-9896(200011)192:3<280::AID-PATH748>3.0.CO;2-L)
- 15 Maurya, P., Meleady, P., Dowling, P. & Clynes, M. Proteomic approaches for serum biomarker discovery in cancer. *Anticancer Res* **27**, 1247-1255 (2007).
- 16 Johnson, A. A., Shokhirev, M. N., Wyss-Coray, T. & Lehallier, B. Systematic review and analysis of human proteomics aging studies unveils a novel proteomic aging clock and identifies key processes that change with age. *Ageing Res Rev* **60**, 101070 (2020). <https://doi.org/10.1016/j.arr.2020.101070>
- 17 Lehallier, B. *et al.* Undulating changes in human plasma proteome profiles across the lifespan. *Nat Med* **25**, 1843-1850 (2019). <https://doi.org/10.1038/s41591-019-0673-2>
- 18 Miike, K. *et al.* Proteome profiling reveals gender differences in the composition of human serum. *Proteomics* **10**, 2678-2691 (2010). <https://doi.org/10.1002/pmic.200900496>
- 19 Couillard, C. *et al.* Plasma leptin concentrations: gender differences and associations with metabolic risk factors for cardiovascular disease. *Diabetologia* **40**, 1178-1184 (1997). <https://doi.org/10.1007/s001250050804>

- 20 Kopkova, A. *et al.* Cerebrospinal Fluid MicroRNA Signatures as Diagnostic Biomarkers
in Brain Tumors. *Cancers (Basel)* **11** (2019). <https://doi.org/10.3390/cancers11101546>
- 21 SOMAscan® v4.1 Data Standardization and File Specification Technical Note. (2021).
- 22 Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human
Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89**,
37-53 (2016). <https://doi.org/10.1016/j.neuron.2015.11.013>
- 23 Kopecki, Z. & Cowin, A. J. The role of actin remodelling proteins in wound healing and
tissue regeneration. *Wound Healing—New Insights into Ancient Challenge. IntechOpen*,
133-154 (2016).
- 24 Protter, D. S. & Parker, R. Principles and properties of stress granules. *Trends in cell
biology* **26**, 668-679 (2016).
- 25 Wang, E. & Aifantis, I. RNA Splicing and Cancer. *Trends Cancer* **6**, 631-644 (2020).
<https://doi.org/10.1016/j.trecan.2020.04.011>
- 26 Hershberger, C. E. *et al.* Complex landscape of alternative splicing in myeloid
neoplasms. *Leukemia* **35**, 1108-1120 (2021). <https://doi.org/10.1038/s41375-020-1002-y>
- 27 Chen, S., Benbarche, S. & Abdel-Wahab, O. Splicing factor mutations in hematologic
malignancies. *Blood* **138**, 599-612 (2021). <https://doi.org/10.1182/blood.2019004260>
- 28 Liu, S. & Cheng, C. Alternative RNA splicing and cancer. *Wiley Interdiscip Rev RNA* **4**,
547-566 (2013). <https://doi.org/10.1002/wrna.1178>
- 29 Tong, X. *et al.* Targeting cell death pathways for cancer therapy: recent developments in
necroptosis, pyroptosis, ferroptosis, and cuproptosis research. *J Hematol Oncol* **15**, 174
(2022). <https://doi.org/10.1186/s13045-022-01392-3>
- 30 Peng, F. *et al.* Regulated cell death (RCD) in cancer: key pathways and targeted
therapies. *Signal Transduct Target Ther* **7**, 286 (2022). <https://doi.org/10.1038/s41392-022-01110-y>
- 31 Carlos, T. M. Leukocyte recruitment at sites of tumor: dissonant orchestration. *J Leukoc
Biol* **70**, 171-184 (2001).
- 32 Dunkelberger, J. R. & Song, W.-C. Complement and its role in innate and adaptive
immune responses. *Cell research* **20**, 34-50 (2010).
- 33 Dominguez-Rodriguez, A., Abreu-Gonzalez, P. & Avanzas, P. Relation of growth-
differentiation factor 15 to left ventricular remodeling in ST-segment elevation myocardial
infarction. *Am J Cardiol* **108**, 955-958 (2011).
<https://doi.org/10.1016/j.amjcard.2011.05.028>
- 34 Wischhusen, J., Melero, I. & Fridman, W. H. Growth/Differentiation Factor-15 (GDF-15):
From Biomarker to Novel Targetable Immune Checkpoint. *Front Immunol* **11**, 951 (2020).
<https://doi.org/10.3389/fimmu.2020.00951>
- 35 Choi, S. *et al.* 31P Magnetic Resonance Spectroscopy Assessment of Muscle
Bioenergetics as a Predictor of Gait Speed in the Baltimore Longitudinal Study of Aging.
J Gerontol A Biol Sci Med Sci **71**, 1638-1645 (2016).
<https://doi.org/10.1093/gerona/glw059>
- 36 Dubrowskaja, N. *et al.* Neurofilament Heavy polypeptide CpG island methylation
associates with prognosis of renal cell carcinoma and prediction of antivasular
endothelial growth factor therapy response. *Cancer Med* **3**, 300-309 (2014).
<https://doi.org/10.1002/cam4.181>
- 37 Wang, J. *et al.* Neurofilament heavy polypeptide protects against reduction in
synaptopodin expression and prevents podocyte detachment. *Sci Rep* **8**, 17157 (2018).
<https://doi.org/10.1038/s41598-018-35465-6>
- 38 Gaetani, L. *et al.* Neurofilament light chain as a biomarker in neurological disorders. *J
Neurol Neurosurg Psychiatry* **90**, 870-881 (2019). <https://doi.org/10.1136/jnnp-2018-320106>

