HHS Public Access

Author manuscript

J Proteome Res. Author manuscript; available in PMC 2024 July 07.

Published in final edited form as:

J Proteome Res. 2023 July 07; 22(7): 2493–2508. doi:10.1021/acs.jproteome.3c00199.

CEREBROSPINAL FLUID PROTEIN BIOMARKER DISCOVERY IN CLN3

An N. Dang Do[1,*], **David E. Sleat**[2,3], **Kiersten Campbell**[4], **Nicholas L. Johnson**[4], **Haiyan Zheng**[2], **Christopher A. Wassif**[5], **Ryan K. Dale**[4], **Forbes D. Porter**[5]

[1]Unit on Cellular Stress in Development and Diseases, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, United States

[2]Center for Advanced Biotechnology and Medicine, Rutgers Biomedical Health Sciences, Piscataway, New Jersey 08854, United States

[3]Department of Biochemistry and Molecular Biology, Robert-Wood Johnson Medical School, Rutgers Biomedical Health Sciences, Piscataway, New Jersey 08854, United States

[4]Bioinformatics and Scientific Programming Core, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, United States

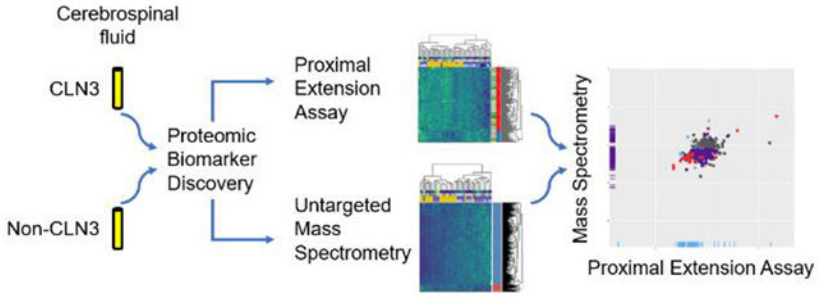
[5]Section on Molecular Dysmorphology, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, United States

Abstract

Syndromic CLN3-Batten is a fatal, pediatric, neurodegenerative disease caused by variants in *CLN3*, which encodes the endolysosomal transmembrane CLN3 protein. No approved treatment for CLN3 is currently available. The protracted and asynchronous disease presentation complicates the evaluation of potential therapies using clinical disease progression parameters. Biomarkers as surrogates to measure progression and effect of potential therapeutics are needed. We performed proteomic discovery studies using cerebrospinal fluid (CSF) samples from 28 CLN3-affected and 32 age-similar non-CLN3 individuals. Proximal extension assay (PEA) of 1467 proteins and untargeted data-dependent mass spectrometry [MS; MassIVE FTP server (ftp://MSV000090147@massive.ucsd.edu)] were used to generate orthogonal lists of protein marker candidates. At an adjusted p-value of <0.1 and threshold CLN3/non-CLN3 fold-change ratio of 1.5, PEA identified 54 and MS identified 233 candidate biomarkers. Some of these (NEFL, CHIT1) have been previously linked with other neurologic conditions. Others (CLPS, FAM217B, QRICH2, KRT16, ZNF333) appear to be novel. Both methods identified 25 candidate biomarkers, including CHIT1, NELL1 and ISLR2 which had absolute fold-change ratios >2. NELL1 and ISLR2 regulate axonal development in neurons and are intriguing new candidates for further investigation in CLN3. In addition to identifying candidate proteins for CLN3 research, this study provides a comparison of two large-scale proteomic discovery methods in CSF.

[*]Corresponding Author: An N. Dang Do, 10 Center Drive, MSC 1103, Bethesda, MD 20892, an.dangdo@nih.gov.

Graphical Abstract



Keywords

Batten; JNCL; proximal extension assay; mass spectrometry; NEFL; CHIT1; CD302; NELL1; ISLR2; protein biomarkers

INTRODUCTION

Since the first descriptions of its clinical presentation in 1894 and identification of the underlying genetic cause in 1994, CLN3 (OMIM #204200, Juvenile Neuronal Ceroid Lipofuscinosis) remains a challenging disease for basic and translational research. CLN3 is a pan-ethnic, recessive, multisystemic, neurodegenerative disorder with an estimated prevalence of 1:100,000 to 1:1,000,000 worldwide.[1] The classic syndromic presentation typically involves vision loss starting around pre-school years that progresses to blindness within the next few years.[2] Neurocognitive involvements begin with development plateauing around 7-9 years of age, followed by progressive short-term memory and skill loss. Seizure onset in the classic CLN3 presentation typically occurs during the pre-teen years. Motor involvements can start with non-functionally limiting symptoms (e.g., in-toeing, crouched gait) in the mid-teens and progress to wheelchair dependency by the late teens to early 20's. There is also a non-syndromic presentation of CLN3 disease that involves later onset and slower progression of vision loss.[3]

CLN3 codes for a 468-amino acid predicted transmembrane protein that is implicated in many cellular pathways but awaiting a clearly defined function.[4] Recent studies identified elevated levels of glycerophosphodiester species in CLN3 mouse and pig models and CLN3-affected individuals[5,6], with the work from Laqtom *et al.*[6] suggesting an essential role of CLN3 protein in the clearance of these compounds from the lysosome. The most common disease-associated variant, a 966-base pair deletion present in ~80% of chromosomes in affected individuals, is predicted to remove exons 7 and 8 and trigger either nonsense mediated mRNA decay[7,8] or production of a truncated product with novel function[9–11].

The protracted and sequential presentation of CLN3 signs and symptoms over years to decades complicates both early diagnosis and the development of clinical outcome measures applicable for therapeutic trials, which are typically designed to be conducted within 1-2 years. Biomarkers may provide useful diagnostic tools and quantifiable surrogate outcome measures of disease state for insights into treatment efficacy. Biomarker discovery

approaches require pathologically relevant, clinically well-characterized samples and high sensitivity, low sample consumption assays. Cerebrospinal fluid (CSF) can provide an important and practical research surrogate for the central nervous system pathology.[12]

In this study, we analyzed CLN3 CSF samples for potential biomarkers using conventional data-dependent mass spectrometry or multiplexed two-antibody recognition of a panel of protein targets coupled with nucleic acid amplification methods (e.g., proximal extension assay, Olink®).[13] We expected that results from the two orthogonal methods will contain both distinct and overlapping candidates, either of which could contribute to the generation of new hypotheses and targets for better understanding of CLN3 pathology.

MATERIALS and METHODS

Ethics Statement

We used samples collected under natural history protocols approved by the National Institutes of Health Institutional Review Board. Prior to study enrollment, we reviewed and obtained written consent for individuals 18 years of age or older. We obtained written permission from parents or legal guardians and assent as appropriate from those younger.

Cerebrospinal Fluid Samples (CSF)

Collection of CSF from healthy pediatric individuals is not ethically permitted. To compare CLN3 (NCT03307304) with age similar samples, we used CSF from 1) anonymized unused residuals from pediatric patients evaluated in an emergency room [pediatric laboratory controls, PLC (NCT00344331)], 2) pediatric participants from natural history protocols conducted by our group who have a non-neurodegenerative condition [Smith-Lemli-Opitz syndrome (SLOS, NCT00001721) or Creatine Transporter Deficiency (CTD, NCT02931682)]. The CLN3, SLOS and CTD CSF samples were collected from the L2-L5 lumbar region and processed in an identical manner. Anonymized PLC samples from a pediatric emergency care facility would have been collected using a standard clinical procedure via lumbar puncture, though we do not have access to the clinical records for verification. PLC, SLOS, and CTD samples were simultaneously used as non-NPC comparison samples in study of Niemann-Pick disease, type C1 biomarker discovery[14]. Additional healthy adult lumbar CSF samples were obtained from PrecisionMed (https://www.precisionmed.com).

To enable data comparisons across publications using biosamples from the same CLN3 participants, we use a consistent identifier (SP_._._) for the participants (Supplemental Table 1). We collected CSF under sedation (except SP29.2.2), following an overnight fast. Samples were aliquoted and frozen at −80 C within 1 hour of collection. Blood-contaminated samples identified by gross visual inspection were centrifuged at 1600 xg for 10 minutes prior to freezing.

Proximal Extension Assay (PEA; Olink® 1536)

We prepared 96-well plates containing CSF samples as previously described[15] and shipped these on dry ice to Olink® (Boston, MA) for Olink® Explore 1536 panel

(Oncology, Cardiometabolic, Neurology, and Inflammation) assay[16]. For analyses, we used the normalized protein expression (NPX), a log2 transformed measure of expression (Supplemental Table 2). A change of one NPX, in Olink® output data, equates to a two-fold change of protein expression level as compared to internal assay controls.

Mass Spectrometry (MS)

The protein content of CSF samples was determined by Bradford assay[17] and 20 μg equivalent of each sample were digested in-gel with trypsin (1:50 ratio, Promega #V5111, specificity, carboxyl side of K and R) and peptides extracted as described previously[18,19]. A total of 45 samples were analyzed in three TMT16-plex studies (Supplemental Table 1). To facilitate comparison between the three TMT16-plex studies, a bridging reference standard was included in each study comprising a pool of all 45 samples analyzed. An equal number of CLN3 and non-CLN3 samples were assigned to each TMT16-plex study to eliminate any potential study-related bias. After digestion, samples were labeled with TMT16-plex reagents using manufacturer's (ThermoFisher Scientific) protocols, pooled and desalted prior to high pH reverse-phase HPLC fractionation[20]. Twenty-two high pH reverse-phase-HPLC fractions from each TMT16-plex study were analyzed by LC-MS/MS.

Chromatography was conducted as described[20]. TMT reporter ion intensities from the TMT16-plex labeled CSF peptides were measured using SPS-MS3[21] on a Thermo Orbitrap Eclipse Tribrid mass spectrometer. Each 3s cycle started with a full MS scan acquired in the Orbitrap (375–1600 m/z, 120 K resolution, RF Lens 60%). For MS2, parent ions were isolated in the quadrupole (2 m/z isolation width), fragmented, and then scanned in the ion trap (normalized collision energy of 35%, 400–1600 m/z). For MS3, parent ions were isolated in the quadrupole, and 10 MS2 fragment ions isolated in the ion trap were fragmented by HCD (collision energy of 55%) and scanned in the Orbitrap detector (2 m/z isolation width, 100–500 m/z, 50K resolution).

Peak lists were generated using Proteome Discoverer version 2.2 with no constraints with respect to retention time, charge state or peak count, and a minimum signal/noise of 1.5 and minimum and maximum precursor masses of 350 Da and 10,000 Da. Data were searched using a local implementation of the Global Proteome Machine version 3[22,23] (GPM Fury with X!Tandem version ALANINE 2017.02.01) (Beavis Informatics Ltd., Winnipeg, Canada) against a human proteome based on ENSEMBL GRCh37.70 (Aug 2012; 56 680 total genes, 20 447 protein coding genes) and a custom database containing laboratory contaminants and other non-murine proteins. TMT MS3 reporter ion intensities were extracted with correction for isotope channel crossover using custom in-house software and exported to Excel, as previously described[24] (Supplemental Tables 3A–C).

Mass spectrometry data normalization and filtering

TMT reporter intensities for each spectra were first divided by the sum of the TMT reporter ion intensity for each channel to correct for differences in labeling efficiency or amount of protein labeled. Intensities for each channel thus corrected were then normalized to the bridging reference channel (134N) to allow cross comparison between the three TMT studies. Protein assignments were filtered to those with two or more unique peptides

(Supplemental Table 3D), and peptides were filtered for those with a total TMT intensity of >10000, a reference channel intensity of >1000, fully tryptic digestion with no missed cleavage sites, and an absence of posttranslational modifications deemed to increase variability in the data (asparagine or glutamine deamidation, methionine dioxidation, tryptophan mono- and dioxidation, and isobaric labeling of tyrosine at positions other than the N terminus) (Supplemental Table 3E).

Raw files, mgf files, GPM result files, Excel workbooks listing protein and peptide assignments, and keys for data interpretation are archived in the MassIVE (http://massive.ucsd.edu) and ProteomeXchange (http://www.proteomexchange.org/) repositories with accession number MSV000090147.

Data Analysis

Computational analyses were conducted using R version 4.0. PCA, Kruskal-Wallis and ANOVA modeling, and multiple testing correction were conducted with R base package stats.

Differential Abundance Analysis—The PEA outputs protein abundance measurements for proteins in targeted panels as a log2 scale NPX value. Prior to differential abundance analysis, protein abundance measurements, originally provided by Olink as Excel spreadsheets with information encoded as cell color, were combined into a cleaned protein-by-sample matrix in tab-delimited format ready for programmatic processing in R. In instances where a protein was measured in more than one panel, a suffix was added to the protein name, allowing both protein measurements to be present in the combined matrix.

A cell in the NPX matrix with an “NA” value or QC warning indicates that there was an error in measuring the abundance of a given protein (row) in a given sample (column). Any protein for which the NPX values were noted with one of those two errors across 75% or more of all samples were removed from the analysis. Five proteins were removed and 1467 retained for downstream analysis. This filtering process was also conducted across samples, identifying cases where 75% or more of the protein measurements for a given sample were noted with one of the above two errors. All samples passed this threshold. Sample PCAs were constructed to visually identify outliers, leading to the exclusion of two PLC samples (Supplemental Table 2).

Differential protein abundance was then analyzed for various sample group comparisons. Fold change ratios between two conditions were calculated as the difference in median log2 protein abundance between the two conditions. A Kruskal-Wallis rank-sum test was conducted for each protein to assess the difference in protein abundance between samples in different conditions. P-values generated for each protein from the Kruskal-Wallis test were then adjusted to account for multiple testing using the Benjamini-Hochberg correction. Proteins with an adjusted p-value < 0.1 were identified as differentially abundant between the two conditions of the contrast.

