Cerebrospinal fluid endo-lysosomal proteins as potential biomarkers for Huntington's disease

Alexander J. Lowe[1], Simon Sjodin[2], Filipe B. Rodrigues[1], Lauren M. Byrne[1], Kaj Blennow[2,3], Rosanna Tortelli[1], Henrik Zetterberg[1,2,3,4] and Edward J. Wild[1*].

[1]UCL Huntington's Disease Centre, UCL Queen Square Institute of Neurology, University College London, London, United Kingdom.

[2]Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden.

[3]Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden.

[4]UK Dementia Research Institute at UCL, London, United Kingdom.

*Corresponding Author

Email: e.wild@ucl.ac.uk (EJW)

Abstract

Molecular markers derived from cerebrospinal fluid (CSF) represent an accessible means of exploring the pathobiology of Huntington's disease (HD) in vivo. The endo-lysosomal/autophagy system is dysfunctional in HD, potentially contributing to disease pathogenesis and representing a potential target for therapeutic intervention. Several endo-lysosomal proteins have shown promise as biomarkers in other neurodegenerative diseases; however, they have yet to be fully explored in HD. We performed parallel reaction monitoring mass spectrometry analysis (PRM-MS) of multiple endo-lysosomal proteins in the CSF of 60 HD mutation carriers and 20 healthy controls. Using generalised linear models controlling for age and CAG, none of the 18 proteins measured displayed significant differences in concentration between HD patients and controls. This was affirmed by principal component analysis, in which no significant difference across disease stage was found in any of the three components representing lysosomal hydrolases, binding/transfer proteins and innate immune system/peripheral proteins. However, several proteins were associated with measures of disease severity and cognition: most notably amyloid precursor protein, which displayed strong correlations with composite Unified Huntington's Disease Rating Scale, UHDRS Total Functional Capacity, UHDRS Total Motor Score, Symbol Digit Modalities Test and Stroop Word Reading. We conclude that although endo-lysosomal proteins are unlikely to have value as disease state CSF biomarkers for Huntington's disease, several proteins demonstrate associations with clinical severity, thus warranting further, targeted exploration and validation in larger, longitudinal samples.

Introduction

Huntington's disease (HD) is an autosomal dominant, neurodegenerative disease characterised by progressive motor, psychiatric and cognitive dysfunction [1]. An extended polyglutamine tract (polyQ) in the ubiquitously-expressed Huntingtin protein (HTT), results in the production of a mutated, pathogenic product (mHTT) which accumulates intracellularly causing toxicity and neuronal death [2,3].

Neuronal survival is dependent, among other things, on intracellular surveillance mechanisms including autophagy, a lysosomal pathway that serves to eliminate toxic substances via two mechanisms: macroautophagy and chaperone-mediated autophagy (CMA) [4,5]. Both of these are disrupted in neurodegenerative diseases including Parkinson's disease (PD), Alzheimer's disease (AD) and polyQ disorders [6–12], potentially resulting in autophagic dysfunction and exacerbation of the neurodegenerative process [13].

Lysosomal-associated membrane protein-2 (LAMP2) has pivotal roles in autophagy including translocation of cargo into the lumen and as a receptor in CMA [14,15]. LAMP2 gene expression levels and total levels of LAMP2 protein have been shown to be reduced and increased in PD and AD respectively [16–19]. Additionally, cerebrospinal fluid (CSF) LAMP2 has been indicated as a potential biomarker in AD with increased concentration compared to controls [18,20] and has been found to correlate with phosphorylated tau, a well-established marker of neuronal pathology [21]. In HD, a compensatory increase in CMA has been described in response to defective macroautophagy which may explain the increased mRNA expression of LAMP2 and increased levels of LAMP2 protein in HD cell models [11].

Deficits in lipid synthesis and metabolism, both of which are reported in HD [22], could contribute towards autophagy failure [23]. Glycosphingolipids endocytosed from the plasma membrane are degraded in the lysosome via the synchronous activity of hydrolases and activator proteins [24]. Ganglioside GM2 activator (GM2A) is a lysosomal protein that together with beta-hexosaminidase-β (HEXB), catalyses the degradation of gangliosides, specifically GM2 [25]. GM2A has shown promise as a CSF biomarker for neurodegeneration in AD, correlating with CSF amyloid-beta levels, and in Lewy body dementia (LBD) with increased concentration [26], whilst the concentration in PD has shown to be reduced [27]. The reason for elevated CSF GM2A in AD and LBD is currently unknown but likely reflects generalised lysosomal dysfunction, as elevated GM2A has been detected via urinary analysis in lysosomal storage disorders [28]. In HD, the reduced expression of genes involved in ganglioside catabolism has been reported [29], in addition to disturbances in ganglioside metabolism and synthesis [29,30]. Furthermore, administration of gangliosides has been found to reduce apoptosis in HD cell lines and restore normal ganglioside concentration in YAK128 mice, resulting in improved motor function [30,31]. Given that gangliosides are involved in regulating white matter integrity [32], and that white matter atrophy is associated with HD [33–35], the exploration of CSF GM2A, a protein pivotal for ganglioside catabolism, is warranted and may further explain white matter pathology in HD.

Lysosomal proteolytic degradation involves the activity of the cathepsin family of proteases [36]. Previous work using CSF has demonstrated significant alterations in the concentration of several cathepsins in other proteopathies such as PD [27]. Both Cathepsin L and Z have been shown to be crucial for the degradation of polyQ proteins within lysosomes [37], suggesting a protective role against toxic aggregates.

The role of additional cathepsins in HD has also been explored, with early work describing an increase in Cathepsin D activity in caudate tissue of HD patients [38]. This has been supported by recent studies showing increased Cathepsin D and L levels in response to mHTT expression in vitro [39], and studies demonstrating overexpression of Cathepsin B and D to reduce mHTT levels and toxicity in multiple cell models, without impacting upon endogenous HTT [40].

CSF is enriched in brain-derived substances, thus biomarkers derived from CSF represent a valid means to assess neuropathology [41]. Given the dysregulation of the autophagy pathway in HD [23], the exploration of endo-lysosomal proteins in HD patients could represent a means of identifying novel biomarkers with prognostic, disease monitoring and pharmacodynamic value [42]. Parallel reaction monitoring mass spectrometry (PRM-MS) is a quantitative approach making use of high resolution instruments and thus offers highly selective and accurate measurements [43,44]. Separation in two dimensions, by physiochemical properties using liquid chromatography and by mass to charge ratio (m/z) using mass spectrometry, facilitates multiplexing capabilities in complex matrices, for example in biofluids. The PRM-MS method employed herein has previously been applied to investigate endo-lysosomal dysfunction in AD and PD patients, with the later demonstrating altered CSF concentrations of multiple cathepsins, GM2A and LAMP2 [27].

We employed PRM-MS to conduct a targeted analysis of 18 proteins associated with endocytosis and lysosomal function in the CSF from the HD-CSF cohort baseline (60 HD mutation carriers and 20 healthy controls). Given the previously described autophagic dysfunction in HD, and their role in other neurodegenerative diseases, we pre-specified 5 lysosomal proteins as primary analytes to study: LAMP1, LAMP2, GM2A, Cathepsin D and F. The remaining 13 proteins, pertaining to other aspects of

the endo-lysosomal and ubiquitin-proteasome system, were assessed in a separate exploratory analysis. We aimed to elucidate the biomarker potential of endo-lysosomal proteins whilst also highlighting targets for future comprehensive analysis, with the aim of facilitating therapeutic developments in HD.

Materials and methods

Participants and study design

HD-CSF was a prospective single-site study with standardised longitudinal collection of CSF, blood and phenotypic data (online protocol: DOI: 10.5522/04/11828448.v1). Ethical approval was given by the London Camberwell St Giles Research Ethics Committee, with all participants providing written informed consent prior to enrolment. The study involved manifest HD, premanifest HD and healthy controls. Manifest HD was defined as UHDRS diagnostic confidence level (DCL) = 4 and CAG repeat length > 36. Premanifest HD had CAG repeat length > 40 and DCL < 4. Healthy controls were contemporaneously recruited, drawn from a population with a similar age to patients, and clinically well, so the risk of incidental neurodegenerative diseases was very low. Consent, inclusion and exclusion criteria, clinical assessment, CSF collection and storage were all as previously described [45,46]. In brief, samples were collected after an overnight fast at the same time of day and centrifuged and aliquoted on ice using a standardised protocol and polypropylene plasticware. Relevant aspects of clinical phenotype were quantified using the Unified Huntington's Disease Rating Scale (UHDRS) [47]. A composite UHDRS (cUHDRS) score was generated for each subject to provide a single measure of motor, cognitive and global functioning decline. This composite score, computed using four measures; Total Functional Capacity (TFC), Total Motor Score (TMS), Symbol Digit

Modality Test (SDMT) and Stroop Word Reading (SWR), has been found to display the strongest relationship to HD brain pathology and enhanced sensitivity to clinical change in early HD [48]. Disease burden score (DBS) was calculated for each HD patient using the formula [CAG repeat length – 35.5] × age [49]. DBS estimates cumulative HD pathology exposure as a function of CAG repeat length and the time exposed to the effects of the expansion, and has been shown to predict several features of disease progression including striatal pathology [49,50]. Baseline samples from HD-CSF have been used for this study.

Sample preparation

Measurement of peptide concentrations was performed as previously described [27], which builds on the original method developed by Brinkmalm *et al*.[51]. However, some minor modifications were introduced. In short, 50 µL CSF was mixed with 50 µL of an internal standard mixture containing stable isotope-labelled peptides (JPT Peptide Technologies GmbH, Berlin, Germany; Thermo Fisher Scientific Inc. Waltham, MA, USA), ^{13}C-labelled ubiquitin (Silantes, GmbH, München, Germany) and bovine serum albumin (Sigma-Aldrich Co., Saint Louis, MO, USA), diluted in 50 mM NH_4HCO_3 (see S1 Table). Reduction and alkylation was performed by the addition of 50 µL 15 mM 1,4-dithiothreitol in 50 mM NH_4HCO_3, shaking for 30 min at + 60 °C, cooling down at room temperature for 30 min, and finally the addition of 25 µL 70 mM iodoacetamide in 50 mM NH_4HCO_3 followed by shaking at room temperature in the dark for 30 min. The samples were digested by the addition of 25 µL 0.08 µg/µL sequencing grade modified trypsin (Promega Co., Madison, WI, USA) diluted in 50 mM NH_4HCO_3 and incubated at + 37 °C shaking for 18 h. Digestion was ended by the addition of 25 µL 10% trifluoroacetic acid. Solid-phase extraction was

performed using Oasis® HLB 96-well µElution Plates (2 mg sorbent and 30 µm particle size; Waters Co., Milford, MA, USA) by conditioning (2x300 µL methanol), equilibration (2 × 300 µL H_2O), loading of samples, washing (2 × 300 µL H_2O), and elution (2 × 100 µL methanol). The samples were then dried by vacuum centrifugation and stored at – 80 °.