- 39 Yuan, A. & Nixon, R. A. Neurofilament Proteins as Biomarkers to Monitor Neurological Diseases and the Efficacy of Therapies. *Front Neurosci* **15**, 689938 (2021). <https://doi.org/10.3389/fnins.2021.689938>
- 40 Byrne, L. M. *et al.* Neurofilament light protein in blood as a potential biomarker of neurodegeneration in Huntington's disease: a retrospective cohort analysis. *Lancet Neurol* **16**, 601-609 (2017). [https://doi.org/10.1016/S1474-4422\(17\)30124-2](https://doi.org/10.1016/S1474-4422(17)30124-2)
- 41 Yilmaz, A. *et al.* Neurofilament light chain protein as a marker of neuronal injury: review of its use in HIV-1 infection and reference values for HIV-negative controls. *Expert Rev Mol Diagn* **17**, 761-770 (2017). <https://doi.org/10.1080/14737159.2017.1341313>
- 42 Lleo, A. *et al.* Longitudinal cerebrospinal fluid biomarker trajectories along the Alzheimer's disease continuum in the BIOMARKAPD study. *Alzheimers Dement* **15**, 742-753 (2019). <https://doi.org/10.1016/j.jalz.2019.01.015>
- 43 Khalil, M. *et al.* Serum neurofilament light levels in normal aging and their association with morphologic brain changes. *Nat Commun* **11**, 812 (2020). <https://doi.org/10.1038/s41467-020-14612-6>
- 44 Rea, I. M. *et al.* Age and Age-Related Diseases: Role of Inflammation Triggers and Cytokines. *Front Immunol* **9**, 586 (2018). <https://doi.org/10.3389/fimmu.2018.00586>
- 45 Maggio, M., Guralnik, J. M., Longo, D. L. & Ferrucci, L. Interleukin-6 in aging and chronic disease: a magnificent pathway. *J Gerontol A Biol Sci Med Sci* **61**, 575-584 (2006). <https://doi.org/10.1093/gerona/61.6.575>
- 46 Ershler, W. B. *et al.* Interleukin-6 and aging: blood levels and mononuclear cell production increase with advancing age and in vitro production is modifiable by dietary restriction. *Lymphokine Cytokine Res* **12**, 225-230 (1993).
- 47 Baird, G. S. *et al.* Age-dependent changes in the cerebrospinal fluid proteome by slow off-rate modified aptamer array. *The American journal of pathology* **180**, 446-456 (2012).
- 48 Tanaka, T. *et al.* Plasma proteomic signature of age in healthy humans. *Aging Cell* **17**, e12799 (2018). <https://doi.org/10.1111/acer.12799>
- 49 Lind, L. *et al.* Longitudinal effects of aging on plasma proteins levels in older adults - associations with kidney function and hemoglobin levels. *PLoS One* **14**, e0212060 (2019). <https://doi.org/10.1371/journal.pone.0212060>
- 50 Signorelli, M. *et al.* Longitudinal serum biomarker screening identifies malate dehydrogenase 2 as candidate prognostic biomarker for Duchenne muscular dystrophy. *J Cachexia Sarcopenia Muscle* **11**, 505-517 (2020). <https://doi.org/10.1002/jcsm.12517>
- 51 Menni, C. *et al.* Circulating Proteomic Signatures of Chronological Age. *J Gerontol A Biol Sci Med Sci* **70**, 809-816 (2015). <https://doi.org/10.1093/gerona/glu121>
- 52 Di Narzo, A. F. *et al.* High-Throughput Characterization of Blood Serum Proteomics of IBD Patients with Respect to Aging and Genetic Factors. *PLoS Genet* **13**, e1006565 (2017). <https://doi.org/10.1371/journal.pgen.1006565>
- 53 Huang, Y. & Mahley, R. W. Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. *Neurobiol Dis* **72 Pt A**, 3-12 (2014). <https://doi.org/10.1016/j.nbd.2014.08.025>
- 54 Liu, C. C., Liu, C. C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol* **9**, 106-118 (2013). <https://doi.org/10.1038/nrneurol.2012.263>
- 55 Raulin, A. C. *et al.* ApoE in Alzheimer's disease: pathophysiology and therapeutic strategies. *Mol Neurodegener* **17**, 72 (2022). <https://doi.org/10.1186/s13024-022-00574-4>
- 56 Yassine, H. N. & Finch, C. E. APOE Alleles and Diet in Brain Aging and Alzheimer's Disease. *Front Aging Neurosci* **12**, 150 (2020). <https://doi.org/10.3389/fnagi.2020.00150>
- 57 de Mezer, M. *et al.* SERPINA3: Stimulator or Inhibitor of Pathological Changes. *Biomedicines* **11** (2023). <https://doi.org/10.3390/biomedicines11010156>