The differential abundance analysis on the MS dataset mirrors the analysis conducted on the PEA dataset, with the following modifications. No protein or sample outliers were identified

in the original matrix. The MS dataset included both NA, indicating measurement error, and true zero values in the input matrix. To properly incorporate the true zero values into the log fold change calculations, the medians of each group in a comparison were calculated, followed by a log2 transformation and a difference of median calculation. If the median protein abundance in either group is 0, we report NA as the log fold change (Supplemental Table 3F).

Metadata Covariate Analysis—Additional analyses were conducted to ensure that the results of the differential abundance analyses were not influenced by covariates that would confound evaluation of the effect of genotype alone. Age and Sex were assessed as potential covariates in the differential abundance analysis. To do so, the distribution of patient age was compared between sample groups in each contrast of interest, and a Wilcox rank-sum test was used to determine if the patient age distributions were significantly different between sample groups, indicating a potential covariate. Similarly, patient sex distributions were compared between sample groups for each contrast of interest, and a Chi-square test was used to assess statistical significance of distributional differences. This was conducted for both the PEA and MS datasets. Age demonstrated distributional differences between CLN3 and non-CLN3 samples in only the MS dataset. However, given that the two dataset results are compared to identify high-confidence biomarkers, introduction of Age as a covariate should be included in the PEA model as well as the MS model. Sex demonstrated distributional differences in only the Olink dataset and only when comparing CLN3 samples to a subset of non-CLN3 samples (CTD+SLOS).

We re-analyzed the PEA and MS data using an ANOVA model, rather than Kruskal-Wallis, which allows for covariates to be introduced. An ANOVA model was constructed for each protein using Age as a covariate, and Benjamini-Hochberg was used for multiple testing correction. Log2 fold change calculations remained the same.

Comparison of PEA and MS Results—We compared the results from the two discovery assays on samples analyzed by both methods. Using this sample subset, the Kruskal-Wallis testing and log2 fold change calculations were executed as above for each assay independently, but only using the samples in common. Not all proteins were quantified in both the PEA and MS assays. However, for those that are quantified in both assays, we can identify high confidence biomarkers as those proteins for which the adjusted p-value < 0.1 when using the PEA *and* the MS protein abundance measurements.

Enrichment Analyses—To identify proteins with similar changes in abundance in the remaining ~1500 proteins assayed, we first created a matrix of protein-by-protein correlation (i.e., Spearman correlation coefficient of pairwise correlation between all proteins in the assay)[25]. We then conducted hierarchical clustering on the matrix to generate a dendrogram. Euclidian distance was used for a complete linkage hierarchical clustering. We used a height cut point of 10 for the PEA and 12 for the MS dendrogram to produce larger sets of more similar clusters. The entire correlation matrix was then plotted as a heatmap, with clusters annotated according to a given cut point.

The 1467 proteins assayed by the PEA method represent a selection of the proteome/genome, and the limited number of differentially expressed proteins make functional enrichment analyses using overrepresentation tests challenging. To assess coordinated, but potentially subtle, functional enrichment, we used the gene set enrichment analysis (GSEA) approach[26]. This evaluates which proteins in the ranked list are annotated with each GO term, and whether these proteins tend to be clustered toward the beginning or end of the list. For the CLN3 versus non-CLN3 rank lists, we used log2 fold change for ranking PEA and MS protein expression levels. The ranked list used as input to GSEA contains the log2 fold change of all proteins, regardless of directionality.

RESULTS

Sample characteristics and sensitivity analyses

Proximal Extension Assay—CSF samples used in the PEA assay and the demographics of study participant donors are shown in Table 1. Comparing CLN3 (n=28) to non-CLN3 (n=32) samples, age median and range values are similar, whereas sex in the latter group skewed males given the inclusion of CTD, an X-linked condition. CLN3 genotype and phenotype (syndromic or vision-only) are provided in Supplemental Table 1. Fourteen (50%) of the CLN3 CSF samples are from individuals homozygous for the common 966-bp (c.461-280_677+382) deletion variant. The remaining 14 samples are from individuals with other disease associated CLN3 genotypes. Two CLN3 individuals (SP5.2.2 and SP29.2.2) have the vision-only phenotype. Age-similar pediatric CSF comparison samples were obtained from study participants with non-neurodegenerative conditions [CTD, SLOS, and emergency room visit residuals (PLC)]. Comparisons of CTD versus SLOS, CTD versus PLC, or SLOS versus PLC showed no protein with expression differences below the 0.1 adjusted p-value. Thus, these non-CLN3 pediatric samples were combined as the comparison group.

In visualizing proteins differentially expressed in CLN3 versus non-CLN3 samples with adjusted p-value < 0.1, 27/28 CLN3 samples clustered separately from the non-CLN3 samples. The remaining CLN3 sample from the youngest participant (SP19.2.1) clustered with a sub-group of the non-CLN3 samples (Figure 1A). An SLOS and a PLC sample cluster with the CLN3 group. While the PLC sample was from one of the older participants in the PLC cohort, the SLOS sample was not. This SLOS sample and those forming the sub-group of non-CLN3 samples have no apparent distinction in demographics or disease presentation in comparison to the other participants within their respective cohort. Removing the two vision-only CLN3 samples effected a more homogenous clustering of CLN3 and non-CLN3 samples, with the exception of the 2nd youngest (SP21.2.1) CLN3 samples now cohorting with the non-CLN3 group (Figure 1B).

Since PLC samples were obtained from residual clinical samples and not processed using the same procedures as those in our studies, we also conducted analyses using samples collected in our studies only. Comparisons of CLN3 versus non-PLC (CTD+SLOS) samples yielded less proteins differentially expressed at adjusted p-value < 0.1. In this heatmap clustering, all CLN3 samples, except for that from the youngest participant (SP19.2.1),

grouped together (Figure 1C). Exclusion of the two vision-only samples showed similar clustering pattern (Figure 1D).

Mass spectrometry—Characteristics of CSF samples used for quantitative isobaric-label mass spectrometry and the demographics of study participant donors are shown in Table 1 and Supplemental Table 1. The CLN3 CSF samples (n=20) were all from individuals with syndromic presentation. Age median and range values are higher in the non-CLN3 group (n=25) given inclusion of commercial healthy adult samples. Similar to PEA, the non-CLN3 cohort skewed males given inclusion of CTD samples. In visualizing proteins differentially expressed in CLN3 versus non-CLN3 (CTD+SLOS+Adult) samples with adjusted p-value < 0.1, 20/25 non-CLN3 samples clustered separately from the CLN3 samples (Figure 2A). The five non-CLN3 samples that clustered with the CLN3 group are from individuals with the highest ages. Restricting the comparison to CLN3 versus non-CLN3 below 24 years of age grouped 17/18 non-CLN3 together, with the sample with the highest age (23.96 years) remaining grouped with the CLN3 samples. The two CLN3 samples that grouped with the non-CLN3 cohort belong to participants aged 7.3 and 7.8 years (Figure 2B).

Sensitivity Analyses—To evaluate further the clustering patterns described above, we compared potential analysis differences of 1) excluding the vision-only samples from the CLN3 cohort and 2) including age as a covariate versus including all CLN3 samples in analyses without age as a covariate. In the PEA data sets, the lists of differentially abundant proteins for the all CLN3 vs. all non-CLN3 and syndromic CLN3 vs. all non-CLN3 contrasts are similar (Supplemental Table 4). The lists of differentially expressed proteins for the all CLN3 vs. all non-CLN3 and all CLN3 vs. non-PLC contrasts displayed greater differences that likely reflect the decreased statistical power given smaller cohort size (Supplemental Table 4). In the MS data sets, the lists of differentially abundant proteins for the all CLN3 vs. all non-CLN3 and all CLN3 vs. all non-CLN3 under 24 years of age contrasts identified similar differentially abundant proteins with the latter contrast also identifying a larger unique set of proteins (Supplemental Table 5).

Given the above clustering of higher age non-CLN3 samples with the CLN3 group, we evaluated the effect of age on protein expression. Re-analyses of the PEA and MS data using an ANOVA model and Age as a covariate, showed that across all contrasts listed in the previous paragraph the two statistical modeling approaches identified > 50% of differentially abundant proteins in common. The ANOVA with covariates analyses consistently identified a larger set of differentially abundant proteins unique from the Kruskal-Wallis analyses (Supplemental Tables 6, 7).

Based on the above sensitivity analyses, and to provide a clear and conservative discussion of candidate biomarkers, we will focus the subsequent data presentation and discussion on the following analyses: 1) all PEA CLN3 vs. all PEA non-CLN3 samples, 2) all MS CLN3 vs. all MS non-CLN3 samples, 3) Kruskal-Wallis statistical model due to its robustness for non-normality.

Differential expression

Proximal Extension Assay—At an adjusted p-value < 0.1, 146 proteins were differentially expressed in all CLN3 compared to non-CLN3 samples, as depicted in a volcano plot (Figure 3A). Of the 146 proteins 54 were 1.5-fold or more (i.e., absolute log2 fold change > 0.585) differentially expressed between the two cohorts, with 49 (33%) higher and 5 (3%) lower in CLN3 (Supplemental Table 8). The top proteins with differential expression at p < 0.1 and absolute log2 fold change ≥ 1 in CLN3 compared to non-CLN3 (Table 2A, Figure 3A) include several previously identified in models of other neurodegenerative and neurometabolic conditions or processes: NEFL (neurofilament light chain[27] and previously identified in CLN3[15]), CHIT1 (chitinase 1) and CHI3L1 (chitinase 3-like 1[28]), CTSH (cathepsin H[29]), CPVL (carboxypeptidase vitellogenic like[14,30]), LEP (leptin[31]), EFEMP1 (EGF containing fibulin extracellular matrix protein 1[32]), CD302 (C-type lectin domain family 13, member A[33,34]). A protein with ≥ 2-fold differential expression, and per our Pubmed search [(complete protein name OR symbol) AND (nerv* OR neuro*)] not previously associated with other neurological conditions, was CLPS (colipase). Though CLPS is more commonly associated with being produced by and regulating the function of the pancreas, CLPS mRNA has been detected in rat hypothalamus[35].

Mass Spectrometry—At an adjusted p-value < 0.1, 662 proteins had differential expression level in CLN3 compared to non-CLN3 samples, as depicted in a volcano plot (Figure 3B). 234 of the 662 proteins were 1.5-fold or more differentially expressed between the two cohorts, with 26 (4%) higher and 208 (31%) lower in CLN3 (Supplemental Table 9). Of the markers with ≥ 2-fold differential expression (Table 2B, Figure 3B), CHIT1, CA8 (carbonic anhydrase 8[36]), BCS1L [BC1 (ubiquinol-cytochrome C reductase) synthesis-like[37]], COL23A1(collagen type XXIII alpha 1 chain[38]), NT5DC2 (5'-nucleotidase domain containing 2[39]), SCG5 (secretogranin V[40]), ASXL3 (additional sex combs like transcriptional regulator 3[41]), C9orf84 (shortage in chiasmata 1[42]), GPR161 (G protein-coupled receptor 161[43]), MSI2 (musashi RNA binding protein 2[44]), ATP6AP2 (ATPase H+ transporting accessory protein 2[45]), FHL1 (four-and-a-half Lin11, Isl-1 and Mec-3 domains 1[46]), HYDIN (hydrocephalus-inducing axonemal central pair apparatus protein[47]), CHGA (chromogranin A[48]), UNC13C (unc-13 homolog C[49]), CA3 (carbonic anhydrase 3[50,51]), CKM (creatine kinase, M-type[52]), CHST10 (carbohydrate sulfotransferase 10[53]), SHISA6 (Shisa family member 6[54]), PCSK1 (proprotein convertase subtilisin/kexin type 1[55]), MGAT3 (beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase[56]), WFDC1 (WAP four-disulfide core domain 1[57,58]), have been associated with neurologic conditions or processes in various models. Biomarkers with ≥ 2-fold differential expression, and per our Pubmed search not previously associated with other neurological conditions or processes include: FAM217B (family with sequence similarity 217 member B), QRICH2 (glutamine rich 2), KRT16 (keratin 16), ZNF333 (zinc finger protein 333). FAM217B has been reported to have repressed expression in ulcerative colitis[59] and differentially expressed in glucokinase-positive neurons[60]. QRICH2 regulates formation of sperm flagella and when absent is associated with male infertility[61]. Variants in its homolog gene, *QRICH1*, is linked to a neurodevelopmental short stature syndrome[62]. QRICH1 is not present or detected in either the PEA or mass spectrometry analyses. KRT16 is a member of the keratin family

involved in inflammatory responses, whose abundance level has been reported to increase in a mouse model of brain ischemia[63] or complex regional pain syndrome[64], and blood sample from individuals with fascioscapulohumeral dystrophy[65]. Variants in the gene encoding ZNF333 have been associated in certain analytical models of Alzheimer samples[66] and with markers of apoptosis[67]. (CD302 and NEFL were not detected in the MS analysis of CSF.)

Assay Comparisons

To identify the most likely biomarker candidates, we compare changes identified by PEA and MS for proteins that were found by both methods to be significantly altered. We focused the analyses on the subsets of CLN3 (n=20) and non-CLN3 (n=10) CSF samples that were used in both assays (Table 1 and Supplemental Table 1). PEA measured 1002 proteins for which the mass spectrometry assay did not consistently identify (i.e., NA values) (Table 3A). Fifty-nine (6%) of these have adjusted p-value < 0.1, with the proteins that have the largest magnitude log2 fold change being NEFL (2.5) and CD302 (−3.4) (Figure 4, Table 3A, Supplemental Table 10). In comparison, mass spectrometry measured 2833 proteins for which PEA did not consistently identify. Of these, 478 (17%) have adjusted p-value < 0.1, with the proteins that have the largest magnitude log2 fold change being CCPG1 (0.96) and ZNF333 (−2.3).