Parallel reaction monitoring mass spectrometry

Prior to analysis by PRM-MS the samples were dissolved by the addition of 50 µL 50 mM NH_4HCO_3, and shaking at room temperature for 1 h. Forty microliters of sample were injected and separated using a Dionex™ UltiMate™ 3000 standard-LC system (Thermo Fisher Scientific Inc., Waltham, MA, USA) and a Kinetex® EVO C18 column (length 150 mm; inner diameter 2.1 mm; particle size 1.7 µm; Phenomenex Inc., Torrance, CA, USA) with a SecurityGuard™ ULTRA cartridge prefilter (Phenomenex Inc.). On a 60 minutes method, with solvents A (0.1% formic acid in H_2O (v/v)) and B (84% acetonitrile and 0.1% fromic acid in H_2O (v/v)), using a flow rate of 300 µL/min, the gradient went from 3 to 5% B over one minute followed by 5 to 26% B over 48 minutes. The column temperature was set to + 50 °C. Separation by high-performance liquid chromatography, as described above, was performed in online mode coupled to a Q Exactive™ Hybrid Quadrupole-Orbitrap™ mass spectrometer (Thermo Fisher Scientific Inc.). Using a HESI-II ionization probe (Thermo Fisher Scientific Inc.) electrospray ionization was performed in positive ion mode with the following settings: spray voltage + 4.1 kV, heater temperature + 400 °C, capillary transfer tube temperature + 380 °C, sheath gas flow rate 25, auxiliary gas flow rate 10, and S-Lens RF level 60. Acquisition of data was performed using single microscans in parallel reaction monitoring (PRM) mode with an isolation window of

193 *m/z* 2 centred on the second isotope of the precursor ion. The resolution setting was
194 70 k with an AGC target of 1×10^6 and a 256 ms injection time. Fragmentation was
195 performed using beam-type collision-induced dissociation (higher energy collision
196 induced dissociation [52] with optimized energies as described before [27]. The PRM
197 method was scheduled using one-minute retention time windows. Peptide related
198 settings are shown in S1 Table.

199 Data extraction

200 Skyline v.19.1 [53] was used to calculate and export fragment ion peak areas.
201 Skyline was also used to monitor and evaluate fragment ion traces and ratios, and to
202 determine which fragment ions to include in the analysis. The ratio between tryptic
203 peptide and isotope-labelled peptide peak area was used for quantification. In total
204 48 peptides from 19 proteins, including added bovine serum albumin as a control
205 protein, were monitored. With each set of samples analysed, four quality control
206 replicates from a CSF pool were run to normalize variation between sets of samples.
207 In this case the samples were split in two sets, however prepared on a single
208 occasion but analysed using PRM-MS at different points in time. The median of the
209 first set's four quality control replicates was used for normalization by dividing the
210 median of the second set's quality control median. Then the samples in the second
211 set were divided by the resulting normalization quotient (one for each peptide). As
212 multiple peptides were monitored from each protein the complexity of the data was
213 reduced by transforming the peptide ratios into a single value, see Equation 1. The
214 transformation was done for proteins with correlating peptides. To create a protein-
215 level estimate, a Mean Peptide Ratio was calculated by dividing the peptide ratio (x)
216 by the mean of all ratios for that peptide in the study ($\bar{x}$). The calculation was made

for peptides 1-n, and was then divided by the number of peptides (n) derived from the protein. Thus, the sample ratios for each peptide were normalized to have a mean of 1, without affecting the relative difference between samples. Additionally, the weight of each peptide in the calculation of the Mean Peptide Ratio became approximately equal.

Equation 1: $$Mean\ Peptide\ Ratio_{1-n} = \frac{{}^{x_1}/_{\bar{x}_1} + {}^{x_2}/_{\bar{x}_2} + \ldots + {}^{x_n}/_{\bar{x}_n}}{n}$$

Precision, shown in S1 Table, was monitored by analysing eight quality control replicates from a CSF pool, which were run with each sample set. The precision and limit of quantification of the method have previously been determined [27]. Given the two sets of samples analysed, the within set variability had coefficients of variation of 1.8-15.8%, depending on peptide. Between sample sets, the coefficients of variation were 2.7-21.0%. For the Mean Peptide Ratio the within set variability coefficients of variation varied between 2.0-13.9% while the between sets variations were 2.1-18.3%.

Statistical analysis

Statistical analysis was performed with Stata IC 15 software (StataCorp, TX, USA). The distribution of all protein concentrations were tested for normality and found to be non-normally distributed. Natural log-transformation was applied and produced an acceptable distribution for all analytes. Based on their putative involvement in the pathogenesis of HD in the literature, we pre-specified 5 proteins (LAMP1, LAMP2, GM2A, and Cathepsins D and F) and designated them as primary analytes (see S1 Table for full protein list). Differences in demographic and clinical characteristics were examined using ANOVA and χ^2 tests. Age, gender and blood contamination

were considered potentially confounding variables, thus their relationship with analyte concentration was examined in a preliminary analysis in controls using independent samples t-tests and Pearson's correlation. Differences across disease stage were tested using general linear models controlling for age. CAG repeat length was also included in the model when assessing differences between premanifest and manifest HD mutation carriers. To test for associations with measures of clinical severity and cognition, Pearson's partial correlation coefficients, bootstrapped with 1000 repetitions, were calculated controlling for age and CAG in all HD gene expansion carriers. Biomarker potential was assessed by controlling relationships first for age, and then for age and CAG. By including both age and CAG as covariates, accurate assessments of associations can be made, independent of known predictors. DBS is a product of age and CAG, as such, the latter two variables were not included as covariates when assessing relationships with DBS.

Principal components analysis (PCA) was employed to reduce the dimensionality of the entire protein dataset. PCA is used to identify the maximum number of uncorrelated principal components that together explain the maximum amount of variance in a data set [54]. We leveraged the Kaiser-Meyer-Olkin measure of sample adequacy and Bartlett's test of sphericity to assess the suitability of our data for PCA. Prior to running the PCA, we controlled each protein for the effect of age using general linear models. When selecting the number of components to use in subsequent analysis, we followed the recommendation to limit this to the smallest number accounting for the most variability in the data [55]. As such, we inspected scree plots and selected components with an eigenvalue of >1. Orthogonal varimax rotation was applied and variables with a loading of >0.3 were deemed significant and used to define the component labels. Participant's original data were then

transformed to create a composite score for each principal component. Group differences could then be analysed using this small number of principal components, rather than the large number of original measures. Mirroring the analysis at the level of individual proteins, general linear models and Pearson's partial correlation were used to assess group differences in component scores and the relationships to measures of clinical severity. Age was not included in the models as it had already been controlled for using linear regression in the generation of PC scores

Significance level was defined as $p < 0.05$ and tests were Bonferroni-corrected for multiple comparisons when required.

A further exploratory analysis was undertaken on the remaining 13 endo-lysosomal proteins using the same hierarchical methodology outlined above.

Results

Participant characteristics

Our cohort consisted of 20 healthy controls and 60 HD mutation carriers. The HD gene expansion carriers comprised of 20 premanifest and 40 manifest HD patients. A single premanifest participant was removed due to missing data. There were no significant differences in the gender distribution ($\chi^2 = 0.34$, $p = 0.84$) among the three groups or CAG repeat length among manifest and premanifest HD participants. A significant difference in age was observed, with both healthy controls and manifest HD patients being significantly older than premanifest, because the controls were recruited to span the entire age range of HD mutation carriers. As expected, there were no differences between controls and premanifest individuals in TFC, TMS,

cUHDRS, SDMT and SWR, but there were differences between premanifest and manifest HD patients (Table 1).

Table 1. Demographics and Baseline Characteristics of Each Cohort.

	Controls (20)	Premanifest (19)	Manifest (40)	ANOVA *p*-value	Control vs Premanifest *p*-value	Premanifest vs Manifest *p*-value
Age (Years)	50.7 ± 11.0	41.8 ± 11.0	56.1 ± 9.4	<0.0001	0.008	<0.0001
Sex (M/F)	10/10	9/10	22/18	NA	NA	NA
CAG	N/A	42.1 ± 1.6	42.7 ± 2.1	NA	NA	0.22
DBS	N/A	265.7 ± 63.3	395.6 ± 94.6	NA	NA	<0.0001
TFC	13 ± 0	13 ± 0	9.4 ± 2.7	<0.0001	1.00	<0.0001
TMS	2.4 ± 2.4	2.5 ± 2.6	37.5 ± 19.4	<0.0001	0.96	<0.0001
cUHDRS	17.4 ± 1.5	18.0 ± 1.0	10.5 ± 3.6	<0.0001	0.46	<0.0001
SDMT	50.9 ± 10.4	55.8 ± 9.5	27.2 ± 12.6	<0.0001	0.18	<0.0001
SWR	100.2 ± 17.4	105.5 ± 11.9	59.6 ± 23.6	<0.0001	0.40	<0.0001

Intergroup differences were assessed using general linear models and Pearson's chi squared test (Gender). P-values are not adjusted for multiple comparisons. Values displayed are mean ±SD unless otherwise stated. DBS, Disease Burden Score; PRE, Premanifest HD mutation carriers; MAN, manifest HD mutation carriers; CAG, CAG triplet repeat count; cUHDRS, composite Unified Huntington's Disease Rating Scale; SDMT, Symbol Digit Modalities Test; SWR, Stroop Word Reading Test; TFC, Total Functional Capacity; TMS, Total Motor Score; NA, not applicable.

Analysis of pre-specified primary analytes

There were no significant differences in protein concentration between genders (LAMP1: Mean Difference (MD) = -0.04, p = 0.75; LAMP2: MD = -0.06, p = 0.61; GM2A: MD = -0.07, p = 0.59; Cathepsin D: MD = -0.07, p = 0.49; Cathepsin F: MD = -0.05, p = 0.54). CSF haemoglobin concentration, used to evaluate effect of blood contamination, displayed no significant associations with any protein (LAMP1: r = 0.16, p = 0.49; LAMP2: r = 0.09, p = 0.70; GM2A: r = 0.15, p = 0.52; Cathepsin D: r = 0.12, p = 0.61; Cathepsin F: r = -0.13, p = 0.59). In addition to significant differences across disease stages, we observed positive trends between CSF protein concentration and age (S1 Fig).

When controlling for age, no significant differences in CSF concentration of LAMP1, LAMP2, GM2A, Cathepsin D or Cathepsin F were observed (group membership main effect: $p = 0.84$; $p = 0.99$, $p = 0.72$; $p = 0.31$; $p = 0.59$, respectively; Fig 1). No significant differences between manifest and premanifest HD patients were observed when also controlling for CAG repeat length (Table 2). Furthermore, we observed no significant differences when grouping together premanifest and manifest HD mutation carriers and comparing with healthy controls (S2 Fig).

Fig 1. Comparison of Analyte Concentration across Disease Stage. No significant differences were observed in the concentration of lysosomal (A) LAMP1, (B) LAMP2, (C) GM2A, (D) Cathepsin (Cat) D and (E) Cathepsin (Cat) F between controls, premanifest and manifest HD patients. P-values were Bonferroni-corrected and generated from general linear models controlling for age. Group membership main effects p-values are displayed in text and Table 2. All CSF protein values have been normalized and log-transformed.

Table 2. Comparison of Analyte Concentration across Disease Stage.

Endo-Lysosomal Proteins	Adjusted for	ANOVA *p* value	Control vs Premanifest *p* value	Manifest vs Premanifest *p* value
LAMP1	Age	0.84	1.00	1.00
	Age and CAG	NA	NA	0.70
LAMP2	Age	0.99	1.00	1.00
	Age and CAG	NA	NA	0.73
GM2A	Age	0.72	0.84	1.00
	Age and CAG	NA	NA	0.64
Cathepsin D	Age	0.31	0.34	0.34
	Age and CAG	NA	NA	0.15
Cathepsin F	Age	0.60	0.65	0.84
	Age and CAG	NA	NA	0.40

Differences in analyte concentration across disease stage were assessed using general linear models controlling for effects of age. P-values are Bonferroni-corrected for multiple comparisons when

required. CAG was also included in the model when assessing differences between manifest and premanifest HD mutation carriers.

Among HD gene expansion carriers, there were no significant correlations between DBS and all measured analytes (Table 3). Furthermore, there were no statistically significant associations between primary analyte concentrations and measures of clinical severity (cUHDRS, TFC, TMS, Fig 2) or cognition (SDMT and SWR, Table 3). Findings remained largely the same when also controlling for CAG repeat length except for LAMP2 which showed a significantly association with TFC (Table 3). Due to LAMP2 demonstrating no significant relationship when controlling for age only, it did not meet our criteria for displaying biomarker potential.

