Analysis of brain and cerebrospinal fluid from mouse models of the three major forms of neuronal ceroid lipofuscinosis reveals changes in the lysosomal proteome.

David E. Sleat‡§†, Jennifer A. Wiseman‡, Mukarram El-Banna‡, Haiyan Zheng‡, Caifeng Zhao‡, Amenah Soherwardy‡, Dirk. F. Moore¶ and Peter Lobel‡§ ǁ

‡ Center for Advanced Biotechnology and Medicine, Piscataway, NJ 08854, USA. § Department of Biochemistry and Molecular Biology, Robert-Wood Johnson Medical School, Rutgers Biomedical Health Sciences, Piscataway, NJ 08854, USA. ¶ Department of Biostatistics, School of Public Health, Rutgers - The State University of New Jersey, Piscataway, NJ 08854.

Running title: Biomarker discovery in NCL mouse models

† To whom correspondence may be addressed: Center for Advanced Biotechnology and Medicine, 679 Hoes Lane, Piscataway, NJ 08854. Tel., 848-445-9832; FAX, 732-235-4466; E-mail: sleat@cabm.rutgers.edu.

ǁ To whom correspondence may be addressed: Center for Advanced Biotechnology and Medicine, 679 Hoes Lane, Piscataway, NJ 08854. Tel., 848-445-9831; FAX, 732-235-5083; E-mail: lobel@cabm.rutgers.edu.

Abbreviations: NCL, neuronal ceroid lipofuscinoses; CSF, cerebrospinal fluid;

Keywords: neuronal ceroid lipofuscinosis; mouse model; brain; cerebrospinal fluid; proteomic; isobaric labeling
ABSTRACT.

Treatments are emerging for the neuronal ceroid lipofuscinoses (NCLs), a group of similar but genetically distinct lysosomal storage diseases. Clinical ratings scales measure long-term disease progression and response to treatment but clinically useful biomarkers have yet to be identified in these diseases. We have conducted proteomic analyses of brain and cerebrospinal fluid (CSF) from mouse models of the most frequently diagnosed NCL diseases: CLN1 (infantile NCL), CLN2 (classical late infantile NCL) and CLN3 (juvenile NCL). Samples were obtained at different stages of disease progression and proteins quantified using isobaric labeling. In total, 8,303 and 4,905 proteins were identified from brain and CSF, respectively. We also conducted label-free analyses of brain proteins that contained the mannose 6-phosphate lysosomal targeting modification. In general, we detect few changes at presymptomatic timepoints but later in disease, we detect a number of proteins whose expression is significantly altered in both brain and CSF of CLN1 and CLN2 animals. Many of these proteins are lysosomal in origin or are markers of neuroinflammation, potentially providing clues to underlying pathogenesis and providing promising candidates for further validation.
INTRODUCTION

The neuronal ceroid lipofuscinoses (NCLs) are a group of clinically-related but genetically distinct lysosomal diseases that typically affect children, resulting in premature death (1). Clinical trials have been conducted or are planned for the three most frequently encountered forms of NCL, infantile (INCL, CLN1 disease), late-infantile (LINCL, CLN2 disease) and juvenile (JNCL, CLN3 disease), referred to here as CLN1, CLN2 and CLN3, respectively. In addition, enzyme replacement therapy has been approved for CLN2 (2). In evaluating potential treatments for NCLs, clinical rating systems (3, 4) are essential for evaluating long-term efficacy but biomarkers that allow measurement of short-term response to treatment would be extremely useful. Such biomarkers could help optimize dose and dosing regimen and also provide clinical endpoint surrogates to accelerate approval of effective treatments and allow trials of ineffective treatments to be terminated in a timely manner.

The goal of this study was to extend earlier studies to identify proteins that are up- or down-regulated in CLN1, CLN2 and CLN3 that could serve as biomarkers for disease progression. Previously, we analyzed NCL cases using human autopsy brain samples and matching CSF samples from the same individuals (5). This study resulted in an extensive database of protein expression changes in these three NCL diseases but there are caveats in the interpretation of this data. First, because the NCLs are predominantly pediatric in nature, it was not possible to obtain age-matched brain and CSF samples for controls thus age-related changes could not be discounted. Second, because these were autopsy samples, there were concerns regarding sample collection, specifically with autolysis times ranging from 3-54 hours and variable amounts of blood contamination of CSF. Third, samples could only be collected at the conclusion of the disease process thus longitudinal data or response to treatment could not be evaluated.

There are various biological samples in which changes in protein expression could be detectable. Given the extensive neuropathology and changes in cell-type associated with NCL disease progression, protein changes are detectable in brain but this is not clinically-accessible
thus we have considered biological fluids that might reflect changes in protein expression in brain. Blood-based tests would have widest clinical applicability but given the complexity of this proteome, detecting changes in brain-derived proteins may prove to be difficult. We have therefore focused on CSF as a biological fluid that is proximal to the brain which is also relatively accessible, especially in LINCL patients undergoing CSF-mediated enzyme replacement therapy (2). A proportion of newly synthesized lysosomal proteins escape the lysosomal targeting pathway and are secreted, providing a route by which these intracellular proteins may be detected in the CSF. However, it is also possible that neuronal death results in leakage of intracellular proteins into the CSF.