Overall, 460 proteins were measured by both assay methods (Table 3B, Supplemental Table 10). Of these, 310 (67%) proteins showed concordant changes (i.e., log2 fold change in the same direction in both assays). Twenty-five proteins have adjusted p-values < 0.1 by both PEA and MS assay methods and concordant log2 fold change (Table 4). The majority (23/25) showed concordant decreases. Sixty percent of these are proteins that are involved in neuronal or ocular processes. Three markers are measured by and have concordant changes as well as have adjusted p-value < 0.1 and log2 fold change > 1 in CLN3 in both assays. CHIT1 is a marker elevated in Gaucher disease and other neurologic conditions[28]. NELL1 (neural epidermal growth factor-like 1), and its homolog NELL2 (PEA p-adjusted = 0.12, log2 fold change = 0.21; MS p-adjusted = 0.01, log2 fold change = −0.71), are involved in directing axon outgrowth in neurons and retinal ganglions.[68,69] ISLR2 (immunoglobulin superfamily containing leucine rich repeat 2, LINX) similarly has been associated with Alzheimer disease[70] and retinal axon guidance[71]. Expression level for NELL1 and ISLR2 have high correlations with each other in both the PEA (rho=0.86) and MS (rho=0.57) hierarchical clustering.

Protein expression patterns

To identify patterns for the differential expression outcomes, we evaluated protein correlations and gene set expression in the CLN3 versus non-CLN3 contrasts. To generate correlation values we used the following Kruskal-Wallis contrasts: 1) all PEA CLN3 vs. all PEA non-CLN3 samples, 2) all MS CLN3 vs. all MS non-CLN3 samples, 3) the subsets of CLN3 (n=20) and non-CLN3 (n=10). We calculated the correlations of differential abundance for each pair of proteins assayed and performed hierarchical clustering to find potential additional proteins with differential abundance not identified by the Kruskal-Wallis analyses[25,72]. We hypothesized that this may provide insights into which proteins are similarly affected by the absence of CLN3, and perhaps potential cellular response patterns.

We also evaluated functional enrichment using GSEA to incorporate the protein expression level data and account for PEA using only a subset of the proteome.

Proximal Extension Assay—Figure 5A shows the protein by protein correlation matrix using PEA data. Each cluster includes a group of proteins with differential abundances in CLN3 versus non-CLN3 that have a similar pattern of pairwise correlation coefficients with other proteins in the group (Supplemental Table 11). NEFL and LEP, as well as CHIT1, CTSH, CPVL, and CHI3L1 have $r > 0.5$ with other proteins that have significant adjusted p-value and log2 fold change, and also with those that do not (e.g., BAG3, CDH3) (Table 2A). CLPS and EFEMP1 each has a relatively large list of correlative proteins involved in various neuronal cell development processes; some of these proteins did not meet the significance criteria in the differential abundance analyses (Table 2A, Supplemental Table 4). The overall size of a cluster does not necessarily dictate the number of correlative proteins (e.g., compare proteins in cluster 8 versus 16). GSEA suggested an effect of CLN3 deficiency on biological processes that include immune response, fusion, and transport (Figure 6A). Of the proteins assessed by PEA, CLN3 deficiency did not appear to affect enrichment for particular cellular components or molecular functions (Supplemental Figures 1A and 2A).

Mass Spectrometry—Figure 5B shows the protein by protein correlation matrix using mass spectrometry data. Multiple proteins with significant differential abundance did not group with any cluster (Table 2B, Supplemental Table 12). SCG5, ATP6AP2, CHST10, and SHISA6 have large correlative clusters containing proteins classified to have binding or catalytic activity function. Some of these proteins did not meet the significance criteria in the differential abundance analyses (Supplemental Table 5). GSEA suggested an effect of CLN3 deficiency on biological processes involving neuronal/axonal development and cellular metabolism (Figure 6B). Of the proteins assessed by mass spectrometry, CLN3 deficiency also affected enrichment for membranous and macromolecule cellular components, and receptor binding molecular functions (Supplemental Figures 1B and 2B).

DISCUSSION

Leveraging technical and methodological advancements in proteomics, we conducted discovery studies for candidate biomarkers for CLN3 disease using unique collections of CSF from pediatric study participants in a longitudinal prospective natural history protocol. Biomarkers identified in this study that have also been associated with other neurologic conditions provide support for the clinical relevance of the sampled cohorts and applied assays. As several of these biomarkers [such as NEFL[73,74]] have been evaluated in more frequently occurring conditions that are amenable to large-scale validation efforts, application of the overlapping biomarkers in interventional trials for rare diseases such as CLN3 holds promise. The overlaps imply that disease-specificity may necessarily derive from different combinations of changes in biomarker expression[25]. In addition to the fluid biomarkers that have also been recognized in other diseases, this discovery work also identified intriguing potentially CLN3-specific proteins, for which further work will be needed to determine their contribution to CLN3 specificity.

The proximal extension assay combines antibody-based detection (thus, theoretically targeting proteins conformationally close to their native state) with signal amplification by next-generation sequencing. The PEA requirement for two antibodies capable of recognizing epitopes within limited distance from each other increases assay specificity and limits the detection of large number of proteins. Comparatively, mass spectrometry assay enables less constrained detection of digested peptides based on their mass and charge. While this allows detection of larger number of inferred proteins, a distinct difference resides with the detection of proteins in PEA versus peptides in MS. Combinatorial application and analysis of both methods for discovery of biomarkers for CLN3 provides an approach to mitigate each method's limitations and identify a set of high-confidence biomarkers. Petrera *et al.*[75] compared PEA and LC-MS/MS methods in human plasma samples and identified 52/736 (7%) overlaps, with 8-14/35 (23-40%) of the overlapping markers having moderate to high correlative findings. These lower percentages as compared to 31% (462/1467) marker overlap from our study likely relate to the smaller number of PEA markers evaluated and the more complex sample matrix in the use of plasma samples in the previous comparison. Panels with greater number of protein markers are now available from Olink®, and may prove instructive for further comparisons with mass spectrometry.

For further hypothesis generation, we looked for proteins with correlative changes in the CLN3 vs. non-CLN3 samples, and for enriched processes. Both approaches provided broad categories of other genes and processes of potential interest, some of which are similar to current information in the field (e.g., involvement of immune response biological process), and all of which are suggestive of much further investigations to pursue. Previous work towards identifying pathways affected in CLN3 used different sample type/method combinations that included human lymphocytes/microarray[76], mouse brain/mass spectrometry[77], Cln3 knock-in mouse cerebellar cells/mass spectrometry[78], and human post-mortem CSF/mass spectrometry[79]. Thus, it is not unexpected that overall, there are very low correlations of the fold change values of each of these published marker list with the PEA or MS candidate biomarker list (Supplemental Table 13). Sample genotype heterogeneity in our cohort may also affect the fold change magnitude for some markers. Recent work by Minnis *et al.*[11] demonstrated differences in regulated pathways and phenotypes in yeast cln3-equivalent mutations that are predicted to have similar cln3 protein outcome.

The orthogonal methods concordantly identified three markers of relevance and interest. CHIT1, a hydrolase of glycosyl-containing compounds such as chitin, was clearly identified by both assay methods to be elevated in CSF of individuals with CLN3. CHIT1 and the non-enzymatic related protein CHI3L1 are hypothesized to protect the central nervous system against damaging innate inflammatory or immunologic responses involving microglia and astrocytes, respectively[28]. While CHIT1 elevation has not been previously reported for CLN3, it has been linked with lysosomal disorders such as Gaucher and Niemann-Pick disease, type C1[80], as well as the infantile form of neuronal lipofuscinosis, CLN1[79]. CHIT1 level correlated with disease state in multiple sclerosis, Alzheimer, and stroke[28]. Its elevation in CLN3 hints at its potential as a marker for tracking progression or treatment responses, perhaps a role similar to that in Gaucher disease type 1[81].

In contrast, to our knowledge neither NELL1 nor ISLR2, the two markers concordantly identified by both assays as being decreased in CLN3, have been examined as fluid biomarkers for diseases. First identified in chicken, nel transcripts were expressed in all embryonic tissues examined but were highest in the brain[82]. Post-natally, nel transcript levels were maintained in the brain, increased in the retina, and significantly decreased in non-neural tissues. Based on sequence, the chicken nel is orthologous to mammalian *Nell2*. NELL2 regulates differentiation and axon development in rat hippocampal neurons[83], the mechanism for axonal growth guidance being through interactions with the Robo family of receptor proteins that direct axon formation[69]. Decreased Nell2 expression in mouse retina resulted in abnormal axonal development of the retinal ganglion cells[68]. In humans, NELL2 is strongly expressed in cortical neurons, apparent in oligodencytes, and insignificantly present in astrocytes[84]. While more prominently linked to regulation of bones and connective tissues, as a decrease in its expression affected skeletal development and led to perinatal death[85], the homolog Nell1 transcripts were shown to be expressed in similar areas of mouse central and peripheral nerve tissues as Nel and Nell1 protein to form complex with Nel[86]. Like NELL2, decreased ISLR2 in CLN3 CSF sample is intriguing and apropos to the pathology. Identified through a screen for genes involved in regulating sensory and motor axonal development, Islr2 was found to be present in mouse embryonic spinal cord and dorsal root ganglia, and its absence to result in disrupted peripheral nerve growth[87]. In a different mouse model, Islr2 was detected in the hippocampus, apical cotex, and an area corresponding to the olfactory nucleus. In mice either homozygous or heterozygous for *Islr2* knockout, severe hydrocephalus and defects in structures connecting the brain hemispheres occurred[88]. In another parallel to NELL2, failure of retinal ganglion cell axons to decussate at the optic chiasm occurred in an *islr2* mutan zebrafish model[71]. NELL1 and ISLR2 will need further study to define their utility as biomarkers in CLN3 and their role in the disease process.

In summary, using two proteomic discovery methods we have identified promising cerebrospinal fluid candidate biomarkers for CLN3, a fatal pediatric neurodegenerative disease currently without treatment. Proteins independently identified by each method include those previously linked to neurological processes or diseases, suggesting biological relevance. Proteins identified by comparison of overlaps between the methods highlight potential new targets for investigations to further understanding of the underlying pathophysiology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

We dedicate this work to the study participants, their families, and the support organizations for the motivation and inspiration they have provided. We thank Mr. Loc Trinh and Mr. Fathy Majadly (NICHD Biorepository), Ms. Kisha Jenkins (study coordinator), Mr. Andrew Smith (post-baccalaureate trainee) and colleagues and staff who enabled the conduct of this study. Funding for this project was provided through an 11th NCL-Stiftung Research Award (DS, ADD, FDP), an NIH Clinical Center Bench-to-Bedside Award (FDP), NICHD Intramural Research Program ZIA009001-01 (ADD), ZIC008986-03 (RKD), and ZIA000892 (FDP). Mass spectrometry was conducted by the

Biological Mass Spectrometry Facility of Robert Wood Johnson Medical School and Rutgers University, partly supported by S10 OD025140.