Endo-Lysosomal Proteins	DBS *r* (95% CI)	Adjusted for	cUHDRS *r* (95% CI)	TFC *r* (95% CI)	TMS *r* (95% CI)	SDMT *r* (95% CI)	SWR *r* (95% CI)
LAMP1	0.27 (-0.05, 0.49)	Age	0.11 (-0.19, 0.39)	0.12 (-0.12, 0.37)	-0.08 (-0.34, 0.18)	0.14 (-0.18, 0.43)	0.07 (-0.20, 0.34)
		Age and CAG	0.18 (-0.10, 0.45)	0.17 (-0.07, 0.41)	-0.14 (-0.36. 0.12)	0.20 (-0.09, 0.49)	0.12 (-0.15, 0.39)
LAMP2	0.31 (-0.02, 0.52)	Age	0.13 (-0.16, 0.40)	0.18 (-0.06, 0.41)	-0.10 (-0.34, 0.18)	0.16 (-0.13, 0.44)	0.08 (-0.19, 0.35)
		Age and CAG	0.22 (-0.04, 0.47)	**0.24 (0.01, 0.46)**	-0.16 (-0.40, 0.11)	0.24 (-0.04, 0.49)	0.15 (-0.11, 0.41)
GM2A	0.23 (-0.13, 0.45)	Age	0.13 (-0.14, 0.40)	0.10 (-0.15, 0.34)	-0.14 (-0.35, 0.10)	0.17 (-0.15, 0.47)	0.10 (-0.15, 0.36)
		Age and CAG	0.15 (-0.12, 0.45)	0.11 (-0.13, 0.33)	-0.15 (-0.36, 0.12)	0.19 (-0.09, 0.50)	0.12 (-0.14, 0.39)
Cathepsin D	0.05 (-0.20, 0.26)	Age	0.15 (-0.17, 0.40)	0.11 (-0.17, 0.38)	-0.16 (-0.39, 0.07)	0.13 (-0.16, 0.41)	0.13 (-0.18, 0.39)
		Age and CAG	0.11 (-0.20, 0.39)	0.07 (-0.21, 0.33)	-0.13 (-0.36, 0.11)	0.09 (-0.22, 0.43)	0.10 (-0.22, 0.36)
Cathepsin F	0.25 (-0.07, 0.47)	Age	0.10 (-0.20, 0.36)	0.05 (-0.21, 0.30)	-0.13 (-0.35, 0.15)	0.16 (-0.15, 0.43)	0.09 (-0.22, 0.34)
		Age and CAG	0.13 (-0.16, 0.40)	0.05 (-0.22, 0.31)	-0.14 (-0.37, 0.14)	0.19 (-0.13, 0.48)	0.12 (-0.18, 0.38)

Table 3. Association of Analytes and Assessed Measures in HD Mutation Carriers.

The relationship between protein concentration and Disease Burden Score (DBS) was computed using Pearson's correlation with unadjusted values displayed. Associations with composite Unified Huntington's Disease Rating Scale (cUHDRS), Total Functional Capacity (TFC), Total Motor Score (TMS), Symbol Digit Modalities Test (SDMT), and Stroop Word Reading (SRW) were assessed using Pearson's partial correlation controlling for age, and age and CAG. Correlation coefficients and 95% confidence intervals were computed using bootstrap testing with 1000 repetitions. Results displayed are unadjusted for multiplicity. Bold text indicates $p<0.05$

Fig 2. Correlation between Primary Analyte Concentrations and Clinical Severity. Association within HD gene expansion carriers between CSF LAMP1 (A-C), LAMP2 (D-F), GM2A (G-I), Cathepsin (Cat) D (J-L), Cathepsin (Cat) F and composite Unified Huntington's Disease Rating Scale (cUHDRS), Total Functional Capacity (TFC) and Total Motor Score (TMS). Scatter plots show unadjusted values. Correlation coefficients and 95% confidence intervals were generated using Pearson's partial correlation controlling for age and bootstrapped with 1000 repetitions. All CSF protein values have been normalized and log transformed. Lighter coloured data points represent premanifest individuals.

Exploratory principal component analysis

An exploratory principal components analysis was performed on the entire dataset. The Kaiser-Meyer-Olkin measure of sample adequacy was 0.92 and Bartlett's test of sphericity was significant ($\chi^2(153) = 1485$, $p < 0.001$) indicating that PCA was an appropriate means of dimensionality reduction. The first three components (PC1, PC2 and PC3) had eigenvalues of >1 and explained 75% of the variance in the data (59%, 9% and 7%, respectively). A screeplot demonstrated the 'levelling off' of eigenvalues after three components, thus a three-component solution was selected.

Composite scores were generated for each of the three components allowing for their use in for subsequent analysis. Based on the protein loadings, the three components were deemed to represent lysosomal hydrolases, membrane binding/transfer proteins and innate immune system/peripheral proteins (PC1, PC2 and PC3, respectively) (Fig 3).

Fig 3. Screeplot and Component Loadings Following PCA. (A) Screeplot displays eigenvalues for all components generated. Red line represents an eigenvalue of 1. The first three components have an eigenvalue of >1, thus a three-component solution was adopted. (B) Proteins with loadings of >0.3 were retained and used to define the component labels. All proteins were controlled for age using the residuals from linear regression models. (C) Line graph displaying loadings on the first three components for all proteins included in the PCA. (D) PCA plot demonstrating the clustering of specific proteins into each of the three principal components.

The principal component scores for each participant represent a composite that can be used to examine disease-related alterations across all proteins while avoiding multiple comparisons. We found no significant differences in component scores between genders (PC1, $p = 0.65$; PC2, $p = 0.84$; PC3, $p = 0.47$). When comparing across disease stage, we found no significant differences in PC1, PC2 or PC3 (Fig 4). We observed similar findings when CAG was included in the model (S2 Table).

Fig 4. Group-wise Comparison of Principal Component Scores. No significant differences were observed in Principal component 1 (PC1), 2 (PC2), or 3 (PC3) scores when comparing between healthy controls and GE carriers (A) or across disease stage (B). P-values were Bonferroni-corrected when required and generated from general linear models.

When controlling for age, PC3 demonstrated a significant relationship with TFC only (S3 Table). Composite scores relating to PC1 were not significantly related to any measure of clinical severity or cognition and although PC2 demonstrated a significant relationship with TFC, this relationship was not present when controlling for age only (Fig 5).

Fig 5. Correlation between Principal Component Scores and Measures of Clinical Severity. Association within HD gene expansion carriers between PC1 (A-C), PC2 (D-F), PC3 (G-I), and composite Unified Huntington's Disease Rating Scale (cUHDRS), Total Functional Capacity (TFC) and Total motor score (TMS). Scatter plots show values adjusted for age with correlation coefficients and confidence intervals generated using Pearson's correlation bootstrapped with 1000 repetitions. Red and yellow data points represent manifest and premanifest HD subjects respectively.

Exploratory analysis of remaining analytes

Pearson's correlation revealed only C9 and lysozyme C to be significantly associated with age. Nevertheless, we controlled for age in the subsequent analysis of each protein. Lysozyme C also demonstrated a significant gender difference and thus gender was additionally controlled for when analysing this protein. No significant associations with haemoglobin concentration were observed (S4 Table).

Despite not showing group-wise alterations (S3 Fig), APP, HEXB, UBQ, Cathepsin B and FUCA were significantly associated with measures of clinical severity within HD mutation carriers when controlling for age. Furthermore, these findings remained significant when additionally controlling for CAG repeat length (Fig 6 and Table 4).

Our exploratory analysis of all the remaining endo-lysosomal proteins found no significant differences in analyte concentration across disease stage (S5 Table) or significant relationships with clinical measures, except for C9 and LYZ which displayed significant associations with DBS (S6 Table).

Fig 6. Significant Relationships between Exploratory Proteins and Measures of Clinical Severity. Correlation analysis between analyte concentration and composite Unified Huntington's Disease Rating Scale (cUHDRS), Total Functional Capacity (TFC) and Total Motor Score (TMS) revealed significant associations between all three analytes and measures of clinical severity. Scatter plots show unadjusted values. Correlation coefficients and 95% confidence intervals were generated using Pearson's partial correlation controlling for age and bootstrapped with 1000 repetitions. All CSF protein values have been normalized and log transformed. Lighter coloured data points represent premanifest individuals. Bold text indicates significance at $p<0.05$.

Table 4. Significant Associations between Exploratory Analytes and Assessed Measures in HD Mutation Carriers.

Endo-Lysosomal Proteins	DBS *r* (95% CI)	Adjusted for	cUHDRS *r* (95% CI)	TFC *r* (95% CI)	TMS *r* (95% CI)	SDMT *r* (95% CI)	SWR *r* (95% CI)
APP	-0.03 (-0.30, 0.24)	Age	**0.34 (0.08 0.57)**	**0.30 (0.07, 0.52)**	**-0.30 (-0.50, -0.07)**	**0.37 (0.09, 0.60)**	**0.32 (0.06, 0.53)**
		Age and CAG	**0.34 (0.06, 0.56)**	**0.27 (0.05, 0.49)**	**-0.27 (-0.46, -0.02)**	**0.36 (0.10, 0.59)**	**0.30 (0.01, 0.53)**
HEXB	0.11 (-0.16, 0.40)	Age	**0.26 (0.01, 0.48)**	**0.22 (0.02, 0.42)**	**-0.22 (-0.43, -0.01)**	**0.27 (0.01, 0.50)**	0.24 (0.00, 0.47)
		Age and CAG	**0.30 (0.04, 0.56)**	**0.23 (0.01, 0.45)**	**-0.23 (-0.45, -0.02)**	0.29 (-0.01, 0.59)	0.26 (-0.04, 0.50)
UBQ	0.13 (-0.17, 0.35)	Age	0.27 (-0.02, 0.53)	**0.27 (0.03, 0.47)**	-0.22 (-0.44, 0.03)	0.28 (-0.02, 0.56)	0.22 (-0.07, 0.47)
		Age and CAG	0.31 (0.04, 0.55)	**0.28 (0.04, 0.50)**	-0.24 (-0.45, 0.02)	**0.32 (0.03, 0.56)**	0.24 (-0.09, 0.49)
Cathepsin B	0.12 (-0.17, 0.36)	Age	0.25 (-0.03, 0.47)	0.21 (-0.28, 0.43)	**-0.26 (-0.45, -0.06)**	0.22 (-0.06, 0.46)	0.22 (-0.05, 0.47)
		Age and CAG	**0.30 (0.02, 0.55)**	0.23 (-0.02, 0.45)	**-0.30 (-0.54, -0.07)**	0.25 (-0.03, 0.53)	0.26 (-0.03, 0.53)
FUCA	0.14 (-0.21, 0.40)	Age	0.20 (-0.05, 0.41)	**0.30 (0.07, 0.51)**	-0.17 (-0.36, 0.05)	0.18 (-0.08, 0.42)	0.15 (-0.10, 0.36)
		Age and CAG	0.23 (-0.03, 0.47)	**0.23 (0.07, 0.51)**	-0.18 (-0.40, 0.50)	0.19 (-0.11, 0.50)	0.16 (-0.12, 0.42)

The relationships between exploratory analytes and Disease Burden Score (DBS) were assessed using Pearson's correlation with unadjusted values shown. Relationships with, composite Unified Huntington's Disease Rating Scale (cUHDRS), Total Functional Capacity (TFC), Total Motor Score (TMS), Symbol Digit Modalities Test (SDMT), and Stroop Word Reading (SRW) were assessed using Pearson's partial correlation controlling for age, and age and CAG. Correlation coefficients and 95% confidence intervals were computed using bootstrap testing with 1000 repetitions. Results displayed are unadjusted multiplicity. Bold text indicates significance at $p<0.05$.