- 58 Padmanabhan, J., Levy, M., Dickson, D. W. & Potter, H. Alpha1-antichymotrypsin, an inflammatory protein overexpressed in Alzheimer's disease brain, induces tau phosphorylation in neurons. *Brain* **129**, 3020-3034 (2006). <https://doi.org/10.1093/brain/awl255>
- 59 Poljak, A. *et al.* Quantitative proteomics of delirium cerebrospinal fluid. *Transl Psychiatry* **4**, e477 (2014). <https://doi.org/10.1038/tp.2014.114>
- 60 Zheng, R. *et al.* The Complement System, Aging, and Aging-Related Diseases. *Int J Mol Sci* **23** (2022). <https://doi.org/10.3390/ijms23158689>
- 61 Gaya da Costa, M. *et al.* Age and Sex-Associated Changes of Complement Activity and Complement Levels in a Healthy Caucasian Population. *Front Immunol* **9**, 2664 (2018). <https://doi.org/10.3389/fimmu.2018.02664>
- 62 Ponting, C. P., Marshall, J. M. & Cederholm-Williams, S. A. Plasminogen: a structural review. *Blood Coagul Fibrinolysis* **3**, 605-614 (1992).
- 63 Plow, E. F. & Hoover-Plow, J. The functions of plasminogen in cardiovascular disease. *Trends Cardiovasc Med* **14**, 180-186 (2004). <https://doi.org/10.1016/j.tcm.2004.04.001>
- 64 Ha, H., Oh, E. Y. & Lee, H. B. The role of plasminogen activator inhibitor 1 in renal and cardiovascular diseases. *Nat Rev Nephrol* **5**, 203-211 (2009). <https://doi.org/10.1038/nrneph.2009.15>
- 65 Sisson, T. H. & Simon, R. H. The plasminogen activation system in lung disease. *Curr Drug Targets* **8**, 1016-1029 (2007). <https://doi.org/10.2174/138945007781662319>
- 66 Kwaan, H. C. Changes in blood coagulation, platelet function, and plasminogen-plasmin system in diabetes. *Diabetes* **41 Suppl 2**, 32-35 (1992). <https://doi.org/10.2337/diab.41.2.s32>
- 67 Lyon, C. J. & Hsueh, W. A. Effect of plasminogen activator inhibitor-1 in diabetes mellitus and cardiovascular disease. *Am J Med* **115 Suppl 8A**, 62S-68S (2003). <https://doi.org/10.1016/j.amjmed.2003.08.014>
- 68 Auwerx, J., Bouillon, R., Collen, D. & Geboers, J. Tissue-type plasminogen activator antigen and plasminogen activator inhibitor in diabetes mellitus. *Arteriosclerosis* **8**, 68-72 (1988). <https://doi.org/10.1161/01.atv.8.1.68>
- 69 Aso, Y. *et al.* Relationship between soluble thrombomodulin in plasma and coagulation or fibrinolysis in type 2 diabetes. *Clin Chim Acta* **301**, 135-145 (2000). [https://doi.org/10.1016/s0009-8981\(00\)00335-1](https://doi.org/10.1016/s0009-8981(00)00335-1)
- 70 Uitte de Willige, S. *et al.* Proteolytic and genetic variation of the alpha-2-antiplasmin C-terminus in myocardial infarction. *Blood* **117**, 6694-6701 (2011). <https://doi.org/10.1182/blood-2010-11-320325>
- 71 Bianchini, E. P., Auditeau, C., Razanakolona, M., Vasse, M. & Borgel, D. Serpins in Hemostasis as Therapeutic Targets for Bleeding or Thrombotic Disorders. *Front Cardiovasc Med* **7**, 622778 (2020). <https://doi.org/10.3389/fcvm.2020.622778>
- 72 Zou, Y., Jung, K. J., Kim, J. W., Yu, B. P. & Chung, H. Y. Alteration of soluble adhesion molecules during aging and their modulation by calorie restriction. *FASEB J* **18**, 320-322 (2004). <https://doi.org/10.1096/fj.03-0849fje>
- 73 Kuuliala, A. *et al.* Circulating soluble E-selectin in early rheumatoid arthritis: a prospective five year study. *Ann Rheum Dis* **61**, 242-246 (2002). <https://doi.org/10.1136/ard.61.3.242>
- 74 Kawamura, T. *et al.* The incidence and characteristics of silent cerebral infarction in elderly diabetic patients: association with serum-soluble adhesion molecules. *Diabetologia* **41**, 911-917 (1998). <https://doi.org/10.1007/s001250051007>
- 75 Matsumoto, K., Fujishima, K., Moriuchi, A., Saishoji, H. & Ueki, Y. Soluble adhesion molecule E-selectin predicts cardiovascular events in Japanese patients with type 2 diabetes mellitus. *Metabolism* **59**, 320-324 (2010). <https://doi.org/10.1016/j.metabol.2009.06.011>