REFERENCES

1. Sleat DE, Gedvilaite E, Zhang Y, Lobel P, Xing J. Analysis of large-scale whole exome sequencing data to determine the prevalence of genetically-distinct forms of neuronal ceroid lipofuscinosis. Gene. Nov 30 2016;593(2):284–91. doi:10.1016/j.gene.2016.08.031 [PubMed: 27553520]
2. Ostergaard JR. Juvenile neuronal ceroid lipofuscinosis (Batten disease): current insights. Degener Neurol Neuromuscul Dis. 2016;6:73–83. doi:10.2147/DNND.S111967 [PubMed: 30050370]
3. Ku CA, Hull S, Arno G, et al. Detailed Clinical Phenotype and Molecular Genetic Findings in CLN3-Associated Isolated Retinal Degeneration. JAMA Ophthalmol. Jul 1 2017;135(7):749–760. doi:10.1001/jamaophthalmol.2017.1401 [PubMed: 28542676]
4. Mirza M, Vainshtein A, DiRonza A, et al. The CLN3 gene and protein: What we know. Mol Genet Genomic Med. Dec 2019;7(12):e859. doi:10.1002/mgg3.859 [PubMed: 31568712]
5. Brudvig JJ, Swier VJ, Johnson TB, et al. Glycerophosphoinositol is Elevated in Blood Samples From CLN3 (Deltaex7-8) pigs, Cln3 (Deltaex7-8) Mice, and CLN3-Affected Individuals. Biomark Insights. 2022;17:11772719221107765. doi:10.1177/11772719221107765 [PubMed: 36212622]
6. Laqtom NN, Dong W, Medoh UN, et al. CLN3 is required for the clearance of glycerophosphodiesters from lysosomes. Nature. Sep 2022;609(7929):1005–1011. doi:10.1038/s41586-022-05221-y
7. Chan CH, Mitchison HM, Pearce DA. Transcript and in silico analysis of CLN3 in juvenile neuronal ceroid lipofuscinosis and associated mouse models. Hum Mol Genet. Nov 1 2008;17(21):3332–9. doi:10.1093/hmg/ddn228 [PubMed: 18678598]
8. Miller JN, Chan CH, Pearce DA. The role of nonsense-mediated decay in neuronal ceroid lipofuscinosis. Hum Mol Genet. Jul 1 2013;22(13):2723–34. doi:10.1093/hmg/ddt120 [PubMed: 23539563]
9. Cotman SL, Vrbanac V, Lebel LA, et al. Cln3(Deltaex7/8) knock-in mice with the common JNCL mutation exhibit progressive neurologic disease that begins before birth. Hum Mol Genet. Oct 15 2002;11(22):2709–21. doi:10.1093/hmg/11.22.2709 [PubMed: 12374761]
10. Kitzmuller C, Haines RL, Codlin S, Cutler DF, Mole SE. A function retained by the common mutant CLN3 protein is responsible for the late onset of juvenile neuronal ceroid lipofuscinosis. Hum Mol Genet. Jan 15 2008;17(2):303–12. doi:10.1093/hmg/ddm306 [PubMed: 17947292]
11. Minnis CJ, Townsend S, Petschnigg J, et al. Global network analysis in Schizosaccharomyces pombe reveals three distinct consequences of the common 1-kb deletion causing juvenile CLN3 disease. Sci Rep. Mar 18 2021;11(1):6332. doi:10.1038/s41598-021-85471-4 [PubMed: 33737578]
12. Gaetani L, Paolini Paoletti F, Bellomo G, et al. CSF and Blood Biomarkers in Neuroinflammatory and Neurodegenerative Diseases: Implications for Treatment. Trends Pharmacol Sci. Dec 2020;41(12):1023–1037. doi:10.1016/j.tips.2020.09.011 [PubMed: 33127098]
13. Ren AH, Diamandis EP, Kulasingam V. Uncovering the Depths of the Human Proteome: Antibody-based Technologies for Ultrasensitive Multiplexed Protein Detection and Quantification. Mol Cell Proteomics. 2021;20:100155. doi:10.1016/j.mcpro.2021.100155 [PubMed: 34597790]
14. Campbell K, Cawley NX, Luke R, et al. Identification of cerebral spinal fluid protein biomarkers in Niemann-Pick disease, type C1. Biomark Res. Jan 31 2023;11(1):14. doi:10.1186/s40364-023-00448-x [PubMed: 36721240]
15. Dang Do AN, Sinaii N, Masvekar RR, et al. Neurofilament light chain levels correlate with clinical measures in CLN3 disease. Genet Med. Apr 2021;23(4):751–757. doi:10.1038/s41436-020-01035-3 [PubMed: 33239751]
16. Wik L, Nordberg N, Broberg J, et al. Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis. Mol Cell Proteomics. 2021;20:100168. doi:10.1016/j.mcpro.2021.100168 [PubMed: 34715355]

17. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. May 7 1976;72:248–54. doi:10.1006/abio.1976.9999 [PubMed: 942051]
18. Sleat DE, Della Valle MC, Zheng H, Moore DF, Lobel P. The mannose 6-phosphate glycoprotein proteome. J Proteome Res. Jul 2008;7(7):3010–21. doi:10.1021/pr800135v [PubMed: 18507433]
19. Sleat DE, Lackland H, Wang Y, et al. The human brain mannose 6-phosphate glycoproteome: a complex mixture composed of multiple isoforms of many soluble lysosomal proteins. Proteomics. Apr 2005;5(6):1520–32. doi:10.1002/pmic.200401054 [PubMed: 15789345]
20. Sleat DE, Sun P, Wiseman JA, et al. Extending the mannose 6-phosphate glycoproteome by high resolution/accuracy mass spectrometry analysis of control and acid phosphatase 5-deficient mice. Mol Cell Proteomics. Jul 2013;12(7):1806–17. doi:10.1074/mcp.M112.026179 [PubMed: 23478313]
21. McAlister GC, Nusinow DP, Jedrychowski MP, et al. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. Anal Chem. Jul 15 2014;86(14):7150–8. doi:10.1021/ac502040v [PubMed: 24927332]
22. Beavis RC. Using the global proteome machine for protein identification. Methods Mol Biol. 2006;328:217–28. doi:10.1385/1-59745-026-X:217 [PubMed: 16785652]
23. Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. J Proteome Res. Nov-Dec 2004;3(6):1234–42. doi:10.1021/pr049882h [PubMed: 15595733]
24. Tannous A, Boonen M, Zheng H, et al. Comparative Analysis of Quantitative Mass Spectrometric Methods for Subcellular Proteomics. J Proteome Res. Apr 3 2020;19(4):1718–1730. doi:10.1021/acs.jproteome.9b00862 [PubMed: 32134668]
25. Jiang Y, Zhou X, Ip FC, et al. Large-scale plasma proteomic profiling identifies a high-performance biomarker panel for Alzheimer's disease screening and staging. Alzheimers Dement. Jan 2022;18(1):88–102. doi:10.1002/alz.12369 [PubMed: 34032364]
26. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. Oct 25 2005;102(43):15545–50. doi:10.1073/pnas.0506580102 [PubMed: 16199517]
27. Delaby C, Alcolea D, Carmona-Iragui M, et al. Differential levels of Neurofilament Light protein in cerebrospinal fluid in patients with a wide range of neurodegenerative disorders. Sci Rep. Jun 8 2020;10(1):9161. doi:10.1038/s41598-020-66090-x [PubMed: 32514050]
28. Pinteac R, Montalban X, Comabella M. Chitinases and chitinase-like proteins as biomarkers in neurologic disorders. Neurol Neuroimmunol Neuroinflamm. Jan 2021;8(1):e921. doi:10.1212/NXI.0000000000000921 [PubMed: 33293459]
29. Drobny A, Prieto Huarcaya S, Dobert J, et al. The role of lysosomal cathepsins in neurodegeneration: Mechanistic insights, diagnostic potential and therapeutic approaches. Biochim Biophys Acta Mol Cell Res. Jul 2022;1869(7):119243. doi:10.1016/j.bbamcr.2022.119243 [PubMed: 35217144]
30. Teunissen CE, Elias N, Koel-Simmelink MJ, et al. Novel diagnostic cerebrospinal fluid biomarkers for pathologic subtypes of frontotemporal dementia identified by proteomics. Alzheimers Dement (Amst). 2016;2:86–94. doi:10.1016/j.dadm.2015.12.004 [PubMed: 27239539]
31. Folch J, Pedros I, Patraca I, et al. Neuroprotective and anti-ageing role of leptin. J Mol Endocrinol. Dec 2012;49(3):R149–56. doi:10.1530/JME-12-0151 [PubMed: 22967480]
32. McGrath ER, Himali JJ, Levy D, et al. Plasma EFEMP1 Is Associated with Brain Aging and Dementia: The Framingham Heart Study. J Alzheimers Dis. 2022;85(4):1657–1666. doi:10.3233/JAD-215053 [PubMed: 34958018]
33. Bolte AC, Shapiro DA, Dutta AB, et al. The meningeal transcriptional response to traumatic brain injury and aging. Elife. Jan 3 2023;12 doi:10.7554/eLife.81154
34. Kuan PF, Clouston S, Yang X, Kotov R, Bromet E, Luft BJ. Molecular linkage between post-traumatic stress disorder and cognitive impairment: a targeted proteomics study of World Trade Center responders. Transl Psychiatry. Aug 4 2020;10(1):269. doi:10.1038/s41398-020-00958-4 [PubMed: 32753605]

35. Rippe C, Erlanson-Albertsson C, Lindqvist A. Consequences of metabolic challenges on hypothalamic colipase and PLRP2 mRNA in rats. Brain Res. Dec 14 2007;1185:152–7. doi:10.1016/j.brainres.2007.09.022 [PubMed: 17936733]
36. Turkmen S, Guo G, Garshasbi M, et al. CA8 mutations cause a novel syndrome characterized by ataxia and mild mental retardation with predisposition to quadrupedal gait. PLoS Genet. May 2009;5(5):e1000487. doi:10.1371/journal.pgen.1000487 [PubMed: 19461874]
37. Hikmat O, Isohanni P, Keshavan N, et al. Expanding the phenotypic spectrum of BCS1L-related mitochondrial disease. Ann Clin Transl Neurol. Nov 2021;8(11):2155–2165. doi:10.1002/acn3.51470 [PubMed: 34662929]
38. Baik JY, Kim M, Bao J, Long Q, Shen L, Alzheimer's Disease Neuroimaging I. Identifying Alzheimer's genes via brain transcriptome mapping. BMC Med Genomics. May 19 2022;15(Suppl 2):116. doi:10.1186/s12920-022-01260-6 [PubMed: 35590321]
39. Mastrokolias A, Ariyurek Y, Goeman JJ, et al. Huntington's disease biomarker progression profile identified by transcriptome sequencing in peripheral blood. Eur J Hum Genet. Oct 2015;23(10):1349–56. doi:10.1038/ejhg.2014.281 [PubMed: 25626709]
40. Chaplot K, Jarvela TS, Lindberg I. Secreted Chaperones in Neurodegeneration. Front Aging Neurosci. 2020;12:268. doi:10.3389/fnagi.2020.00268 [PubMed: 33192447]
41. Balasubramanian M, Schirwani S. ASXL3-Related Disorder. 2020 Nov 5. In: Adam MP, Everman DB, Mirzaa GM, et al., eds. GeneReviews [Internet]. Seattle (WA): University of Washington, Seattle; 1993–2023.
42. Sarnowski C, Satizabal CL, DeCarli C, et al. Whole genome sequence analyses of brain imaging measures in the Framingham Study. Neurology. Jan 16 2018;90(3):e188–e196. doi:10.1212/WNL.0000000000004820 [PubMed: 29282330]
43. Wright GEB, Caron NS, Ng B, et al. Gene expression profiles complement the analysis of genomic modifiers of the clinical onset of Huntington disease. Hum Mol Genet. Sep 29 2020;29(16):2788–2802. doi:10.1093/hmg/ddaa184 [PubMed: 32898862]
44. Sengupta U, Montalbano M, McAllen S, Minuesa G, Kharas M, Kayed R. Formation of Toxic Oligomeric Assemblies of RNA-binding Protein: Musashi in Alzheimer's disease. Acta Neuropathol Commun. Oct 26 2018;6(1):113. doi:10.1186/s40478-018-0615-0 [PubMed: 30367664]
45. Dubos A, Castells-Nobau A, Meziane H, et al. Conditional depletion of intellectual disability and Parkinsonism candidate gene ATP6AP2 in fly and mouse induces cognitive impairment and neurodegeneration. Hum Mol Genet. Dec 1 2015;24(23):6736–55. doi:10.1093/hmg/ddv380 [PubMed: 26376863]
46. Windpassinger C, Schoser B, Straub V, et al. An X-linked myopathy with postural muscle atrophy and generalized hypertrophy, termed XMPMA, is caused by mutations in FHL1. Am J Hum Genet. Jan 2008;82(1):88–99. doi:10.1016/j.ajhg.2007.09.004 [PubMed: 18179888]
47. Nielsen JE, Honore B, Vestergard K, et al. Shotgun-based proteomics of extracellular vesicles in Alzheimer's disease reveals biomarkers involved in immunological and coagulation pathways. Sci Rep. Sep 16 2021;11(1):18518. doi:10.1038/s41598-021-97969-y [PubMed: 34531462]
48. Lopez-Perez O, Bernal-Martin M, Hernaiz A, et al. BAMBI and CHGA in Prion Diseases: Neuropathological Assessment and Potential Role as Disease Biomarkers. Biomolecules. May 2 2020;10(5):706. doi:10.3390/biom10050706 [PubMed: 32370154]
49. Miller JA, Woltjer RL, Goodenbour JM, Horvath S, Geschwind DH. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. Genome Med. 2013;5(5):48. doi:10.1186/gm452 [PubMed: 23705665]
50. Ayoglu B, Chaouch A, Lochmuller H, et al. Affinity proteomics within rare diseases: a BIO-NMD study for blood biomarkers of muscular dystrophies. EMBO Mol Med. Jul 2014;6(7):918–36. doi:10.15252/emmm.201303724 [PubMed: 24920607]
51. Nogradi A, Kelly C, Carter ND. Localization of acetazolamide-resistant carbonic anhydrase III in human and rat choroid plexus by immunocytochemistry and in situ hybridisation. Neurosci Lett. Mar 19 1993;151(2):162–5. doi:10.1016/0304-3940(93)90011-9 [PubMed: 8506074]