Discussion

In this cross-sectional study, we successfully quantified 18 endo-lysosomal proteins in high-quality CSF obtained under strictly standardised conditions, from HD mutation carriers and controls, by condensing peptide-level data from 48 peptides

quantified using mass spectrometry. Our pre-specified analysis of the five endo-lysosomal proteins most likely to show relevant HD-related alterations (Cathepsin D, Cathepsin F, GM2A, LAMP1 and LAMP2) found no discernible differences in concentration between HD mutation carriers and controls. Nor, did we observe any significant relationships between the concentrations of these proteins and measurements of clinical severity or cognition. These findings were supported by an exploratory unbiased PCA of the entire dataset which also showed no groupwise differences in three principal components. The findings of our exploratory analysis of the remaining 13 proteins, were also negative for group-wise differences. However, we observed significant negative associations between CSF APP and all measures of clinical severity and cognitive decline within HD mutation carriers, suggesting that APP, and its cleaved product beta-amyloid (Aβ), may be an important avenue to be explored in HD.

Lower levels of CSF APP were associated with worse clinical phenotype and lower cognitive performance. The strongest relationship was observed with cUHDRS score, a powerful measure of clinical progression that predicts corticostriatal atrophy [48]; this relationship, and all others tested, remained significant when controlling for both age and CAG, indicating that there is predictive value independent from well-known predictors of HD progression [50]. APP is a transmembrane protein with multiple physiological functions, including regulating brain iron homeostasis [56]. In HD, mHTT expression has been linked to brain iron accumulation, particularly within neurons [57], potentially exacerbating disease pathology via reactive oxygen species production and oxidative stress [58]. APP is known to facilitate neuronal iron export [56] and has been shown to be decreased in the R6/2 mouse brain [59]. It has been hypothesised that an inadequate APP response to brain iron accumulation may

contribute to iron homeostatic dysfunction [60]. The association between reduced CSF APP and clinical worsening in this study provides some support for APP dysfunction in HD and a possible impact on disease progression.

APP is cleaved by β- and γ-secretase to form Aβ peptides [61,62]. Although we are not measuring Aβ in this study, our findings also raise interesting questions regarding the biomarker potential of CSF Aβ, a biomarker most associated with AD [63], in HD. Reduced CSF Aβ is well described in the AD literature [64–67], likely as a result of increased amyloid deposition in the brain and reduced clearance into the CSF [68]. The CSF level often demonstrates an inverse relationship with whole brain amyloid load and CSF tau concentration [67,69,70]. However, Aβ in CSF has not been studied in HD to our knowledge. Though amyloid deposition is not a typical feature of HD pathology, our APP findings suggest it is possible that Aβ could also represent a novel monitoring or prognostic biomarker in HD.

Similarly, we observed reduced levels of beta-hexosaminidase-β and Cathepsin B tended to predict a more severe clinical phenotype. Cathepsin B is a lysosomal cysteine protease implicated in the pathology of several neurodegenerative diseases, most notably AD [71] in which it has been shown to contribute to increased Aβ load [72], yet also offers potential neuroprotective and anti-amyloidogenic properties [73,74]. Contrary to previous studies demonstrating increased levels of CSF and plasma Cathepsin B in PD and AD respectively [27,75], we found Cathepsin B to offer little value as a state biomarker in HD. However, given its significant relationship with TMS and previous work showing reduced mHTT in response to CTSB overexpression [40], it may possess potential for monitoring disease progression.

Together with the co-factor GM2 activator protein (GM2A), beta-hexosaminidase-β is responsible for the degradation of ganglioside GM2 [76]. Mutations in *HEXB*, resulting in reduced levels of the β-subunit and subsequent accumulation of GM2 in neuronal tissue, are the cause of three fatal, neurodegenerative disorders known as the GM2 Gangliosidoses [77]. In this study, we did not observe any differences in beta-hexosaminidase-β across disease stage; however, given its strong association with all three measures of clinical severity, and the reported dysfunction in lipid synthesis, metabolism and catabolism in HD [29,30], CSF beta-hexosaminidase-β represents an interesting avenue for future research and could help shed light on the role of generalised lysosomal dysfunction in HD pathogenesis.

Furthermore, we observed significant relationships between ubiquitin and complement component C9 and measures of clinical severity. The ubiquitin-proteasome system (UPS) is a key mechanism of intracellular protein clearance, in which misfolded proteins are polyubiquitinated by ligases, thus targeting the substrate for degradation [78,79]. Previous proteomic work has demonstrated differences in CSF ubiquitin levels between HD patients and controls, whilst also showing a negative relationship with TFC [80]. However, we did not observe any discernible group differences and found lower CSF ubiquitin to be indicative of worsening clinical phenotype. Given these contrary findings and the abundance of literature implicating UPS alterations in the context of Huntington's disease [81–84], further exploration of CSF ubiquitin in HD is required. C9 is a constituent protein of the innate immune system and is highly expressed by astrocytes, microglia and neurons [85–87]. In HD, mHTT activates the complement system resulting in a cascade of neuroinflammatory responses [88]. Neuroinflammation remains a promising area in the field of biomarker research with additional complement

components shown to be upregulated in the plasma of HD patients [89] and CSF YKL-40, a microglial marker, showing disease related elevations and the ability to independently predict clinical severity and neuronal death [90]. We found increased levels of C9 and Lysozyme C (LYZ), another cornerstone of innate immunity, to be associated with a higher DBS. This finding was strengthened by our PCA results in which a single component (PC3) correlated negatively with TFC when controlling for age. Interestingly, the protein which loaded highest onto this component was C9, with LYZ also loading highly, thus further supporting the involvement of the innate immune system in HD.

By measuring several peptides per protein, a more accurate approximation of the abundance of the intact protein can be obtained. Our decision to combine the peptides was influenced by our desire to generate an accurate protein-level estimate. However, it should be noted that individual peptides can be derived from different endogenous fragments of the protein or may belong to different functional domains, therefore there is value in studying individual peptides in future studies.

Our study has some limitations that should be acknowledged. First, the cross-sectional nature of this study means we cannot fully understand how the measured analytes may vary with disease progression; to do this requires longitudinal data collection. Secondly, HD-CSF was principally designed to study manifest HD, so it has a relatively small number of premanifest HD and control subjects. Future studies should recruit larger numbers of subjects within these groups to help improve generalisability of results across the entire disease course. The HDClarity CSF collection initiative [91] represents a large collection of CSF with longitudinal repeat sampling underway. Furthermore, patients with juvenile HD were not recruited in HD-CSF; thus we cannot extend our findings to this sub-population of HD mutation

carriers. Finally, all CSF sampling visits were undertaken at the same time of day following an overnight fast; while this minimises the effect of diurnal variation and diet, it may limit the generalisability of our findings.

In conclusion, out of 5 primary and 13 exploratory endo-lysosomal proteins derived from CSF, we could find no alterations in HD patients compared with healthy controls. In our exploratory analyses, we found interesting associations with disease severity for several proteins of potential pathogenic relevance namely HEXB, Cathepsin B, UBQ, C9 and perhaps most notably, APP. These observations link HD severity to several mechanisms, including lipid catabolism deficits, proteostasis network dysfunction, enhanced neuroinflammatory response and dysregulation of iron homeostasis, and suggest a means for beginning to explore these pathways quantitatively in mutation carriers.

Our overall negative groupwise findings in CSF do not exclude a role of lysosomal dysfunction in the pathogenesis of HD; only that major discernible differences in their concentrations could not be observed in the CSF of HD patients. It remains likely that the endo-lysosomal/autophagy system is implicated in the pathology of, and CNS response to, Huntington's disease. However, our work suggests that endo-lysosomal proteins measured in human CSF are unlikely to be state biomarkers in HD but may show promise as tools for exploring pathways of interest and as pharmacodynamic markers for future drug candidates targeting this system.

Acknowledgements

We would like to thank all the participants from the HD community who donated samples and gave their time to take part in this study.

References

1. Ross CA, Aylward EH, Wild EJ, Langbehn DR, Long JD, Warner JH, et al. Huntington disease: Natural history, biomarkers and prospects for therapeutics. Nat Rev Neurol. 2014;10(4):204–16.

2. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell. 1993;72(6):971–83.

3. Ross CA. Polyglutamine pathogenesis: Emergence of unifying mechanisms for Huntington's disease and related disorders. Neuron. 2002;35(5):819–22.

4. Menzies FM, Moreau K, Rubinsztein DC. Protein misfolding disorders and macroautophagy. Curr Opin Cell Biol. 2011;23(2):190–7.

5. Koga H, Cuervo AM. Chaperone-mediated autophagy dysfunction in the pathogenesis of neurodegeneration. Neurobiol Dis. 2011;43(1):29–37.

6. Wong E, Cuervo AM. Autophagy gone awry in neurodegenerative diseases. Nat Neurosci. 2010;13(7):805–11.

7. Martinez-Vicente M, Cuervo AM. Autophagy and neurodegeneration: when the cleaning crew goes on strike. Lancet Neurol. 2007;6(4):352–61.

8. Zhang L, Sheng R, Qin Z. Review: The lysosome and neurodegenerative diseases. Acta Biochim Biophys Sin (Shanghai). 2009;41(6):437–45.

9. Martinez-Vicente M, Talloczy Z, Kaushik S, Massey AC, Mazzulli J, Mosharov

E V., et al. Dopamine-modified α-synuclein blocks chaperone-mediated autophagy. J Clin Invest. 2008;118(2):777–8.

10. Wang Y, Martinez-Vicente M, Krüger U, Kaushik S, Wong E, Mandelkow EM, et al. Tau fragmentation, aggregation and clearance: The dual role of lysosomal processing. Hum Mol Genet. 2009;18(21):4153–70.

11. Koga H, Martinez-Vicente M, Arias E, Kaushik S, Sulzer D, Cuervo AM. Constitutive upregulation of chaperone-mediated autophagy in Huntington's disease. J Neurosci. 2011;31(50):18492–505.

12. Martinez-Vicente M, Talloczy Z, Wong E, Tang G, Koga H, Kaushik S, et al. Cargo recognition failure is responsible for inefficient autophagy in Huntington's disease. Nat Neurosci. 2010;13(5):567–76.

13. Hara T, Nakamura K, Matsui M, Yamamoto A, Nakahara Y, Suzuki-Migishima R, et al. Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice. Nature. 2006;441(7095):885–9.

14. Eskelinen EL. Roles of LAMP-1 and LAMP-2 in lysosome biogenesis and autophagy. Mol Aspects Med. 2006;27(5–6):495–502.

15. Huynh KK, Eskelinen EL, Scott CC, Malevanets A, Saftig P, Grinstein S. LAMP proteins are required for fusion of lysosomes with phagosomes. EMBO J. 2007;26(2):313–24.

16. Pang S, Chen D, Zhang A, Qin X, Yan B. Genetic analysis of the LAMP-2 gene promoter in patients with sporadic Parkinson's disease. Neurosci Lett. 2012;526(1):63–7.

17. Alvarez-Erviti L, Rodriguez-Oroz MC, Cooper JM, Caballero C, Ferrer I, Obeso

627 JA, et al. Chaperone-mediated autophagy markers in Parkinson disease
628 brains. Arch Neurol. 2010;67(12):1464–72.

629 18. Armstrong A, Mattsson N, Appelqvist H, Janefjord C, Sandin L, Agholme L, et
630 al. Lysosomal network proteins as potential novel CSF biomarkers for
631 Alzheimer's disease. NeuroMolecular Med. 2014;16(1):150–60.