- 76 Fassbender, K., Bertsch, T., Mielke, O., Muhlhauser, F. & Hennerici, M. Adhesion molecules in cerebrovascular diseases. Evidence for an inflammatory endothelial activation in cerebral large- and small-vessel disease. *Stroke* **30**, 1647-1650 (1999). <https://doi.org/10.1161/01.str.30.8.1647>
- 77 Du Clos, T. W. & Mold, C. Pentraxins (CRP, SAP) in the process of complement activation and clearance of apoptotic bodies through Fcγ receptors. *Curr Opin Organ Transplant* **16**, 15-20 (2011). <https://doi.org/10.1097/MOT.0b013e32834253c7>
- 78 Doni, A. *et al.* Serum amyloid P component is an essential element of resistance against. *Nature Communications* **12** (2021). <https://doi.org/ARTN> 3739
- 10.1038/s41467-021-24021-y
- 79 Gate, D., Saligrama, N., Levanthal, O., Davis, M. & Wyss-Coray, T. Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease: Molecular and cell biology: Inflammatory and microglia-mediated mechanisms in AD. *Alzheimer's & Dementia* **16**, e044082 (2020).
- 80 Gate, D. *et al.* CD4+ T cells contribute to neurodegeneration in Lewy body dementia. *Science* **374**, 868-874 (2021).
- 81 Morgan, B. P. & Walport, M. J. Complement deficiency and disease. *Immunol Today* **12**, 301-306 (1991). [https://doi.org/10.1016/0167-5699\(91\)90003-C](https://doi.org/10.1016/0167-5699(91)90003-C)
- 82 Sjöholm, A. G., Jonsson, G., Braconier, J. H., Sturfelt, G. & Truedsson, L. Complement deficiency and disease: an update. *Mol Immunol* **43**, 78-85 (2006). <https://doi.org/10.1016/j.molimm.2005.06.025>
- 83 Unsworth, D. J. Complement deficiency and disease. *J Clin Pathol* **61**, 1013-1017 (2008). <https://doi.org/10.1136/jcp.2008.056317>
- 84 Ross, S. C. & Densen, P. Complement deficiency states and infection: epidemiology, pathogenesis and consequences of neisserial and other infections in an immune deficiency. *Medicine (Baltimore)* **63**, 243-273 (1984).
- 85 Mayilyan, K. R. Complement genetics, deficiencies, and disease associations. *Protein Cell* **3**, 487-496 (2012). <https://doi.org/10.1007/s13238-012-2924-6>
- 86 Mollah, F. & Tam, S. in *StatPearls* (2023).
- 87 Protter, D. S. W. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell Biol* **26**, 668-679 (2016). <https://doi.org/10.1016/j.tcb.2016.05.004>
- 88 Horwitz, M. A. Phagocytosis of microorganisms. *Rev Infect Dis* **4**, 104-123 (1982). <https://doi.org/10.1093/clinids/4.1.104>
- 89 Aderem, A. Phagocytosis and the inflammatory response. *J Infect Dis* **187 Suppl 2**, S340-345 (2003). <https://doi.org/10.1086/374747>
- 90 Cerutti, C. & Ridley, A. J. Endothelial cell-cell adhesion and signaling. *Exp Cell Res* **358**, 31-38 (2017). <https://doi.org/10.1016/j.yexcr.2017.06.003>
- 91 Mateo, M., Generous, A., Sinn, P. L. & Cattaneo, R. Connections matter--how viruses use cell-cell adhesion components. *J Cell Sci* **128**, 431-439 (2015). <https://doi.org/10.1242/jcs.159400>
- 92 Quispe Calla, N. E., Vicetti Miguel, R. D., Aceves, K. M., Torres, A. & Cherpes, T. L. Depot-medroxyprogesterone acetate reduces genital cell-cell adhesion molecule expression and increases genital herpes simplex virus type 2 infection susceptibility in a dose-dependent fashion. *Contraception* **100**, 397-401 (2019). <https://doi.org/10.1016/j.contraception.2019.07.003>
- 93 Kurpios, N. A. *et al.* The direction of gut looping is established by changes in the extracellular matrix and in cell:cell adhesion. *Proc Natl Acad Sci U S A* **105**, 8499-8506 (2008). <https://doi.org/10.1073/pnas.0803578105>
- 94 Liu, Y., Yao, Z., Lian, G. & Yang, P. Biomolecular phase separation in stress granule assembly and virus infection. *Acta Biochim Biophys Sin (Shanghai)* **55**, 1099-1118 (2023). <https://doi.org/10.3724/abbs.2023117>

- 95 Rasmussen, M. & Bjorck, L. Proteolysis and its regulation at the surface of *Streptococcus pyogenes*. *Mol Microbiol* **43**, 537-544 (2002).
<https://doi.org/10.1046/j.1365-2958.2002.02766.x>
- 96 Frees, D., Brondsted, L. & Ingmer, H. Bacterial proteases and virulence. *Subcell Biochem* **66**, 161-192 (2013). https://doi.org/10.1007/978-94-007-5940-4_7
- 97 Deming, Y. *et al.* Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol* **133**, 839-856 (2017). <https://doi.org/10.1007/s00401-017-1685-y>
- 98 Morris, J. C. *et al.* APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging. *Ann Neurol* **67**, 122-131 (2010).
<https://doi.org/10.1002/ana.21843>
- 99 Knight Alzheimer Disease Research Center, <<https://knightadrc.wustl.edu/>> (2023).
- 100 Alzheimer's Disease Neuroimaging Initiative, <www.adni-info.org> (2023).
- 101 Ace Alzheimer Center Barcelona, <www.fundacioace.com/en> (2023).
- 102 Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **5**, e15004 (2010). <https://doi.org/10.1371/journal.pone.0015004>
- 103 Yang, C. *et al.* Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. *Nat Neurosci* **24**, 1302-1312 (2021).
<https://doi.org/10.1038/s41593-021-00886-6>
- 104 Timsina, J. *et al.* Comparative Analysis of Alzheimer's Disease Cerebrospinal Fluid Biomarkers Measurement by Multiplex SOMAscan Platform and Immunoassay-Based Approach. *J Alzheimers Dis* **89**, 193-207 (2022). <https://doi.org/10.3233/JAD-220399>
- 105 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289-300 (1995).
- 106 Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software* **33**, 1 (2010).
- 107 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 1-13 (2008).
- 108 Song, L., Langfelder, P. & Horvath, S. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC bioinformatics* **13**, 1-21 (2012).
- 109 Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719-720 (2008).
- 110 Lakens, D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology* **4**, 863 (2013).
- 111 Cleveland, W., Grosse, E. & Shyu, W. Local regression models. Chapter 8 in *Statistical models in S* (JM Chambers and TJ Hastie eds.), 608 p. *Wadsworth & Brooks/Cole, Pacific Grove, CA* (1992).
- 112 Yu, G., Wang, L.-G., Yan, G.-R. & He, Q.-Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608-609 (2014).
- 113 Bodenreider, O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research* **32**, D267-D270 (2004).
- 114 Lamy, J.-B. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. *Artificial intelligence in medicine* **80**, 11-28 (2017).
- 115 Hirsch, J. *et al.* ICD-10: history and context. *American Journal of Neuroradiology* **37**, 596-599 (2016).
- 116 Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).
- 117 Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one* **6**, e21800 (2011).