52. Loh SH, Francescut L, Lingor P, Bahr M, Nicotera P. Identification of new kinase clusters required for neurite outgrowth and retraction by a loss-of-function RNA interference screen. Cell Death Differ. Feb 2008;15(2):283–98. doi:10.1038/sj.cdd.4402258 [PubMed: 18007665]
53. Liang WS, Dunckley T, Beach TG, et al. Neuronal gene expression in non-demented individuals with intermediate Alzheimer's Disease neuropathology. Neurobiol Aging. Apr 2010;31(4):549–66. doi:10.1016/j.neurobiolaging.2008.05.013 [PubMed: 18572275]
54. Ramos J, Caywood LJ, Prough MB, et al. Genetic variants in the SHISA6 gene are associated with delayed cognitive impairment in two family datasets. Alzheimers Dement. May 1 2022;19(2):611–620. doi:10.1002/alz.12686 [PubMed: 35490390]
55. Ramos-Molina B, Molina-Vega M, Fernandez-Garcia JC, Creemers JW. Hyperphagia and Obesity in Prader(−)Willi Syndrome: PCSK1 Deficiency and Beyond? Genes (Basel). Jun 7 2018;9(6):288. doi:10.3390/genes9060288 [PubMed: 29880780]
56. Kizuka Y, Kitazume S, Fujinawa R, et al. An aberrant sugar modification of BACE1 blocks its lysosomal targeting in Alzheimer's disease. EMBO Mol Med. Feb 2015;7(2):175–89. doi:10.15252/emmm.201404438 [PubMed: 25592972]
57. Higginbotham L, Ping L, Dammer EB, et al. Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic Alzheimer's disease. Sci Adv. Oct 2020;6(43):eaaz9360. doi:10.1126/sciadv.aaz9360 [PubMed: 33087358]
58. Sathe G, Na CH, Renuse S, et al. Quantitative Proteomic Profiling of Cerebrospinal Fluid to Identify Candidate Biomarkers for Alzheimer's Disease. Proteomics Clin Appl. Jul 2019;13(4):e1800105. doi:10.1002/prca.201800105 [PubMed: 30578620]
59. Kang K, Bae JH, Han K, Kim ES, Kim TO, Yi JM. A Genome-Wide Methylation Approach Identifies a New Hypermethylated Gene Panel in Ulcerative Colitis. Int J Mol Sci. Aug 9 2016;17(8):1291. doi:10.3390/ijms17081291 [PubMed: 27517910]
60. Gaspari S, Quenneville S, Rodriguez Sanchez-Archidona A, Thorens B, Croizier S. Structural and molecular characterization of paraventricular thalamic glucokinase-expressing neuronal circuits in the mouse. J Comp Neurol. Aug 2022;530(11):1773–1949. doi:10.1002/cne.25312 [PubMed: 35303367]
61. Shen Y, Zhang F, Li F, et al. Loss-of-function mutations in QRICH2 cause male infertility with multiple morphological abnormalities of the sperm flagella. Nat Commun. Jan 25 2019;10(1):433. doi:10.1038/s41467-018-08182-x [PubMed: 30683861]
62. Kumble S, Levy AM, Punetha J, et al. The clinical and molecular spectrum of QRICH1 associated neurodevelopmental disorder. Hum Mutat. Feb 2022;43(2):266–282. doi:10.1002/humu.24308 [PubMed: 34859529]
63. Yang YX, Wu ML, Du O, Duan W, Du JR. Bioinformatic identification of key biomarkers in the brain and blood involved in ischemia preconditioning protection against acute ischemic stroke in mice. Int J Clin Exp Med. 2021;14(4):1628-+.
64. Tajerian M, Hung V, Khan H, et al. Identification of KRT16 as a target of an autoantibody response in complex regional pain syndrome. Exp Neurol. Jan 2017;287:14–20. doi:10.1016/j.expneurol.2016.10.011 [PubMed: 27773721]
65. Heier CR, Zhang AP, Nguyen NY, et al. Multi-Omics Identifies Circulating miRNA and Protein Biomarkers for Facioscapulohumeral Dystrophy. J Pers Med. Dec 2020;10(4):236. doi: 10.3390/jpm10040236 [PubMed: 33228131]
66. Zhang XL, Zhu CC, Beecham G, et al. A rare missense variant of CASP7 is associated with familial late-onset Alzheimer's disease. Alzheimers & Dementia. Mar 2019;15(3):441–452. doi:10.1016/j.jalz.2018.10.005
67. Mattisson IY, Bjorkbacka H, Wigren M, et al. Elevated Markers of Death Receptor-Activated Apoptosis are Associated with Increased Risk for Development of Diabetes and Cardiovascular Disease. Ebiomedicine. Dec 2017;26:187–197. doi:10.1016/j.ebiom.2017.11.023 [PubMed: 29208468]
68. Nakamoto C, Durward E, Horie M, Nakamoto M. Nell2 regulates the contralateral-versus-ipsilateral visual projection as a domain-specific positional cue. Development. Feb 18 2019;146(4): dev170704. doi:10.1242/dev.170704 [PubMed: 30745429]

69. Pak JS, DeLoughery ZJ, Wang J, et al. NELL2-Robo3 complex structure reveals mechanisms of receptor activation for axon guidance. Nat Commun. Mar 20 2020;11(1):1489. doi:10.1038/s41467-020-15211-1 [PubMed: 32198364]
70. Dayon L, Nunez Galindo A, Wojcik J, et al. Alzheimer disease pathology and the cerebrospinal fluid proteome. Alzheimers Res Ther. Jul 18 2018;10(1):66. doi:10.1186/s13195-018-0397-4 [PubMed: 30021611]
71. Panza P, Sitko AA, Maischein HM, et al. The LRR receptor Islr2 is required for retinal axon routing at the vertebrate optic chiasm. Neural Dev. Oct 22 2015;10:23. doi:10.1186/s13064-015-0050-x [PubMed: 26492970]
72. Shen Q, Polom K, Williams C, et al. A targeted proteomics approach reveals a serum protein signature as diagnostic biomarker for resectable gastric cancer. EBioMedicine. Jun 2019;44:322–333. doi:10.1016/j.ebiom.2019.05.044 [PubMed: 31151932]
73. Ashton NJ, Janelidze S, Al Khleifat A, et al. A multicentre validation study of the diagnostic value of plasma neurofilament light. Nat Commun. Jun 7 2021;12(1):3400. doi:10.1038/s41467-021-23620-z [PubMed: 34099648]
74. Benatar M, Zhang L, Wang L, et al. Validation of serum neurofilaments as prognostic and potential pharmacodynamic biomarkers for ALS. Neurology. Jul 7 2020;95(1):e59–e69. doi:10.1212/WNL.0000000000009559 [PubMed: 32385188]
75. Petrera A, von Toerne C, Behler J, et al. Multiplatform Approach for Plasma Proteomics: Complementarity of Olink Proximity Extension Assay Technology to Mass Spectrometry-Based Protein Profiling. J Proteome Res. Jan 1 2021;20(1):751–762. doi:10.1021/acs.jproteome.0c00641 [PubMed: 33253581]
76. Lebrun AH, Moll-Khosrawi P, Pohl S, et al. Analysis of potential biomarkers and modifier genes affecting the clinical course of CLN3 disease. Mol Med. 2011;17(11-12):1253–61. doi:10.2119/molmed.2010.00241 [PubMed: 21863212]
77. Llavero Hurtado M, Fuller HR, Wong AMS, et al. Proteomic mapping of differentially vulnerable pre-synaptic populations identifies regulators of neuronal stability in vivo. Sci Rep. Sep 29 2017;7(1):12412. doi:10.1038/s41598-017-12603-0 [PubMed: 28963550]
78. Schmidtke C, Tiede S, Thelen M, et al. Lysosomal proteome analysis reveals that CLN3-defective cells have multiple enzyme deficiencies associated with changes in intracellular trafficking. J Biol Chem. Jun 14 2019;294(24):9592–9604. doi:10.1074/jbc.RA119.008852 [PubMed: 31040178]
79. Sleat DE, Tannous A, Sohar I, et al. Proteomic Analysis of Brain and Cerebrospinal Fluid from the Three Major Forms of Neuronal Ceroid Lipofuscinosis Reveals Potential Biomarkers. J Proteome Res. Oct 6 2017;16(10):3787–3804. doi:10.1021/acs.jproteome.7b00460 [PubMed: 28792770]
80. Ries M, Schaefer E, Luhrs T, et al. Critical assessment of chitotriosidase analysis in the rational laboratory diagnosis of children with Gaucher disease and Niemann-Pick disease type A/B and C. J Inherit Metab Dis. Oct 2006;29(5):647–52. doi:10.1007/s10545-006-0363-3 [PubMed: 16972172]
81. Weinreb NJ, Aggio MC, Andersson HC, et al. Gaucher disease type 1: revised recommendations on evaluations and monitoring for adult patients. Semin Hematol. Oct 2004;41(4 Suppl 5):15–22. doi:10.1053/j.seminhematol.2004.07.010
82. Matsuhashi S, Noji S, Koyama E, et al. New gene, nel, encoding a M(r) 93 K protein with EGF-like repeats is strongly expressed in neural tissues of early stage chick embryos. Dev Dyn. Jun 1995;203(2):212–22. doi:10.1002/aja.1002030209 [PubMed: 7655083]
83. Kim HR, Kim DH, An JY, et al. NELL2 Function in Axon Development of Hippocampal Neurons. Mol Cells. Jun 2020;43(6):581–589. doi:10.14348/molcells.2020.0032 [PubMed: 32597395]
84. Shaker MR, Kahtan A, Prasad R, et al. Neural Epidermal Growth Factor-Like Like Protein 2 Is Expressed in Human Oligodendroglial Cell Types. Front Cell Dev Biol. Feb 21 2022;10doi:ARTN 803061 10.3389/fcell.2022.803061
85. Desai J, Shannon ME, Johnson MD, et al. Nell1-deficient mice have reduced expression of extracellular matrix proteins causing cranial and vertebral defects. Hum Mol Genet. Apr 15 2006;15(8):1329–41. doi:10.1093/hmg/ddl053 [PubMed: 16537572]

86. Nakamura R, Nakamoto C, Obama H, Durward E, Nakamoto M. Structure-Function Analysis of Nel, a Thrombospondin-1-like Glycoprotein Involved in Neural Development and Functions. J Biol Chem. Jan 27 2012;287(5):3282–3291. doi:10.1074/jbc.M111.281485 [PubMed: 22157752]

87. Mandai K, Guo T, St Hillaire C, et al. LIG family receptor tyrosine kinase-associated proteins modulate growth factor signals during neural development. Neuron. Sep 10 2009;63(5):614–27. doi:10.1016/j.neuron.2009.07.031 [PubMed: 19755105]

88. Abudureyimu S, Asai N, Enomoto A, et al. Essential Role of Linx/Islr2 in the Development of the Forebrain Anterior Commissure. Sci Rep-Uk. May 8 2018;8(1):7292. doi: 10.1038/s41598-018-24064-0

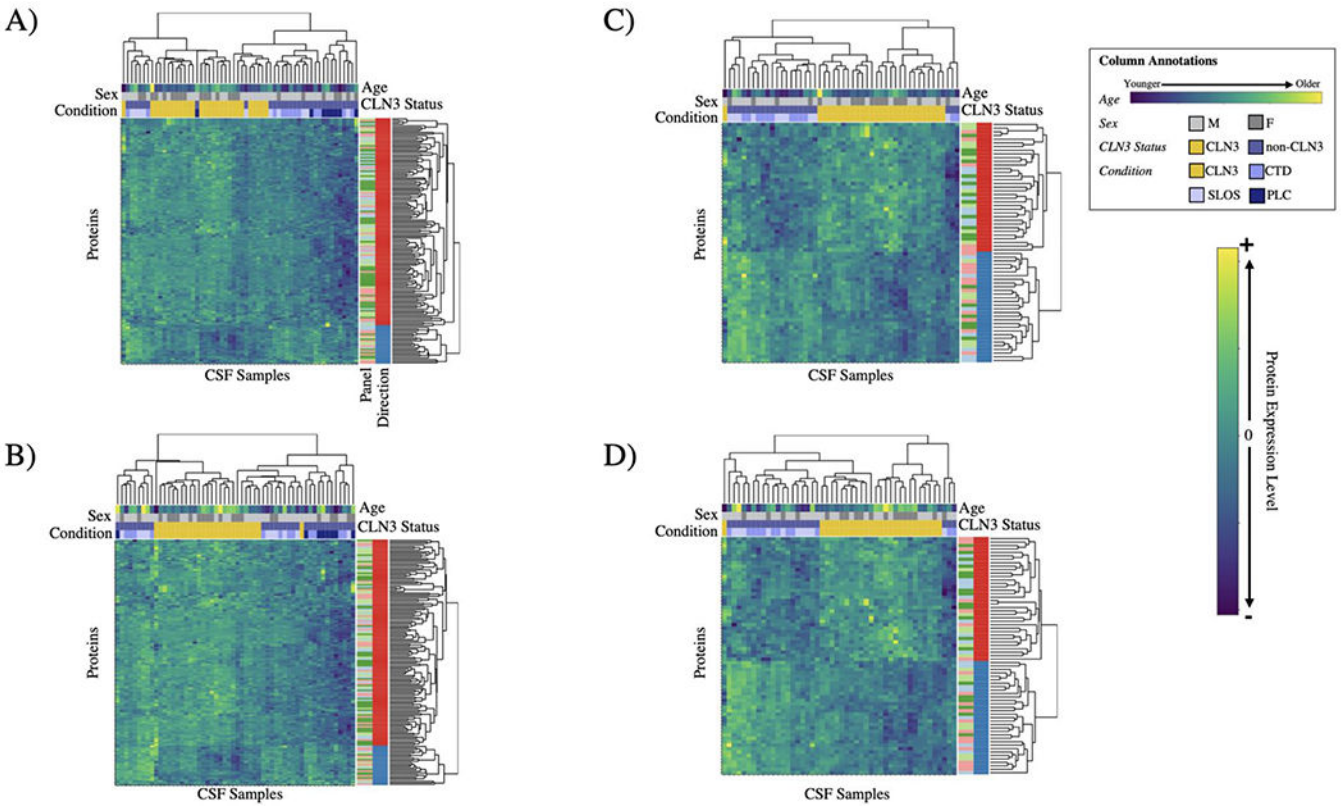


Figure 1.

Clustering of cerebrospinal fluid samples analyzed using proximal extension assay. Samples that have adjusted p-value < 0.1 are grouped based on expression level pattern. A) All CLN3 versus all non-CLN3. B) Syndromic CLN3 versus all non-CLN3. C) All CLN3 versus non-PLC. D) Syndromic CLN3 versus non-PLC. Color gradient bar depicts ranges of age or protein expression levels. Panel: light green = inflammatory; pink = oncology; dark green = neurology; light blue = cardiometabolic. Direction: red = increased expression, blue = decreased expression in CLN3 compared to non-CLN3.