632 19. Youn J, Lee S Bin, Lee HS, Yang HO, Park J, Kim JS, et al. Cerebrospinal
633 Fluid Levels of Autophagy-related Proteins Represent Potentially Novel
634 Biomarkers of Early-Stage Parkinson's Disease. Sci Rep. 2018;8(1).

635 20. Sjödin S, Öhrfelt A, Brinkmalm G, Zetterberg H, Blennow K, Brinkmalm A.
636 Targeting LAMP2 in human cerebrospinal fluid with a combination of
637 immunopurification and high resolution parallel reaction monitoring mass
638 spectrometry. Clin Proteomics. 2016;13(1):1–14.

639 21. Buerger K, Ewers M, Pirttilä T, Zinkowski R, Alafuzoff I, Teipel SJ, et al. CSF
640 phosphorylated tau protein correlates with neocortical neurofibrillary pathology
641 in Alzheimer's disease. Brain. 2006;129(11):3035–41.

642 22. Block RC, Dorsey ER, Beck CA, Brenna JT, Shoulson I. Altered cholesterol
643 and fatty acid metabolism in Huntington disease. J Clin Lipidol. 2010;4(1):17–
644 23.

645 23. Cortes CJ, La Spada AR. The many faces of autophagy dysfunction in
646 Huntington's disease: From mechanism to therapy. Drug Discov Today.
647 2014;19(7):963–71.

648 24. Kolter T, Sandhoff K. Lysosomal degradation of membrane lipids. FEBS Lett.
649 2010;584(9):1700–12.

650 25. Cordeiro P, Hechtman P, Kaplan F. The GM2 gangliosidoses databases:
651 Allelic variation at the HEXA, HEXB, and GM2A gene loci. Genet Med.
652 2000;2(6):319–27.

653 26. Heywood WE, Galimberti D, Bliss E, Sirka E, Paterson RW, Magdalinou NK, et
654 al. Identification of novel CSF biomarkers for neurodegeneration and their
655 validation by a high-throughput multiplexed targeted proteomic assay. Mol
656 Neurodegener. 2015;10(1):1–16.

657 27. Sjödin S, Brinkmalm G, Öhrfelt A, Parnetti L, Paciotti S, Hansson O, et al.
658 Endo-lysosomal proteins and ubiquitin CSF concentrations in Alzheimer's and
659 Parkinson's disease. Alzheimer's Res Ther. 2019;11(1).

660 28. Manwaring V, Heywood WE, Clayton R, Lachmann RH, Keutzer J, Hindmarsh
661 P, et al. The identification of new biomarkers for identifying and monitoring
662 kidney disease and their translation into a rapid mass spectrometry-based test:
663 Evidence of presymptomatic kidney disease in pediatric fabry and type-I
664 diabetic patients. J Proteome Res. 2013;12(5):2013–21.

665 29. Desplats PA, Denny CA, Kass KE, Gilmartin T, Head SR, Sutcliffe JG, et al.
666 Glycolipid and ganglioside metabolism imbalances in Huntington's disease.
667 Neurobiol Dis. 2007;27(3):265–77.

668 30. Maglione V, Marchi P, Di Pardo A, Lingrell S, Horkey M, Tidmarsh E, et al.
669 Impaired ganglioside metabolism in Huntington's disease and neuroprotective
670 role of GM1. J Neurosci. 2010;30(11):4072–80.

671 31. Di Pardo A, Maglione V, Alpaugh M, Horkey M, Atwal RS, Sassone J, et al.
672 Ganglioside GM1 induces phosphorylation of mutant huntingtin and restores
673 normal motor behavior in Huntington disease mice. Proc Natl Acad Sci U S A.

2012;109(9):3528–33.

32. Schnaar RL. Brain gangliosides in axon-myelin stability and axon regeneration. FEBS Lett. 2010;584(9):1741–7.

33. Zhang J, Gregory S, Scahill RI, Durr A, Thomas DL, Lehericy S, et al. In vivo characterization of white matter pathology in premanifest huntington's disease. Ann Neurol. 2018;84(4):497–504.

34. Rosas HD, Tuch DS, Hevelone ND, Zaleta AK, Vangel M, Hersch SM, et al. Diffusion tensor imaging in presymptomatic and early Huntington's disease: Selective white matter pathology and its relationship to clinical measures. Mov Disord. 2006;21(9):1317–25.

35. Gregory S, Johnson E, Byrne LM, Rodrigues FB, Henderson A, Moss J, et al. Characterizing White Matter in Huntington's Disease. Mov Disord Clin Pract. 2020;7(1):52–60.

36. Stoka V, Turk V, Turk B. Lysosomal cathepsins and their regulation in aging and neurodegeneration. Ageing Res Rev. 2016;32:22–37.

37. Bhutani N, Piccirillo R, Hourez R, Venkatraman P, Goldberg AL. Cathepsins L and Z are critical in degrading polyglutamine-containing proteins within lysosomes. J Biol Chem. 2012;287(21):17471–82.

38. MANTLE D, FALKOUS G, ISHIURA S, PERRY RH, PERRY EK. Comparison of Cathepsin Protease Activities in Brain-Tissue From Normal Cases and Cases With Alzheimers-Disease, Lewy Body Dementia, Parkinsons-Disease and Huntingtons-Disease. J Neurol Sci [Internet]. 1995;131:65–70. Available from:

http://apps.isiknowledge.com.proxy.bib.uottawa.ca/full_record.do?product=WOS&search_mode=GeneralSearch&qid=1&SID=2F1g4g79DbbOIlFfbiA&page=1&doc=1

39. Kim YJ, Sapp E, Cuiffo BG, Sobin L, Yoder J, Kegel KB, et al. Lysosomal proteases are involved in generation of N-terminal huntingtin fragments. Neurobiol Dis. 2006;22(2):346–56.

40. Liang Q, Ouyang X, Schneider L, Zhang J. Reduction of mutant huntingtin accumulation and toxicity by lysosomal cathepsins D and B in neurons. Mol Neurodegener. 2011;6(1):1–12.

41. Reiber H. Proteins in cerebrospinal fluid and blood: Barriers, CSF flow rate and source-related dynamics. Restor Neurol Neurosci. 2003;21(3–4):79–96.

42. Byrne LM, Wild EJ. Cerebrospinal Fluid Biomarkers for Huntington's Disease. J Huntingtons Dis. 2016;5(1):1–13.

43. Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. Mol Cell Proteomics. 2012;11(11):1475–88.

44. Gallien S, Duriez E, Demeure K, Domon B. Selectivity of LC-MS/MS analysis: Implication for proteomics experiments. J Proteomics. 2013;81:148–58.

45. Byrne LM, Rodrigues FB, Johnson EB, De Vita E, Blennow K, Scahill R, et al. Cerebrospinal fluid neurogranin and TREM2 in Huntington's disease. Sci Rep. 2018;8(1):1–7.

46. Byrne LM, Rodrigues FB, Johnson EB, Wijeratne PA, De Vita E, Alexander DC, et al. Evaluation of mutant huntingtin and neurofilament proteins as

720 potential markers in Huntington's disease. Sci Transl Med. 2018;10(458):1–11.

721 47. Huntington Study Group. Unified Huntington's Disease Rating Scale: Reliability
722 and-Consistency. Mov Disord. 1996;11:136–42.

723 48. Schobel SA, Palermo G, Auinger P, Long JD, Ma S, Khwaja OS, et al. Motor,
724 cognitive, and functional declines contribute to a single progressive factor in
725 early HD. Neurology. 2017;89(24):2495–502.

726 49. Penney JB, Vonsattel JP, MacDonald ME, Gusella JF, Myers RH. CAG repeat
727 number governs the development rate of pathology in huntington's disease.
728 Ann Neurol. 1997;41(5):689–92.

729 50. Tabrizi SJ, Scahill RI, Owen G, Durr A, Leavitt BR, Roos RA, et al. Predictors
730 of phenotypic progression and disease onset in premanifest and early-stage
731 Huntington's disease in the TRACK-HD study: Analysis of 36-month
732 observational data. Lancet Neurol. 2013;12(7):637–49.

733 51. Brinkmalm G, Sjödin S, Simonsen AH, Hasselbalch SG, Zetterberg H,
734 Brinkmalm A, et al. A Parallel Reaction Monitoring Mass Spectrometric Method
735 for Analysis of Potential CSF Biomarkers for Alzheimer's Disease. Proteomics
736 - Clin Appl. 2018;12(1).

737 52. De Graaf EL, Altelaar AFM, Van Breukelen B, Mohammed S, Heck AJR.
738 Improving SRM assay development: A global comparison between triple
739 quadrupole, Ion trap, and higher energy CID peptide fragmentation spectra. J
740 Proteome Res. 2011;10(9):4334–41.

741 53. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B,
742 et al. Skyline: An open source document editor for creating and analyzing

743 targeted proteomics experiments. Bioinformatics. 2010;26(7):966–8.

744 54. Joliffe IT, Morgan B. Principal component analysis and exploratory factor
745 analysis. Stat Methods Med Res. 1992;1(1):69–95.

746 55. Preacher KJ, MacCallum RC. Exploratory factor analysis in behavior genetics
747 research: Factor recovery with small sample sizes. Behav Genet.
748 2002;32(2):153–61.

749 56. Duce JA, Tsatsanis A, Cater MA, James SA, Robb E, Wikhe K, et al. Iron-
750 Export Ferroxidase Activity of β-Amyloid Precursor Protein is Inhibited by Zinc
751 in Alzheimer's Disease. Cell. 2010;142(6):857–67.

752 57. Rosas HD, Chen YI, Doros G, Salat DH, Chen NK, Kwong KK, et al.
753 Alterations in brain transition metals in Huntington disease: An evolving and
754 intricate story. Arch Neurol. 2012;69(7):887–93.

755 58. Ward RJ, Zucca FA, Duyn JH, Crichton RR, Zecca L. The role of iron in brain
756 ageing and neurodegenerative disorders. Lancet Neurol. 2014;13(10):1045–
757 60.

758 59. Fox JH, Kama JA, Lieberman G, Chopra R, Dorsey K, Chopra V, et al.
759 Mechanisms of copper ion mediated Huntington's disease progression. PLoS
760 One. 2007;2(3):1–12.

761 60. Berggren K, Agrawal S, Fox JA, Hildenbrand J, Nelson R, Bush AI, et al.
762 Amyloid precursor protein haploinsufficiency preferentially mediates brain iron
763 accumulation in mice transgenic for the Huntington's disease mutation. J
764 Huntingtons Dis. 2017;6(2):115–25.

765 61. Müller UC, Zheng H. Physiological functions of APP family proteins. Cold

766 Spring Harb Perspect Med. 2012;2(2):1–18.

767 62. Haass C, Kaether C, Thinakaran G, Sisodia S. Trafficking and proteolytic
768 processing of APP. Cold Spring Harb Perspect Med. 2012;2(5):1–25.

769 63. Blennow K, Hampel H, Weiner M, Zetterberg H. Cerebrospinal fluid and
770 plasma biomarkers in Alzheimer disease. Nat Rev Neurol. 2010;6(3):131–44.

771 64. Hampel H, Teipel SJ, Fuchsberger T, Andreasen N, Wiltfang J, Otto M, et al.
772 Value of CSF β-amyloid 1-42 and tau as predictors of Alzheimer's disease in
773 patients with mild cognitive impairment. Mol Psychiatry. 2004;9(7):705–10.

774 65. Moonis M, Swearer JM, Dayaw MPE, St. George-Hyslop P, Rogaeva E,
775 Kawarai T, et al. Familial Alzheimer disease: Decreases in CSF Aβ42 levels
776 precede cognitive decline. Neurology. 2005;65(2):323–5.

777 66. Sunderland T, Linker G, Mirza N, Putnam KT, Friedman DL, Kimmel LH, et al.
778 Decreased β-Amyloid1-42 and Increased Tau Levels in Cerebrospinal Fluid of
779 Patients with Alzheimer Disease. J Am Med Assoc. 2003;289(16):2094–103.