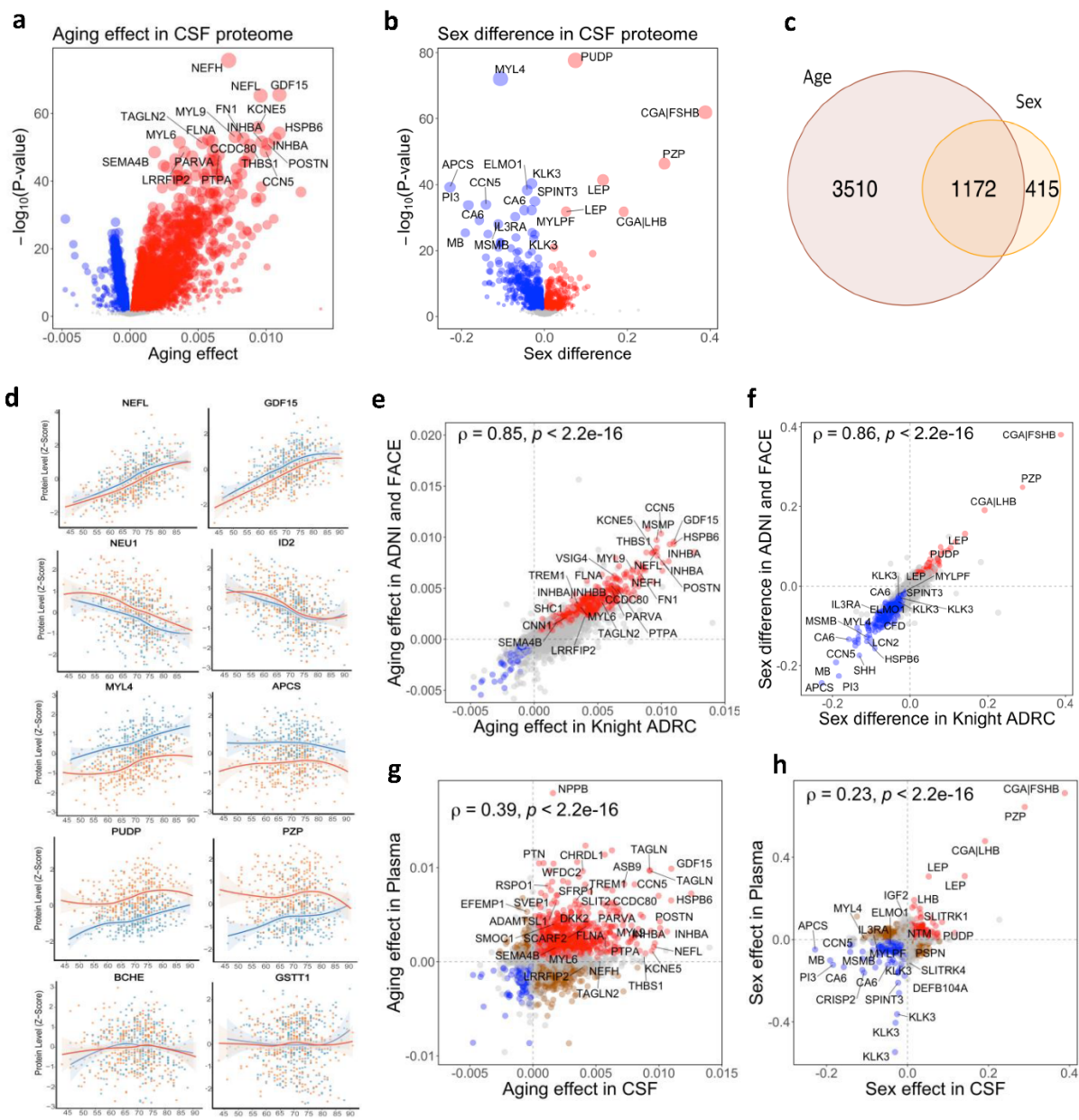


Fig. 1: Age effects and sex differences in CSF proteomics.

(a) Age effects on CSF proteins in Knight ADRC. (b) Sex differences in CSF proteins in Knight ADRC. (c) Significant proteins associated with age and/or sex at FDR < 0.05. (d) Protein abundance across ages for selected proteins in males (blue dots) and females (orange dots). The locally estimated scatterplot smoothing (LOESS) lines were separately in males (blue) and females (red). (e) Comparison of age effects on CSF proteins in independent ADNI and FACE (y-axis) versus those in Knight ADRC (x-axis). (f) Comparison of sex differences in CSF proteins in ADNI and FACE versus those in Knight ADRC. (g) Comparison between aging effects on plasma proteins versus those in CSF proteins. (h) Comparison between sex differences in plasma proteins versus those in CSF proteins