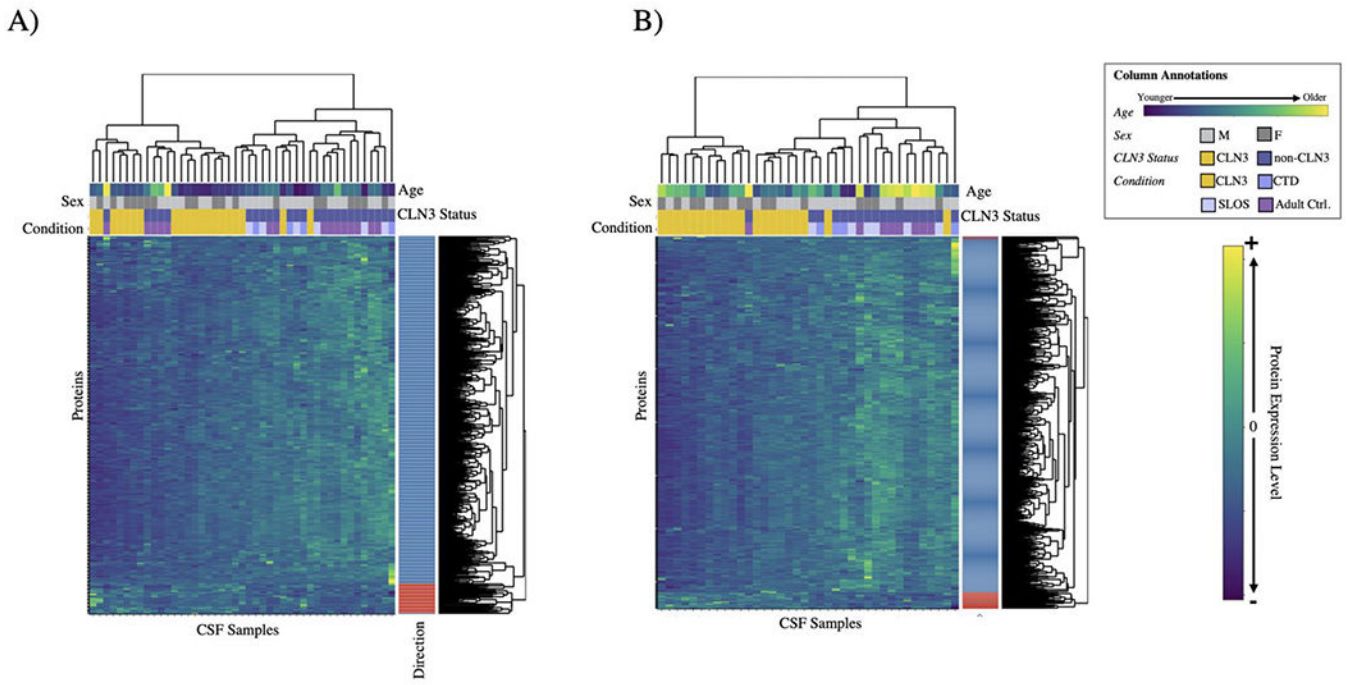


Figure 2.

Clustering of cerebrospinal fluid samples analyzed using mass spectrometry. Samples that have adjusted p-value < 0.1 are grouped based on expression level pattern. A) CLN3 versus all non-CLN3. B) CLN3 versus non-CLN3 < 24 years of age. Color gradient bar depicts ranges of age or protein expression levels. Direction: red = increased expression in CLN3 compared to non-CLN3; blue = decreased expression in CLN3 compared to non-CLN3.

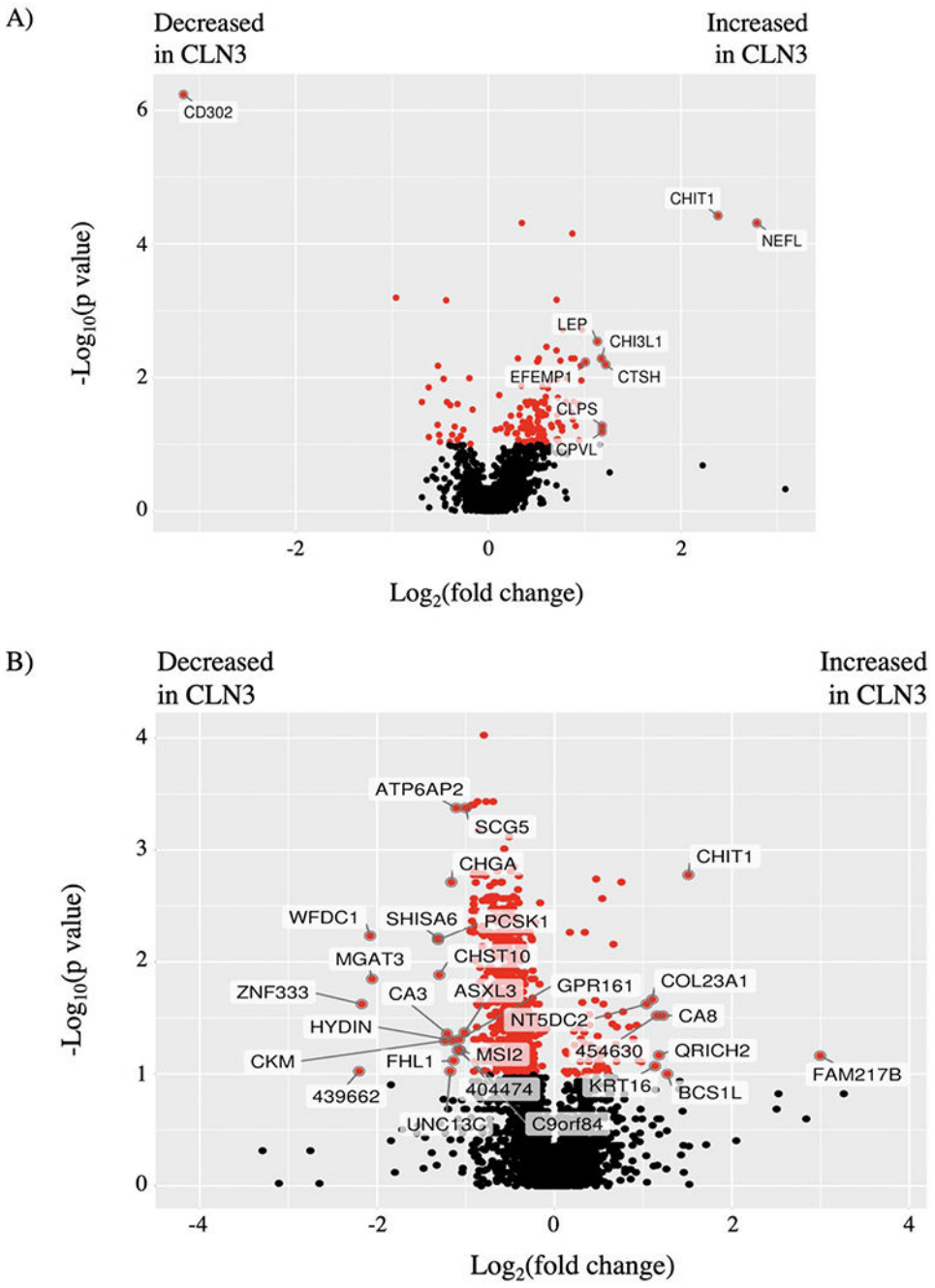


Figure 3.

Volcano plot depicting differential level of biomarkers in CLN3 versus non-CLN3 cerebrospinal fluid samples analyzed using A) proximal extension assay and B) mass spectrometry. Protein level in CSF samples plotted as log2 fold change on the x-axis and −log10 adjusted p-value on the y-axis. Red: adjusted p-value < 0.1. Black: adjusted p-value ≥ 0.1. Labeled proteins are those with adjusted p-value < 0.1 and absolute log2 fold change ≥ 1. For visual clarity, protein labels containing “ENSP00000…” are shortened to their unique numbers only.

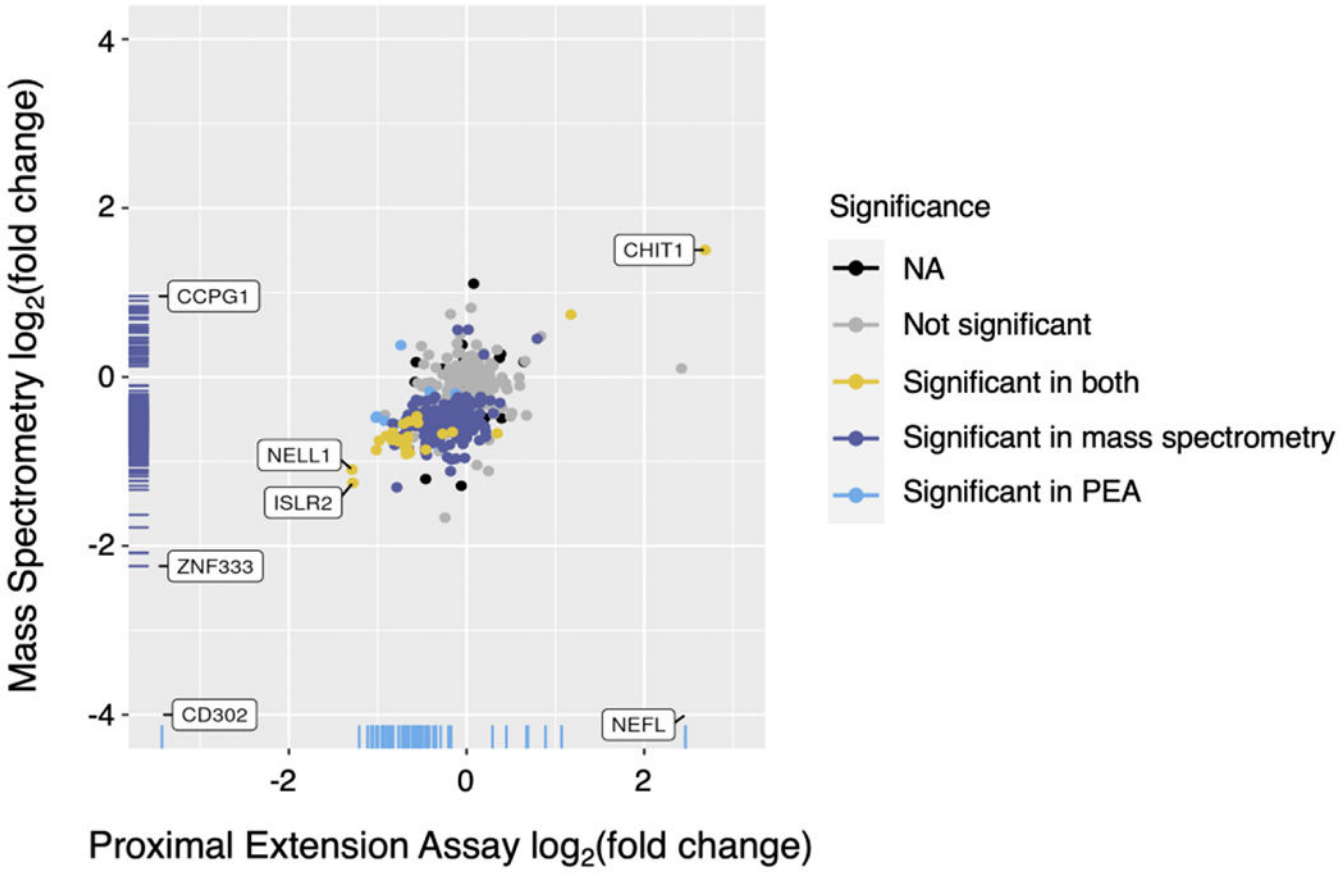


Figure 4.

Differential expression of biomarkers in cerebrospinal fluid samples analyzed using Olink® proximal extension assay (PEA) or mass spectrometry. The scatterplot shows the log2(fold change) for each protein using data from PEA (x-axis) and from mass spectrometry (y-axis). Internal points are for proteins found in *both* PEA and mass spectrometry datasets. Proteins found only in the PEA dataset are illustrated in the rug plot on the x-axis; proteins found only in the mass spectrometry dataset are illustrated in the rug plot on the y-axis. Points within the NA significance group reported an adjusted p-value of NA (not available) in at least one of the two methods. Significance for biomarker discovery is $p < 0.1$.