780 67. Grimmer T, Riemenschneider M, Förstl H, Henriksen G, Klunk WE, Mathis CA,
781 et al. Beta Amyloid in Alzheimer's Disease: Increased Deposition in Brain Is
782 Reflected in Reduced Concentration in Cerebrospinal Fluid. Biol Psychiatry.
783 2009;65(11):927–34.

784 68. Weller RO. How well does the CSF inform upon pathology in the brain in
785 Creutzfeldt-Jakob and Alzheimer's diseases? J Pathol. 2001;194(1):1–3.

786 69. Seppälä TT, Nerg O, Koivisto AM, Rummukainen J, Puli L, Zetterberg H, et al.
787 CSF biomarkers for Alzheimer disease correlate with cortical brain biopsy
788 findings. Neurology. 2012;78(20):1568–75.

789 70. Palmqvist S, Zetterberg H, Mattsson N, Johansson P, Minthon L, Blennow K,
790 et al. Detailed comparison of amyloid PET and CSF biomarkers for identifying
791 early Alzheimer disease. Neurology. 2015;85(14):1240–9.

792 71. Cataldo AM, Barnett JL, Pieroni C, Nixon RA. Increased neuronal endocytosis
793 and protease delivery to early endosomes in sporadic Alzheimer's disease:
794 Neuropathologic evidence for a mechanism of increased β-amyloidogenesis. J
795 Neurosci. 1997;17(16):6142–51.

796 72. Klein DM, Felsenstein KM, Brenneman DE. Cathepsins B and L differentially
797 regulate amyloid precursor protein processing. J Pharmacol Exp Ther.
798 2009;328(3):813–21.

799 73. Sun B, Zhou Y, Halabisky B, Lo I, Cho SH, Mueller-Steiner S, et al. Cystatin C-
800 Cathepsin B Axis Regulates Amyloid Beta Levels and Associated Neuronal
801 Deficits in an Animal Model of Alzheimer's Disease. Neuron. 2008;60(2):247–
802 57.

803 74. Mueller-Steiner S, Zhou Y, Arai H, Roberson ED, Sun B, Chen J, et al.
804 Antiamyloidogenic and Neuroprotective Functions of Cathepsin B: Implications
805 for Alzheimer's Disease. Neuron. 2006;51(6):703–14.

806 75. Sundelöf J, Sundström J, Hansson O, Eriksdotter-Jönhagen M, Giedraitis V,
807 Larsson A, et al. Higher cathepsin B levels in plasma in Alzheimer's disease
808 compared to healthy controls. J Alzheimer's Dis. 2010;22(4):1223–30.

809 76. Bateman KS, Cherney MM, Mahuran DJ, Tropak M, James MNG. Crystal
810 structure of β-hexosaminidase B in complex with pyrimethamine, a potential
811 pharmacological chaperone. J Med Chem. 2011;54(5):1421–9.

812 77. Mahuran DJ. Biochemical consequences of mutations causing the GM2
813 gangliosidoses. Biochim Biophys Acta - Mol Basis Dis. 1999;1455(2–3):105–
814 38.

815 78. Ciechanover A, Brundin P. The ubiquitin proteasome system in
816 neurodegenerative diseases: Sometimes the chicken, sometimes the egg.
817 Neuron. 2003;40(2):427–46.

818 79. DeMartino GN, Gillette TG. Proteasomes: Machines for All Reasons. Cell.
819 2007;129(4):659–62.

820 80. Vinther-Jensen T, Simonsen AH, Budtz-Jørgensen E, Hjermind LE, Nielsen
821 JE. Ubiquitin: A potential cerebrospinal fluid progression marker in
822 Huntington's disease. Eur J Neurol. 2015;22(10):1378–84.

823 81. Wang J, Wang CE, Orr A, Tydlacka S, Li SH, Li XJ. Impaired ubiquitin-
824 proteasome system activity in the synapses of Huntington's disease mice. J
825 Cell Biol. 2008;180(6):1177–89.

826 82. Hunter JM, Lesort M, Johnson GVW. Ubiquitin-proteasome system alterations
827 in a striatal cell model of Huntington's disease. J Neurosci Res.
828 2007;85(8):1774–88.

829 83. Bennett EJ, Shaler TA, Woodman B, Ryu KY, Zaitseva TS, Becker CH, et al.
830 Global changes to the ubiquitin system in Huntington's disease. Nature.
831 2007;448(7154):704–8.

832 84. Ortega Z, Lucas JJ. Ubiquitin-proteasome system involvement in huntington's
833 disease. Front Mol Neurosci. 2014;7(SEP).

834 85. Sarma JV, Ward PA. The complement system. Cell Tissue Res.

2011;343(1):227–35.

86. Bonifati DM, Kishore U. Role of complement in neurodegeneration and neuroinflammation. Mol Immunol. 2007;44(5):999–1010.

87. Gasque P, Dean YD, McGreal EP, Vanbeek J, Morgan BP. Complement components of the innate immune system in health and disease in the CNS. Immunopharmacology. 2000;49(1–2):171–86.

88. Janssen BJC, Huizinga EG, Raaijmakers HCA, Roos A, Daha MR, Nilsson-Ekdahl K, et al. Structures of complement component C3 provide insights into the function and evolution of immunity. Nature. 2005;437(7058):505–11.

89. Dalrymple A, Wild EJ, Joubert R, Sathasivam K, Björkqvist M, Petersén Å, et al. Proteomic profiling of plasma in Huntington's disease reveals neuroinflammatory activation and biomarker candidates. J Proteome Res. 2007;6(7):2833–40.

90. Rodrigues FB, Byrne LM, McColgan P, Robertson N, Tabrizi SJ, Zetterberg H, et al. Cerebrospinal fluid inflammatory biomarkers reflect clinical severity in huntington's disease. PLoS One. 2016;11(9):1–10.

91. Wild EJ, Borowsky B Hdc investigators. HDClarity: A new multi-site cerebrospinal fluid collection initiative to facilitate therapeutic development for Huntington's disease. In: 11th Annual CHDI HD Therapeutics Conference. Palm Spring, CA; 2016.

Supporting information

S1 Table. Complete List of Lysosomal Proteins. Information pertaining to all CSF endo-lysosomal proteins used in the study.

S1 Fig. Correlation Analysis between Main CSF Analytes and Age. Pearson's correlation revealed positive trends between the concentrations of lysosomal (A) LAMP1, (B) LAMP2, (C) GM2A, (D) Cathepsin (Cat) D, (E) Cathepsin (Cat) F and age in healthy controls. All CSF protein values have been normalized and log transformed.

S2 Fig. Comparison of Analyte Concentration between Gene Expansion Carriers and Controls. No significant differences in the concentration of lysosomal (A) LAMP1, (B) LAMP2, (C) GM2A, (D) Cathepsin (Cat) D and (E) Cathepsin (Cat) F, was observed between healthy controls and GE carriers. All CSF protein values have been normalized and log transformed. P-values were Bonferroni-corrected and generated from general linear models controlling for age.

S2 Table. Comparison of Principal Component Scores across Disease Stage. Differences in scores across disease stage. P-values were Bonferroni-corrected and generated from general linear models. CAG was included in the model when assessing differences between manifest and premanifest HD mutation carriers.

S3 Table. Association of Principal Components and Assessed Measures in HD Mutation Carriers. The relationships between principal components 1 (PC1), 2 (PC2) and 3 (PC3) and Disease Burden Score (DBS) were computed using Pearson's correlation with unadjusted values shown. Relationships with composite Unified Huntington's Disease Rating Scale (cUHDRS), Total Functional Capacity (TFC), Total Motor Score (TMS), Symbol Digit Modalities Test (SDMT), and Stroop Word Reading (SWR) were assessed using Pearson's partial correlation controlling for age, and age and CAG. Correlation coefficients and 95% confidence intervals were computed using bootstrap testing with 1000 repetitions. Results shown are unadjusted for multiplicity. Bold text indicates significance at $p<0.05$.

S4 Table. Assessments for Potential Confounding Variables in Exploratory Proteins. Values are Pearson's r and t-test statistic. Bold indicates significance at the $p<0.05$ level.

884 **S3 Fig. Comparison of Exploratory Analyte Concentration across Disease**

885 **Stage.** We observed no significant differences across disease stage in APP, HEBX, UBQ,

886 Cathepsin (Cat) B and FUCA**.** P-values were Bonferroni-corrected and generated from general linear

887 models controlling for age.

888 **S5 Table. Exploratory Comparison of Additional Proteins across Disease**

889 **Stage.** Differences in analyte concentration across disease stage were Bonferroni-corrected and

890 generated from general linear models controlling for age, or age and CAG. Gender was also included

891 in the model for LYZ. Bold indicates significance at the p<0.05 level.

892 **S6 Table. Exploratory Correlation Analysis between Lysosomal Proteins and**

893 **Measures of Clinical Severity and Cognition.** Associations between analyte concentration

894 and Disease Burden Score (DBS) were assessed using Pearson's correlation with unadjusted values

895 displayed. Associations with composite Unified Huntington's Disease Rating Scale (cUHDRS), Total

896 Functional Capacity (TFC), Total Motor Score (TMS), Symbol Digit Modalities Test (SDMT), and

897 Stroop Word Reading (SWR) were assessed using partial correlation with age, and age and CAG

898 included in the model. For LYZ, the effects of gender were also controlled for. Significant associations

899 are highlighted in bold. Correlation coefficients and confidence intervals were both generated using

900 bootstrapping with 1000 repetitions. Bold indicates significance at the p<0.05 level.