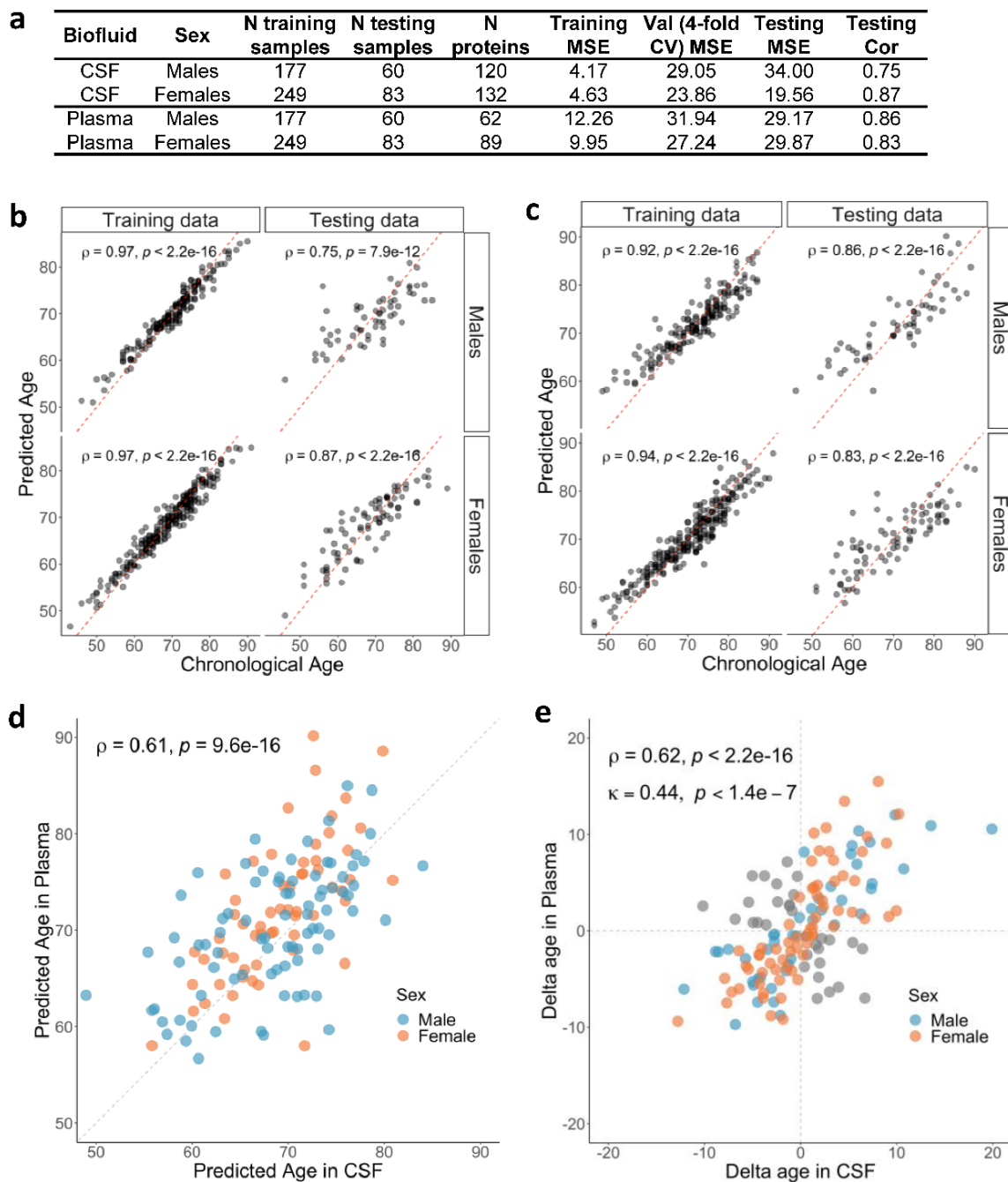


Fig. 2: Sex-specific proteomic clocks in CSF and plasma

(a) Summary of creating sex-specific proteomic clocks. (b) Performance of CSF proteomic clock with chronological ages (x-axis) and predicted biological ages in CSF (y-axis). (c) Performance of plasma proteomic clock. (d) Comparison between CSF predicted age (x-axis) and plasma predicted age (y-axis) in males (blue) and females (red). (e) Comparison of delta ages in CSF versus in plasma. The delta age is a departure of the predicted age from the chronological age.