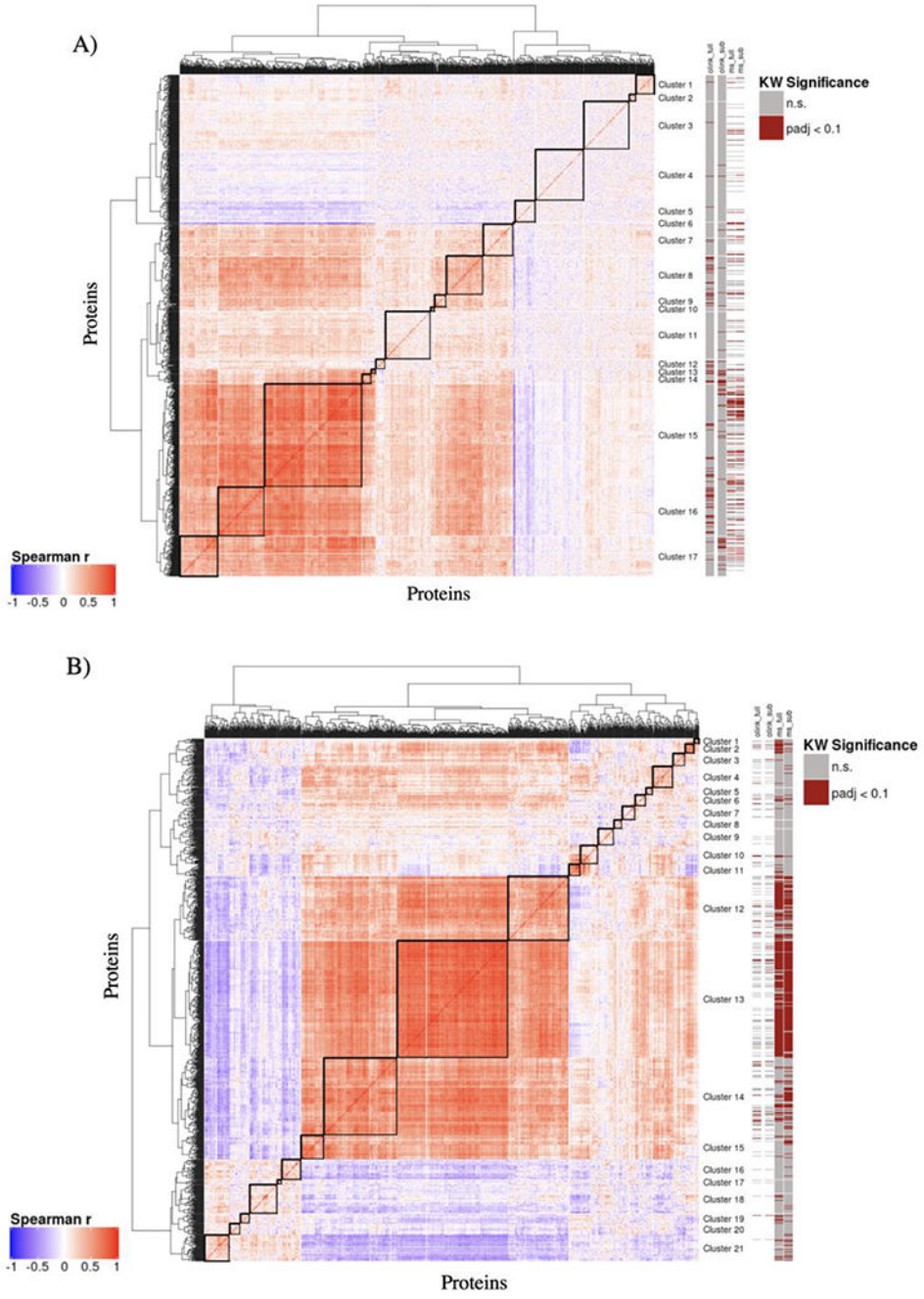


Figure 5.

Hierarchical clustering of expression level by protein-protein correlations. CSF protein expression measured by A) proximal extension assay and B) mass spectrometry. Clusters represent group of proteins with similar correlative patterns of change in levels when comparing CLN3 versus non-CLN3 samples. The significance vertical bars on the right depict the assay type, contrasts used to calculate the correlation coefficients for proteins within each cluster, and the Kruskal-Wallis adjusted p-value significance. olink/ms_full: all CLN3 vs. all non-CLN3. olink/ms_sub: subset of CLN3 vs. non-CLN3 used in both

proximal extension assay and mass spectrometry. White spaces in the vertical bars denote that the protein(s) were not included in the assay method.

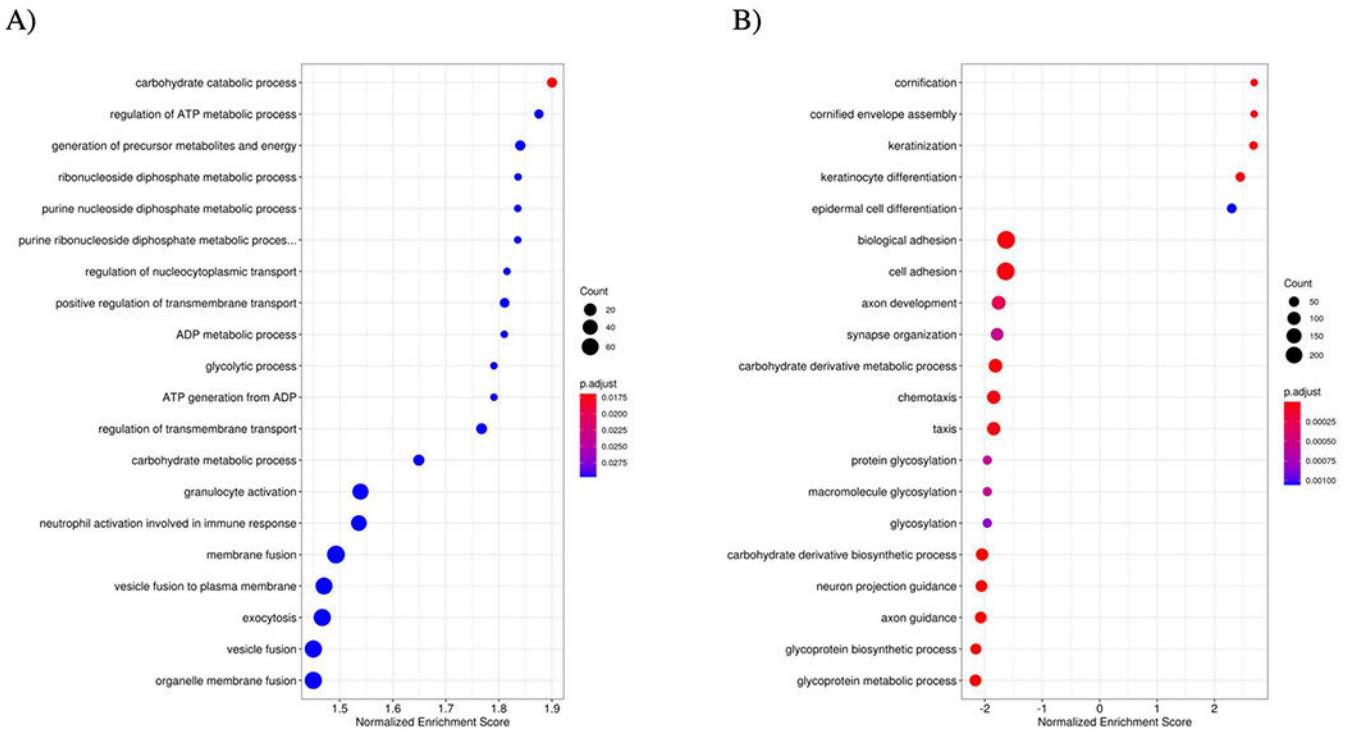


Figure 6.

Gene set enrichment analysis for the top 20 enriched biological processes based on protein differential abundances as measured by A) proximal extension analysis and B) mass spectrometry. Count: number of proteins annotated within the given category. Normalized Enrichment Score (NES) quantifies the degree of overrepresentation in the ranked list of protein log2 fold change. Sign of the NES corresponds to differential abundance directionality of the majority of proteins in the category (e.g. positive NES = primarily positive log2 fold change). P.adjust: likelihood that the enrichment scores based on the rank list of protein differential abundances is different from a random permutation (smaller adjusted p-value = more likely to be different).

Table 1.

Demographics of CLN3 and non-CLN3 individuals from whom cerebrospinal fluid samples were collected.

Group	Proximal Extension Assay					Mass Spectrometry				
	n	Age (years)		Sex (n)		n	Age (years)		Sex (n)	
		Median	Range	Female	Male		Median	Range	Female	Male
CLN3	28	10.2	0.1–29.1	14	14	20	10.7	6.0–20.7	10	10
Non-CLN3	32	10.1	0.4–21.0	10	22	25	22.3	2.6–57	9	16
Pediatric Laboratory	10	11.5	0.4–20.0	5	5	n/a	n/a	n/a	n/a	n/a
Adult Commercial	n/a	n/a	n/a	n/a	n/a	15	24	21–57	6	9
Non-PLC	22	8.0	2.6–21.0	5	17	10	10.3	2.6–21.0	3	7
CTD	12	7.0	3.2–14.3	0	12	4	10.2	3.2–14.3	0	4
SLOS	10	9.4	2.6–21.0	5	5	6	10.3	2.6–21.0	3	3
Less than 24 years	n/a	n/a	n/a	n/a	n/a	18	19.0	2.6–24	6	12

CTD: creatine transporter deficiency. n/a: not applicable. PLC: pediatric laboratory control. SLOS: Smith-Lemli-Opitz syndrome.

Table 2.

Proteins differentially abundant in CLN3 as compared to non-CLN3 cerebrospinal fluid samples as identified by A) Olink® proximal extension assay and B) mass spectrometry. Proteins included in the table are those with absolute log2 fold change ≥ 1 and adjusted p-value < 0.1. Adjusted p-values are derived from Kruskal-Wallis method with multiple comparison corrections applied. Correlated proteins are derived from comparisons of differential abundance and hierarchical clustering based on the correlation coefficients.

A)			
Protein	**Log2 FC**	**Adjusted p-value**	**Correlated proteins**
NEFL	2.79	4.9E-05	Cluster 12: CD63, FCAR, SCG3, LEP
CHIT1	2.39	3.8E-05	Cluster 8: CD300LF, MSLN, SIGLEC9, ASAH2, CC2D1A, FRZB, EDAR, LILRB4, LY6D, SCRN1, CHI3L1
CTSH	1.22	6.4E-03	Cluster 8: CALB1, KLK13, MIA, BAG3, CDH3, FRZB, GLB1, MYOC, CCL23, CTRC, CXCL3, CXCL6, LILRB4, SCRN1, CXCL5, EDIL3, GLRX, HYAL1, SFTPD
CLPS	1.18	5.3E-02	Cluster 16: AMIGO2, ANGPTL7, CAPG, CDNF, CPXM1, CREG1, CTSF, EBI1_IL27, EDA2R, EPS8L2, GFAP, IDUA, IL13RA1, INPP1, KDR, KLK11, LEFTY2, NT5E, SMAD5, TNFRSF12A, XCL1, CCL19, CD109, CD300C, CPA2, CPM, CTSS, CXCL13, DDR1, DSG2, EFNA1, EFNA4, F11R, FOLR2, FUT3_FUT5, GGT1, GUCA2A, IL17RA, IL1RAP, LAMP2, LILRA2, MESD, MSR1, OGN, OXT, PDCD5, PILRA, PLA2G7, SERPINB1, SUSD2, THBS2, TMPRSS5, CCL22, CLEC7A, EPHA1, ERBB3, FLT3LG, FSTL3, IL15, IL17D, IL18R1, IL1R2, LAIR1, LAMA4, LGALS9, MATN2, MGLL, MILR1, OMD, OSCAR, PRELP, PRSS8, SCGB3A2, SIGLEC1, SIGLEC10, TNFRSF11B, TNFRSF13B, TREMS, VEGFD, ACP5, ADA2, ADAMTS13, ANGPTL1, BMP6, BPIFB1, C2, CCL15, CCL16, CD163, CD209, CPB1, CST6, CSTB, CTSD, CXC16, DKK3, DPP4, EFEMP1, GDF15, GUSB, ICAM1, IGFBP2, IGFBP6, IL18BP, IL6R, ITGB2, LGALS3, LILRB1, LILRB2, LTBP2, MARCO, MFAP5, OLR1, OSMR, PCOLCE, PCSK9, PLTP, PLXNB3, PRCP, PROC, RARRES2, RNASET2, SOST, TCN2, TFF3, TFPI, TGFBI, TIMP1, VCAM1
CPVL	1.18	6.5E-02	Cluster 8: ENO1, FRZB, WARS, CXCL6, LGMN, TPP1, DPP7, HYAL1
CHI3L1	1.17	5.2E-03	Cluster 8: CD300LF, CEACAM1, MIA, MSLN, SIGLEC6, SIGLEC9, ASAH2, CDCP1, CGA, ENO1, FRZB, ITGAM, LPO, PIGR, SIGLEC5, TNFRSF8, VSIG4, WARS, CXCL6, EDAR, LILRB4, LY6D, TLR3, TPP1, ANG, ANPEP, CCL14, CDH1, CHIT1, FCGR2A, ICAM3, LCN2, PGLYRP1, PILRB, QDPR, SIGLEC7, THBD, TNFRSF10C, TNFSF13B
LEP	1.13	2.9E-03	Cluster 12: NEFL, FCAR, SCG3
EFEMP1	1.01	5.9E-03	Cluster 16: ADM, AMIGO2, ANGPTL7, CAPG, CDNF, CPXM1, CREG1, CTSF, EBI3_IL27, EDA2R, EPS8L2, GFAP, IDUA, INPP1, KDR, KLK11, LEFTY2, LYPD3, NT5E, SERPINA9, SMAD5, TNFRSF12A, XCL1, CCL19, CD109, CD300C, CLPS, CPA2, CPM, CTSS, CXCL13, DDR1, DSG2, EFNA1, EFNA4, F11R, FOLR2, FUT3_FUT5, GGT1, GRN, GUCA2A, IL17RA, IL1RAP, LAMP2, LILR2, MESD, MSR1, OGN, OXT, PDCDS, PECAM1, PILRA, PLA2G7, SERPINB1, SUSD2, THBS2, TMPRSSS, TNFRSF10A, TNFRSF1B, CCL22, CD48, CLEC7A, CRKL, EPHA1, ERBB2, FLT3LG, FSTL3, GLOD4, IL15, IL17D, IL18R1, IL1R2, LAIR1, LAMA4, LGALS9, MATN2, MGLL, MILR1, OMD, OSCAR, PRELP, PRSS8, SCG3A2, SELPLG, SIGLEC1, SIGLEC10, TNFRSF11B, TNFRSF13B, TREM2, VEGFD, ACP5, ADA2, ADAMTS13, ANGPTL1, BMP6, BPIFB1, C1, CCL15, CCL16, CD14, CD163, CD209, CPB1, CST6, CSTB, CTSB, CTSD, CTSZ, CXCL16, DKK3, DPP4, GDF15, GUSB, ICAM1, IGFBP2, IGFBP6, IL18BP, IL6R, ITGB2, LGALS3, LILRB1, LILRB2, LTBP2, MARCO, MFAP5, OLR1, OSMR, PCOLCE, PCSK9, PLTP, PLXNB3, PRCP, PROC, RARRES2, RNASET2, SOST, TCN2, TFF3, TFPI, TGFBI, TIMP1, VCAM1
CD302	−3.17	5.8E-07	Cluster 13: AREG, HS3ST3B1, ITGAV, SPINK6, YES1, SLC16A1, TNFRSF9, VCAN, FGF5, PAG1
B)			
Protein	**Log2 FC**	**Adjusted p-value**	**Correlated proteins**