901

902

903

904

905

906

907

908

909

910

911

912

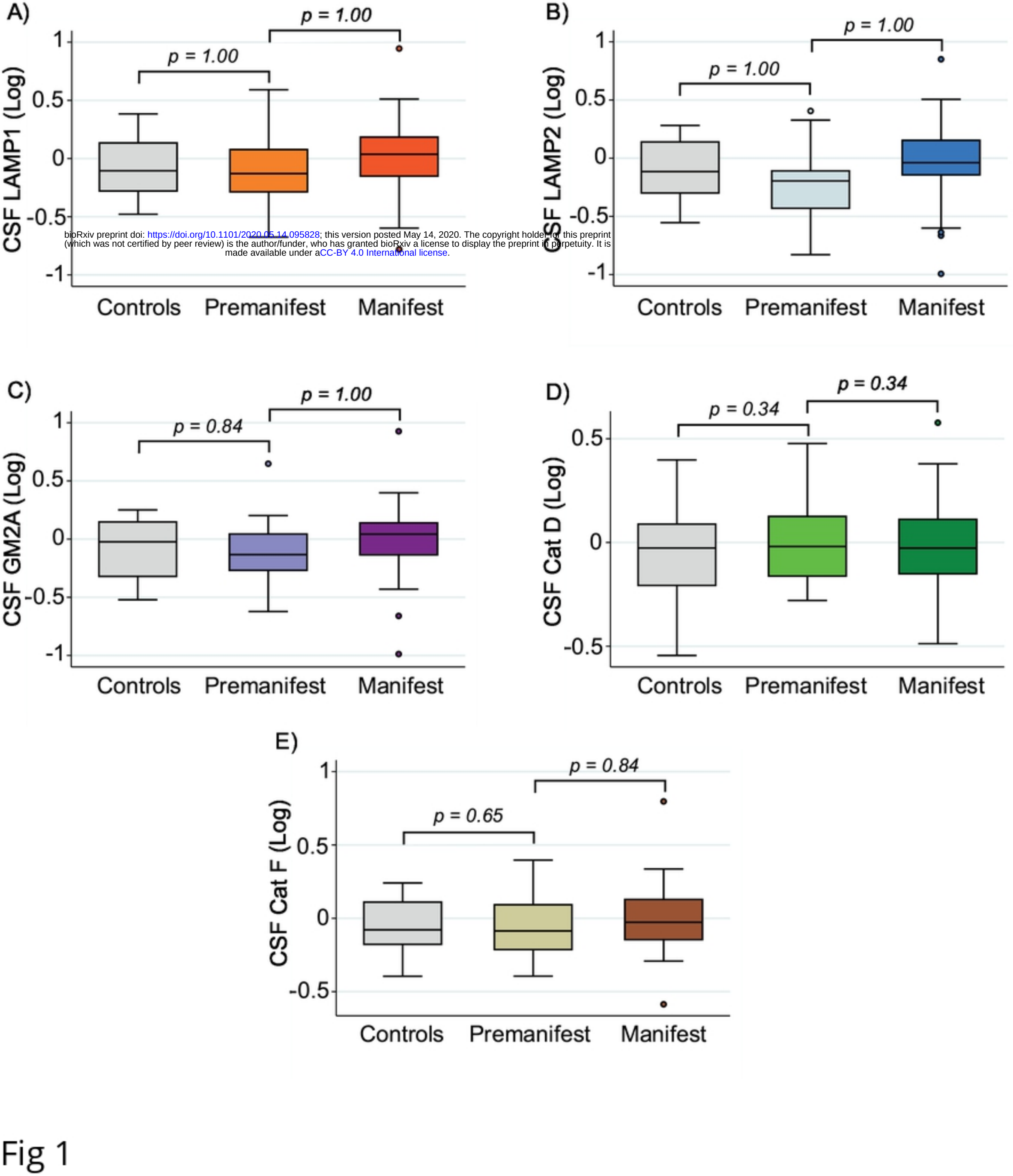


Fig 1

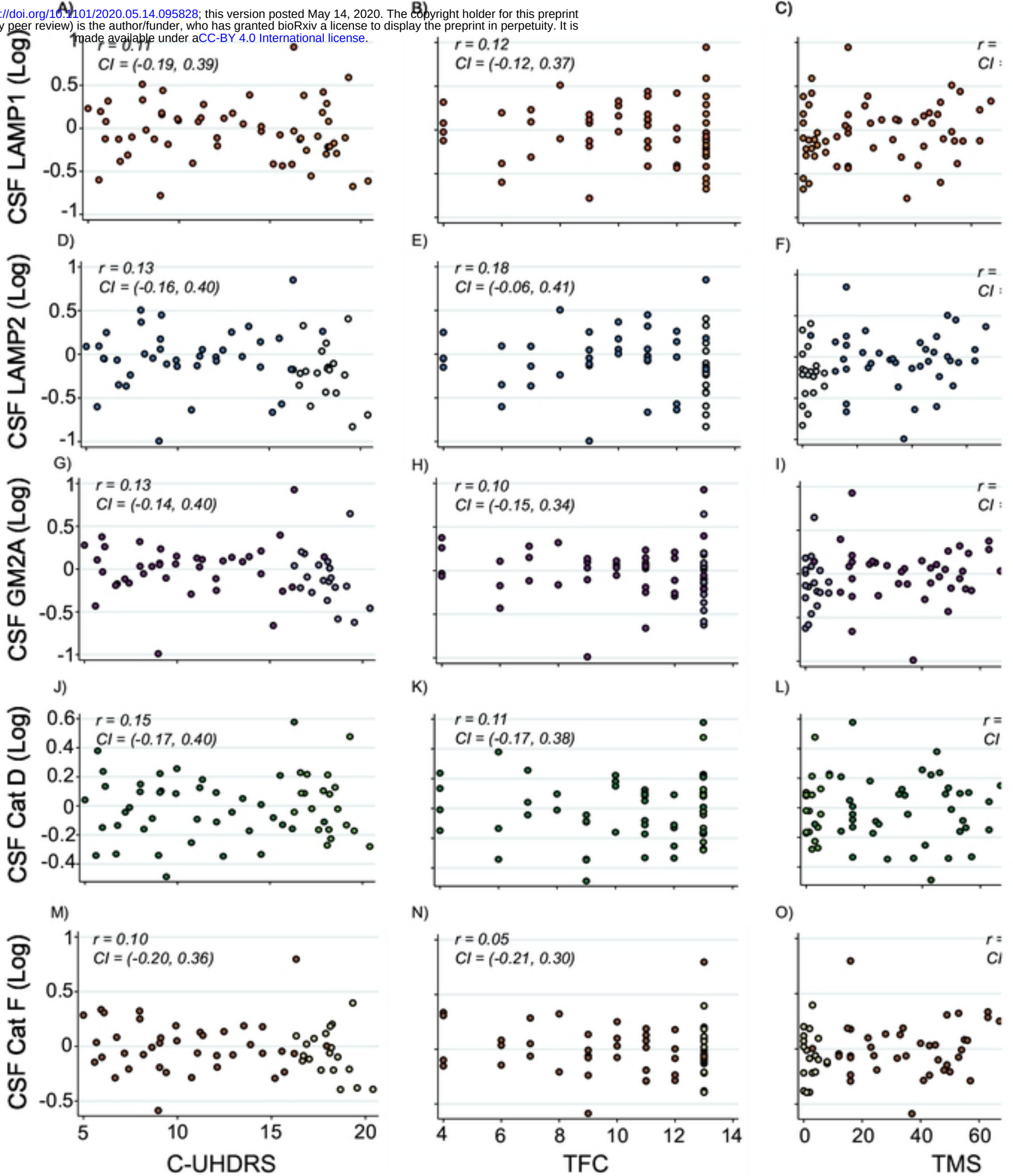
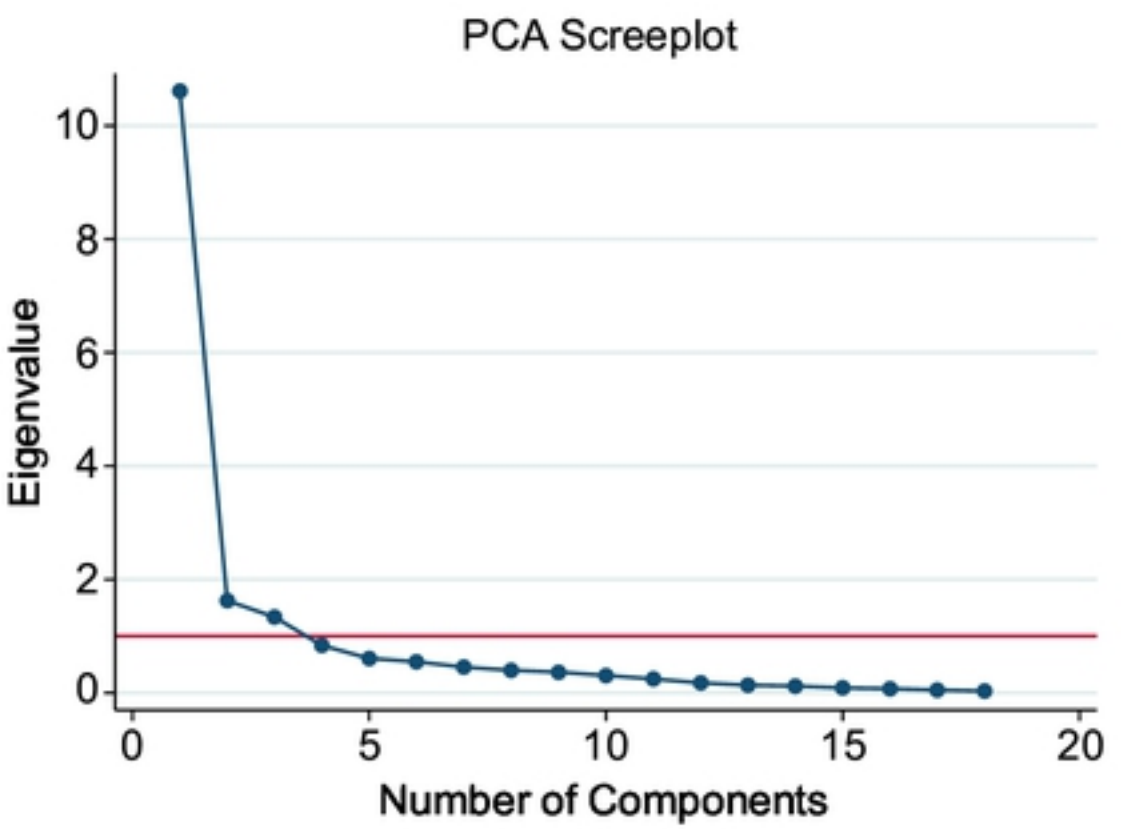


Fig 2

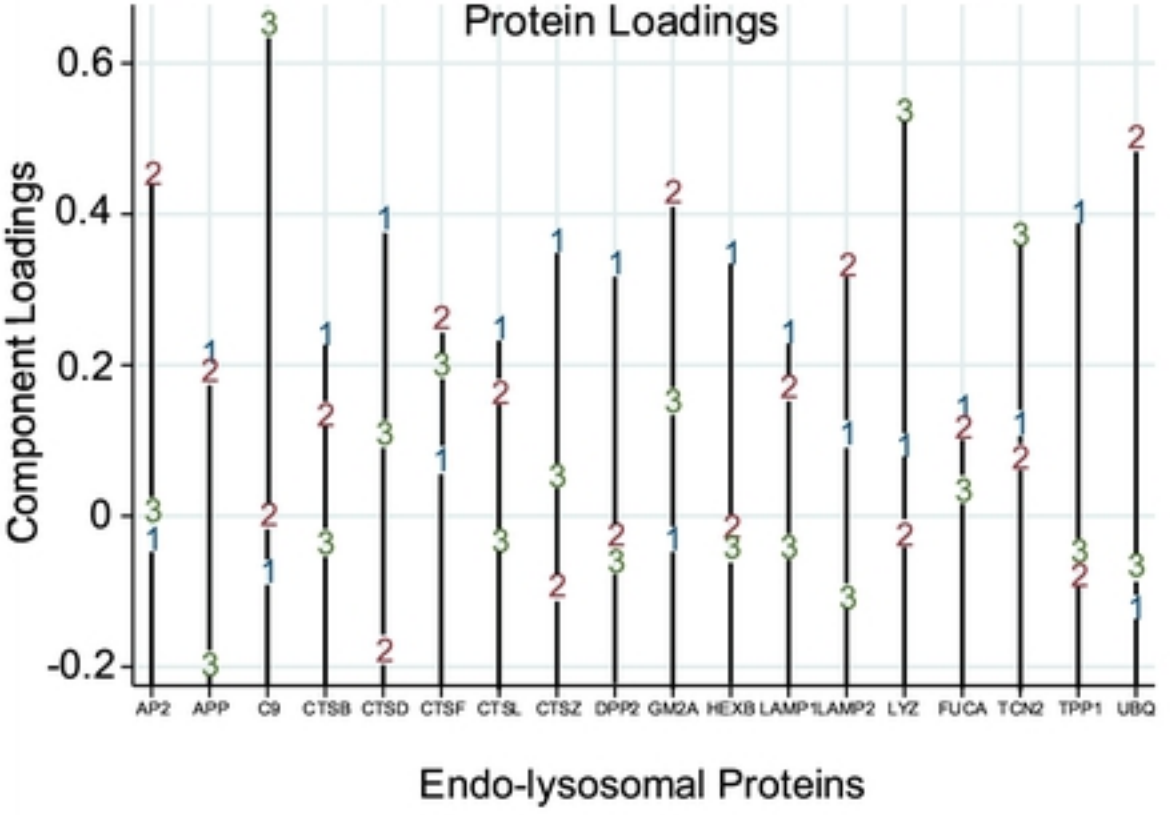
A)