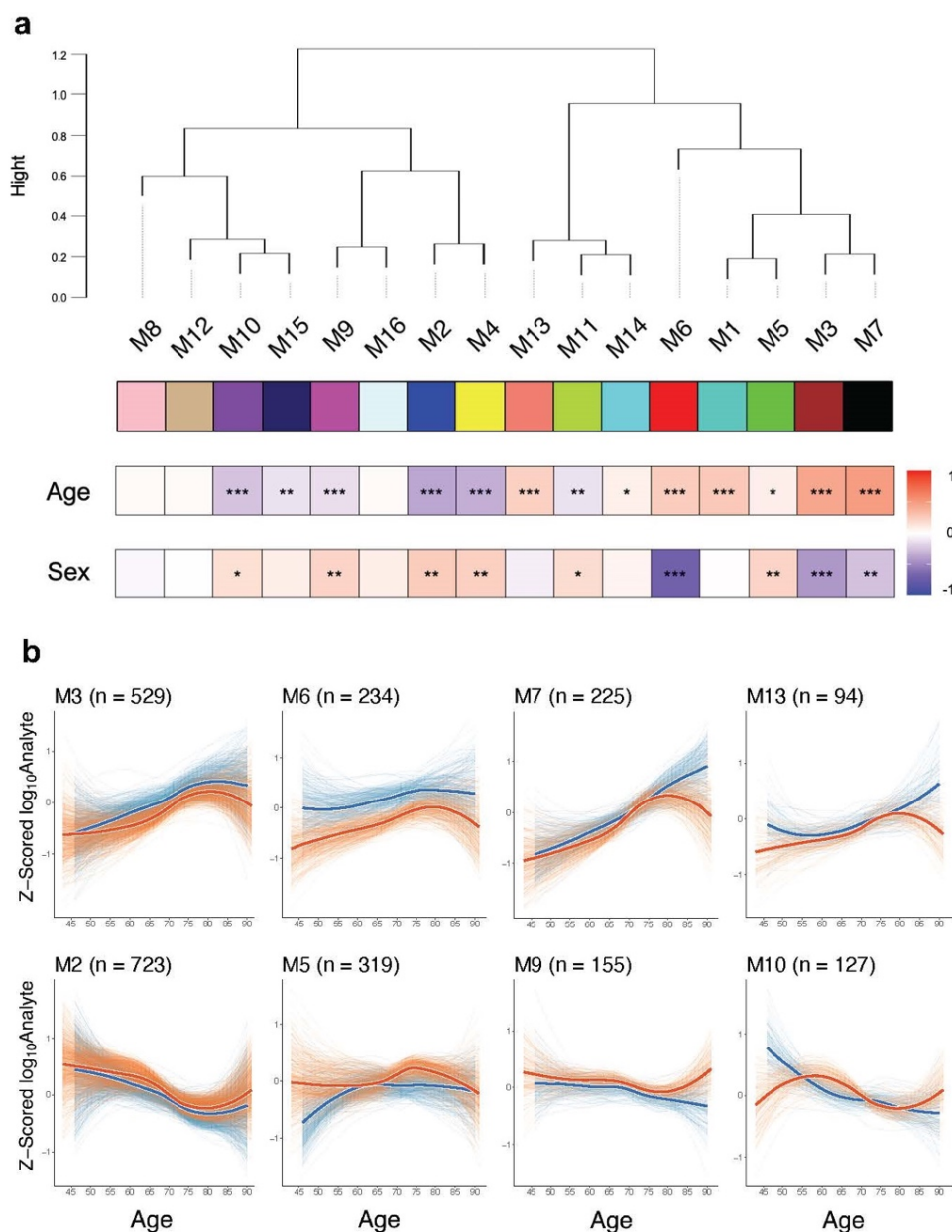


Fig. 3 Clustering of CSF proteins into 16 modules using WGCNA

(a) WGCNA identified the 16 proteins modules. The age row shows the correlation coefficient between proteins and ages (red for increasing with age; blue for decreasing). The sex row shows Cohen's distance between proteins in females vs those in males (red for higher in females; blue for higher in males). *** for $P < 0.0001$, ** for $P < 0.001$, and * for $P < 0.05$. (b) Aging trajectory for the select modules. LOESS lines for individual proteins in males (thin blue lines) and females (thin orange lines). The thick lines are the average across them separately in males (blue) and females (red). The number of included analytes (n) is shown. Aging trajectory for all modules is shown in Supplementary Fig. 3.