FAM217B	3.00	6.9E-02	not in a correlation cluster
CHIT1	1.51	1.7E-03	Cluster 19: CHI3L1
BCS1L	1.27	1.0E-01	not in a correlation cluster
CA8	1.22	3.0E-02	not in a correlation cluster
QRICH2	1.18	6.8E-02	not in a correlation cluster
ENSP00000454630 (protein)	1.15	3.0E-02	Cluster 11: ALOX12B, ARG1, DSG1, DSP, ENSP00000357762, GSDMA, KRT1, KRT10, KRT14, KRT16, KRT17, KRT1, KRT3, KRT5, KRT6A, KRT6B, KRT77, KRT78, KRT80 KRT9, LGALSL, PSMA7, PSMB5, S100A8, S100A9, TGM3
KRT16	1.13	8.5E-02	Cluster 11: ALOX12B, ARG1, DSG1, DSP, ENSP00000454630, GSDMA, KRT1, KRT10, KRT14, KRT17, KRT2, KRT3, KRT5, KRT6A, KRT6B, KRT77, KRT78, KRT80, KRT9, S100A8, TGM3
COL23A1	1.10	2.2E-02	not in a correlation cluster
NT5DC2	1.04	2.4E-02	not in a correlation cluster
SCG5	−1.00	4.2E-04	Cluster 12: CTSF, A4GALT, ASTN2, ATP6AP2, B3GAT3, B4GALT1, CBLN2, CHRDL1, CLUL1, CNDP1, CORO1A, CPD, CPZ, CSPG4, DDR1, DKK3, DNAJC3, ECM1, ECM2, EFEMP2, ENSP00000431883, ENSP00000453973, ENSP00000470520, EPHA6, F13A1, FAM19A2, FGFR3, FRZB, GALNT6, GLG1, GNPTG, GPX3, GRIA4, GSN, HECTD3, HSP90B1, HTRA1, IGF2, IGSF21, IGSF8, KIAA0319L, KIAA1549L, KLK6, LAMAS, LCP1, LGALS1, LMOD1, LTBP1, LTBP4, MAN1B1, MAN1C1, MASP1, MDGA1, MEGF10, MFSD6L, MAT1, MGAT2, MGAT5, MSN, NELL1, NPDC1, OGN, OMD, OMG, PCDH17, PCDH8, PLEKHB1, PODXL2, PRNP, PROS1, SDF4, SEMA3C, SEPP1, SHISA6, SLITRK2, SLITRK3, SMO, SNED1, SORT1, SPOCK2, SSC5D, SUSD5, TF, TMEM25, TNFRSF21, TYRO3, WIPF2, ZMYND11
ASXL3	−1.01	4.3E-02	not in a correlation cluster
C9orf84	−1.06	6.2E-02	not in a correlation cluster
GPR161	−1.07	4.9E-02	not in a correlation cluster
MSI2	−1.07	6.1E-02	not in a correlation cluster
ENSP00000404474 (protein)	−1.09	6.2E-02	not in a correlation cluster
ATP6AP2	−1.11	4.2E-04	Cluster 12: CTSF, DPP7, GLA, GNS, A4GALT, ASTN2, B3GAT3, B4GALT1, CHRDL1, CLUL1, CNDP1, CORO1A, CPD, CPZ, CSPG4, DDR1, DKK3, DNAJC3, DSC3, ECM1, ECM2, EFEMP2, ENSP00000431883, ENSP00000453973, ENSP00000461497, ENSP00000470520, F13A1, FRZB, GALNT6, GLG1, GNPTG, GPX3, GRIA4, GSN, HECTD3, HSP90B1, HTRA1, IGF2, IGSF21, IGSF8, KIAA0319L, KLK6, LAMA5, LCP1, LGALS1, LMOD1, LTBP1, MAN1B1, MAN1C1, MASP1, MDGA1, MEGF10, MFSD6L, MGAT1, MGAT2, MGAT5, MSN, NELL1, OMD, OMG, PCDH8, PLEKHB1, PRNP, PROS1, SCG5, SDF4, SEMA3C, SEMA3F, SEPP1, SERPIND1, SH2B1, SHISA6, SMO, SNED1, SORT1, SPOCK2, SSC5D, SUSD5, TENM2, TF, TMEM25, TYRO3, WIPF2, ZNF217
FHL1	−1.14	7.6E-02	not in a correlation cluster
HYDIN	−1.15	5.0E-02	not in a correlation cluster
CHGA	−1.16	1.9E-03	Cluster 13: all proteins in cluster
UNC13C	−1.18	9.5E-02	not in a correlation cluster
CA3	−1.21	4.4E-02	Cluster 1: MB, PYGL
CKM	−1.23	5.1E-02	Cluster 6: WDR7
CHST10	−1.30	1.3E-02	Cluster 14: ASAH1, CTBS, CTSH, GALNS, GBA, GGH, GUSB, NCSTN, NEU1, NPC2, PLD3, RNASET2, TPP1, ADAM22, AEBP1, ALDOA, ANGPTL2, APOE, ASPH, ATP1B1, AXL, BST1, C12orf49, C1QC, C1QTNF1, C1QTNF3, CADM1, CCDC88A, CD109,

			CD163, CD248, CDH11, CDH15, CDH8, CGREF1, CLSTN3, CNTN2, COL12A1, COL4A1, CRKL, CRYM, CTHRC1, DDAH1, DDB1, DLD, DPP3, DSC2, EDIL3, EFNB2, ENO1, ENO2, ENPP2, ENPP6, ENSP00000256861, ENSP00000262325, ENSP00000290575, ENSP00000353656, ENSP00000370505, ENSP00000441566, FAT1, FBLN1, FBLN5, FMN2, FOLR1, FSCN1, GALNT16, GPC4, HDAC2, HEPH, HMCN1, INPP1, JAM2, KAT6B, LAMA4, LINGO3, LRRN1, MANEAL, MATN2, ME1, MET, MFAP4, MMP2, NDST1, NID1, NRP1, NSF, NTK3, NUCB1, OBSCN, OLMFL3, P4HB, PAPPA2, PCDHAC2, PCMT1, PCOLCE, PEBP1, PGLS, PLOD1, PLTP, PPPSC, PSAT1, PTGDS, RAB14, RGN, RUNX2, SCN3B, SCUBE2, SERPINB1, SERPINF1, SIL1, SOD1, SPARC, SPINT1, TENM1, TGFBI, TIMP1, TIMP4, TMEM235, TNXB, TPI1, UNCSC, VCAM1
SHISA6	−1.31	6.1E-03	Cluster 12: AGA, CTSF, GLA, PPT1, ADAM23, ADAMTS8, ANXAS, ATP6AP2, ATRN, B4GALT1, BGN, C4orf48, CAMK2A, CAMK2B, CBLN2, CBR1, CD14, CHRDL1, CKB, CLUL1, CNDP1, COCH, COL6A2, CORO1A, CPD, CSPG4, DCN, DDR1, DKK3, DNAJC3, DSC3, ECM1, ECM2, EFEMP2, ENSP00000360454, ENSP00000431883, ENSP00000453973, ENSP00000461497, EPHA6, FAM19A2, FGFR3, FGL2, FRZB, GALNT5, GALNT6, GLG1, GOLIM4, GPX3, GRIA4, GSN, HSP90B1, HTR1, IGF2, IGSF1, IGSF21, IGSF8, ISLR2, KIAA0319L, KIAA1549L, KLK6, LAMA5, LCP1, LTBP1, LTBP4, LYVE1, MAN1B1, MAN1C1, MASP1, MDGA1, MEGF10, MFSD6L, MGAT1, MGAT2, MGAT5, MSN, NPDC1, OGN, OLMF2, OMD, OMG, PCDH17, PCDH8, PLEKHB1, PLXND1, PODXL2, PRNP, PROC, RNASE1, SCG5, SDF4, SEMA3C, SEMA3F, SEPP1, SLITRK2, SLITRK3, SMO, SNED1, SOD3, SORL1, SORT1, SPOCK2, SSC5D, STC2, SUSD5, TCN2, TENM2, TF, TGFBR3, TIMP2, TMEM25, TNFRSF21, TYRO3, VASN, WIPF2, ZMYND11
PCSK1	−1.31	6.4E-03	Cluster 13: all proteins in cluster, except PRUNE2
MGAT3	−2.06	1.4E-02	not in a correlation cluster
WFDC1	−2.08	5.8E-03	not in a correlation cluster
ZNF333	−2.17	2.4E-02	not in a correlation cluster
ENSP00000439662 (protein)	−2.20	9.5E-02	not in a correlation cluster

Table 3.

Overlapping biomarkers identified by both proximal extension assay and mass spectrometry with A) any adjusted p value and B) adjusted p value < 0.1 in both methods. NA: no value. −: decreased level in CLN3 samples. +: increased level in CLN3 samples.

A)

	Proximal Extension Assay log2(fold change)		
Mass Spectrometry log2(fold change)	**−**	**+**	**NA**
−	267	121	1719
+	29	43	1114
NA	607	395	18

B)

	Proximal Extension Assay log2(fold change)		
Mass Spectrometry log2(fold change)	**−**	**+**	**NA**
−	23	1	434
+	0	2	44
NA	52	7	NA

Table 4.

Biomarkers identified by both proximal extension assay (PEA) and mass spectrometry (MS) with adjusted p-value (adj p-val) < 0.1 in both methods and concordant in the log2 fold change (FC) direction. **Bold**: protein markers with concordant changes, adjusted p-value < 0.1, and absolute log2 fold change ≥ 1 in both assays. Protein function is derived from NCBI database (https://www.ncbi.nlm.nih.gov/gene) and www.genecards.org (https://www.genecards.org. Accessed January 3, 2023).

Protein Symbols	PEA		MS		Protein Descriptions
	adj p-val	log2 FC	adj p-val	log2 FC	
CHI3L1	5.6E-02	1.18	1.4E-02	0.74	member of glycosyl hydrolase 18 family, no chitinase activity
CHIT1	**7.4E-03**	**2.69**	**8.7E-03**	**1.50**	**chitotriosidase**
ACAN	2.0E-02	−0.68	9.1E-02	−0.91	extracellular matrix protein of aggrecan/versican proteoglycan family
BSG	3.3E-02	−0.67	9.4E-02	−0.84	plasma membrane protein involved in neural network formation
CLSTN2	2.0E-02	−0.76	5.7E-03	−0.74	predicted Ca^{2+}-binding protein, synapse assembly/synaptic transmission
CRTAC1	8.4E-02	−0.67	1.1E-02	−0.72	glycosylated extracellular matrix protein, chondrocyte marker
DFFA	7.6E-02	−0.27	7.2E-02	−0.67	DNA fragmentation factor subunit A
EPHA10	7.2E-02	−0.99	5.7E-03	−0.75	receptor tyrosine kinase, mediator of neuronal cell mobility
GFRA2	6.5E-02	−0.67	6.8E-03	−0.75	cell surface receptor for neurotrophic factors
IGF2R	6.9E-02	−0.55	6.7E-03	−0.54	Insulin-like growth factor 2 and mannose 6-phosphate receptor
ISLR2	**1.7E-02**	**−1.28**	**5.7E-03**	**−1.25**	**predicted integral membrane protein, axon extension regulation**
JAM2	7.4E-02	−0.66	1.5E-02	−0.71	type I membrane protein of the junctional adhesion molecule family
KIT	2.3E-02	−1.02	6.7E-03	−0.87	glycosylated transmembrane receptor tyrosine kinase
L1CAM	6.9E-02	−0.77	7.6E-03	−0.76	axonal glycoprotein of the immunoglobulin supergene family
LRRN1	5.5E-02	−0.71	2.5E-02	−0.56	predicted membrane protein, synapse assembly regulation
NELL1	**6.5E-02**	**−1.29**	**1.4E-02**	**−1.10**	**cytoplasmic protein, cell growth regulation and differentiation**
NPTN	5.6E-02	−0.56	2.5E-02	−0.47	type I transmembrane protein, cell-cell/substrate interactions
NTRK2	4.8E-02	−0.65	2.4E-02	−0.53	neurotrophic tyrosine receptor kinase
NXPH1	9.2E-02	−0.69	1.0E-02	−0.84	secreted protein promoting dendrite-axon adhesion
PLXDC1	9.2E-02	−0.16	8.7E-03	−0.65	predicted to be involved in spinal cord development and angiogenesis
PTPRS	7.2E-02	−0.83	5.7E-03	−0.66	protein tyrosine phosphatase family, primary axonogenesis involvement
SMOC1	8.4E-03	−0.64	5.7E-03	−0.90	secreted protein, role in ocular and limb development
ST6GAL1	3.5E-02	−0.91	1.8E-02	−0.70	glycosyltransferase, generation of cell-surface carbohydrate determinants
TNR	7.4E-03	−0.46	5.7E-03	−0.86	extracellular matrix protein, neurite outgrowth/neural cell adhesion
VCAN	2.1E-02	−0.84	5.7E-03	−0.75	extracellular matrix chondroitin sulfate proteoglycan