B)

Endo-lysosomal Proteins	PC1 Loadings	PC2 Loadings	PC3 Loadings
AP2		0.455	
APP			
C9			0.653
CTSB			
CTSD	0.395		
CTSL			
CTSZ	0.364		
DPP2	0.336		
GM2A		0.429	
HEXB	0.350		
LAMP1			
LAMP2		0.334	
LYZ			0.538
FUCA			
TCN2			0.372
TPP1	0.403		
UBQ		0.502	

C)



D)

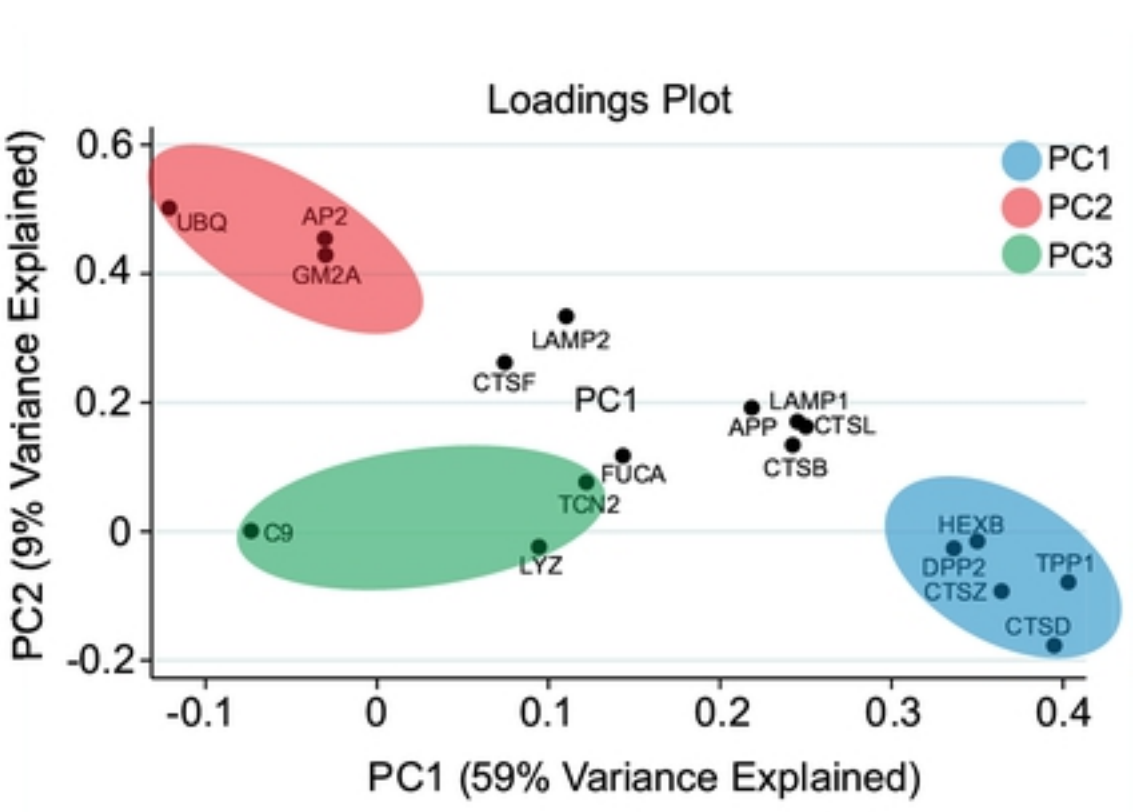


Fig 3

Fig 4

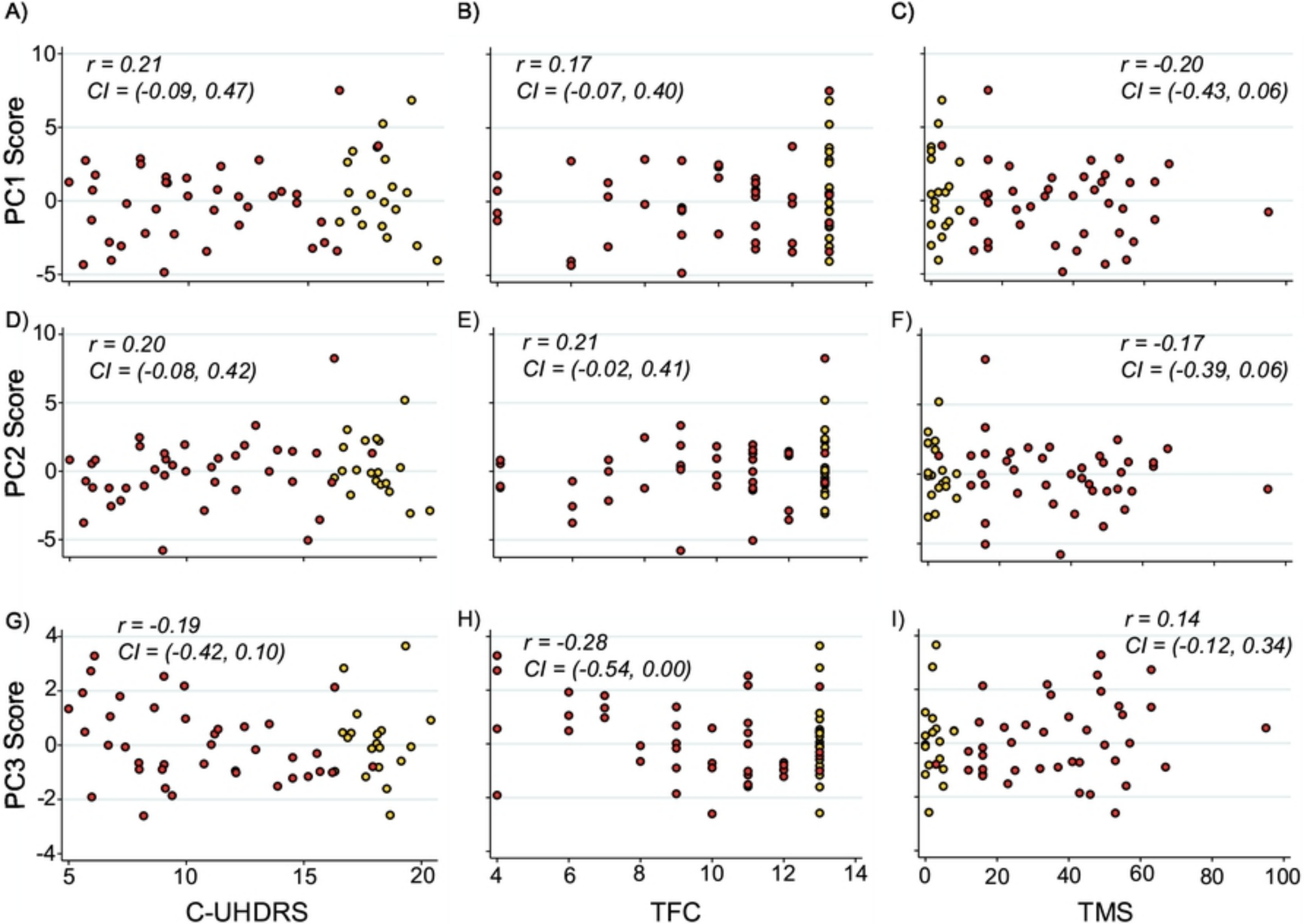

A)
B)
C)
D)
E)
F)
G)
H)
I)
PC1 Score
PC2 Score
PC3 Score
r = 0.21
CI = (-0.09, 0.47)
r = 0.17
CI = (-0.07, 0.40)
r = -0.20
CI = (-0.43, 0.06)
r = 0.20
CI = (-0.08, 0.42)
r = 0.21
CI = (-0.02, 0.41)
r = -0.17
CI = (-0.39, 0.06)
r = -0.19
CI = (-0.42, 0.10)
r = -0.28
CI = (-0.54, 0.00)
r = 0.14
CI = (-0.12, 0.34)
C-UHDRS
TFC
TMS

Fig 5

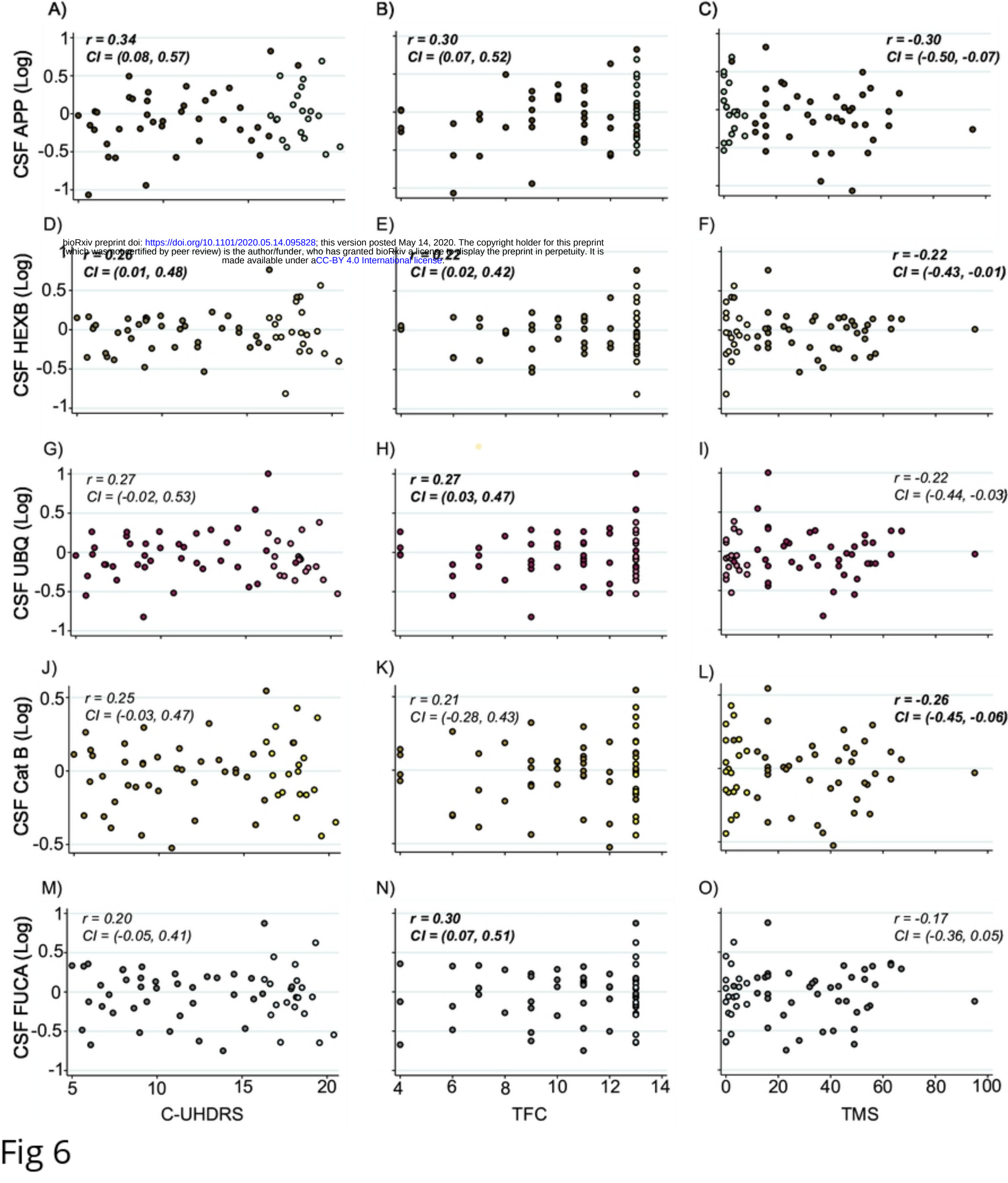


Fig 6