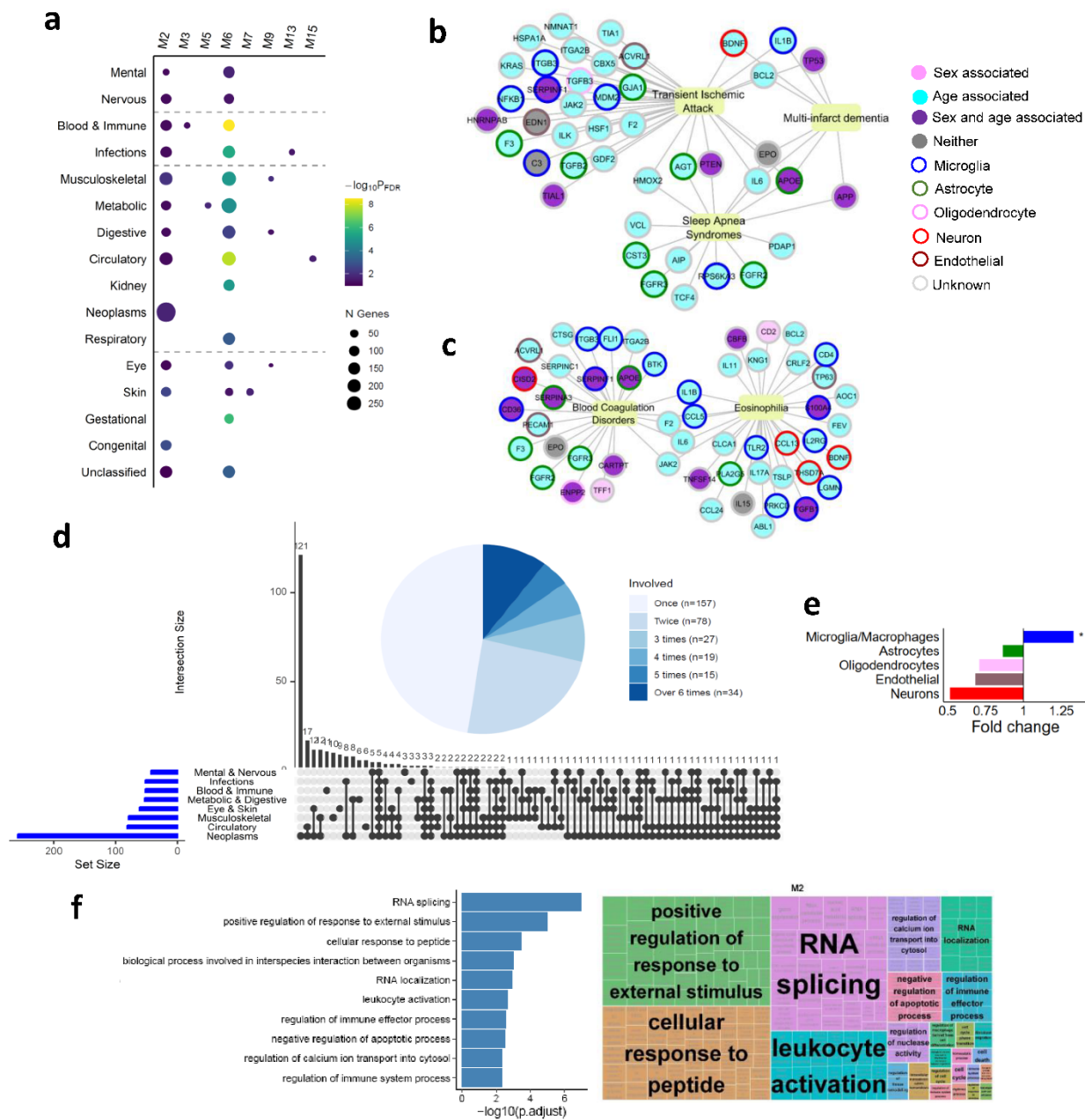


Fig. 4: Enrichment of module M2.

(a) The enrichment of 8 modules for the 16 ICD-10 disease chapters. (b) M2 proteins involved with mental and neuro-diseases. (c) M2 proteins involved with blood and immune mechanisms. (d) Upset plot for M2 proteins enriched with the 11 chapters. Congenital and unclassified are excluded. The pie chart shows the number of M2 proteins overlapping across the ICD-10 chapters (e) Cell-type specificity of M2 (f) Gene ontology enrichment of M2.

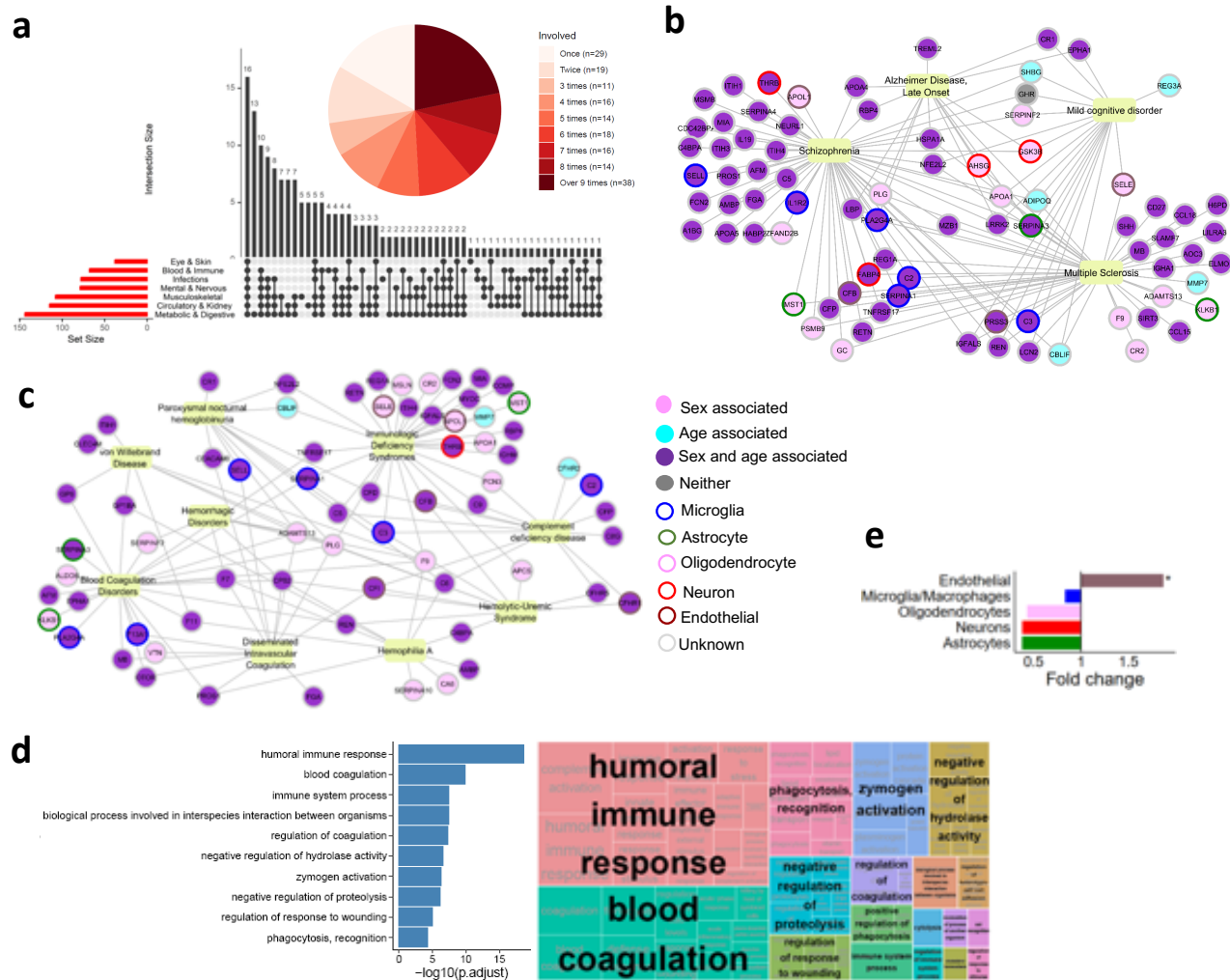


Fig. 5: Enrichment of module M6

(a) Upset plot for M6 proteins enriched with the 11 chapters. Respiratory, congenital and unclassified were excluded. The pie chart shows the number of M6 proteins overlapping across the ICD-10 chapters. (b) M2 proteins involved with mental and neuro-diseases. (c) M2 proteins involved with blood and immune mechanisms. (d) Gene ontology enrichment of M6 (e) Cell-type specificity of M6.