The use of mouse models should provide complementary data to that obtained with human autopsy samples while circumventing many of the problems encountered with the latter. In this study, we have conducted a proteomic analysis of brain and CSF from mouse models of CLN1, CLN2 and CLN3. These analyses have provided an extensive catalog of proteomic changes in the NCL mouse models, extending and corroborating findings from earlier studies and providing a discovery basis for future studies to identify clinically useful biomarkers.
EXPERIMENTAL PROCEDURES

Animals - Mice were maintained and used following protocols approved by the Rutgers University and Robert Wood Johnson Medical School Institutional Animal Care and Use Committee (“Preclinical evaluation of therapy in an animal model for LINCL,” protocol I09-0274-4). All mouse models were in a C57BL/6 background. $Tpp1^{-/}$ (CLN2) mice were as described previously (6, 7) and were analyzed at 4 and ~19 weeks of age. $Ppt1^{-/-}$ (CLN1) mice (8) were obtained from Dr. Mark Sands (Washington University School of Medicine, St Louis) and were analyzed at 4 and ~26 weeks of age. $Cln3^{-/-}$ (CLN3) mice (9) were obtained from David Pearce (Sanford Research, Sioux Falls) and were analyzed at 4 and 52 weeks of age. Wild-type animals were obtained from the Jackson Laboratory and were analyzed at 4 weeks (early stage control for all NCLs), 19-25 weeks (late-stage control for CLN1 and CLN2) and 44 weeks (late-stage control for CLN3) of age. All groups contained ~equal numbers of males and females. CSF was collected from the cisterna magna (10) after deep anesthesia induced by intraperitoneal administration of 0.1-0.2 ml of mixture of sodium pentobarbital and sodium phenytoin (Euthasol, Delmarva Laboratories, Milton, DE) and animals were euthanized by exsanguination following transcardiac perfusion with saline. CSF was visually inspected for signs of blood contamination (red or pink color) and if uncontaminated, centrifuged and frozen on dry ice. Brains were dissected and frozen on dry ice. Animal details are given in Table S1. Pathology of disease has been well characterized in CLN1, CLN2 and CLN3 mouse models (reviewed in (11, 12)).

Experimental design and Statistical Rationale – This study consists of several analyses of three NCL mouse models and wild-type controls at early and late-stage disease. Proteomic analysis of brain samples was using isobaric labeling with n=3 animals per experimental group, chosen to facilitate analysis of a total of 27 animals in three labeling experiments using a pool bridge sample. Samples were analyzed using reporter ion measurement at both MS2 and MS3 level. For CSF, we determined point estimates using pooled samples from 10 animals per experimental group due to limitations of protein availability. Mannose 6-phosphate (M6P)
glycoprotein analysis was conducted using 5-6 animals per experimental group. Statistical approaches employed for each study are described in respective sections detail below.

**M6P glycoproteins** - Frozen brains were powdered using a Bessman tissue disruptor and stored at -80 °C prior to use. M6P glycoproteins were visualized in brain extracts using radioiodinated bovine soluble cation-independent M6P receptor in a western blot-style assay (13, 14). Proteins containing M6P were isolated from soluble brain extracts prepared using PBS/0.1% Triton X100 (CLN1) or PBS/0.2% Tween-20 (CLN2 and CLN3) by affinity chromatography on immobilized bovine soluble cation-independent M6P receptor as described previously (15).

**Mass spectrometry** – The protein concentration of purified M6P glycoprotein samples was determined and 5 μg per animal were used for in-gel trypsin (specificity, carboxyl side of K and R) digests (16) and analyzed in duplicate by LC-MS/MS (~0.5 μg/run) using an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher) as described (17). Five or 6 animals per genotype and time-point were analyzed for CLN1/3 and CLN2, respectively.

Whole brain extracts were prepared by homogenization of frozen brain in PBS/0.2% Tween (1:10 w/v), and cleared supernatants prepared by centrifugation at 14,000 x g at 4°C for 30 min. Protein concentrations were determined Bradford protein assay (18) (BioRad) and tryptic digests were generated by sequential trypsin and endoprotease LysC (specificity, carboxyl side of K) digestion using the FASP method (19). CSF samples (2.5 μl) from 10-15 animals per experimental group were pooled, protein concentration determined and 10 μg of each pool digested with trypsin in-gel (16). TMT10-plex (Thermo Fisher) labeling of in-gel tryptic digests was conducted using manufacturer’s methods. Pooled brain or CSF samples were fractionated by alkaline RP-HPLC and ~ 20 fractions analyzed by nanospray LC-MS using column and gradient conditions as described (17). Labeled CSF samples were analyzed using a Thermo Q Exactive HF. Each cycle consisted of an MS1 scan (300-2000 m/z, 120 K resolution) followed by data-dependent MS/MS on the 20 most intense peaks (normalized collision energy of 38%, 110-2000 m/z, 30 K resolution). Labeled brain samples were analyzed using the Q Exactive HF as described above and on a
Thermo Fusion Lumos Tribrid instrument using multinotch MS3 (20). One full MS scan was acquired in the Orbitrap (380-1600 m/z, 120K resolution). For MS2, parent ions were isolated in the quadrupole (1.3 m/z isolation width), fragmented and then scanned in the ion trap (collision energy of 35%, 400-1200 m/z). For MS3, parent ions were isolated in the quadrupole, and five MS2 fragment ions generated and isolated in the ion trap were fragmented by HCD (collision energy of 65%) and scanned in the Orbitrap detector (100-500 m/z, 50K resolution).

**Generation of peak lists** - Peak lists were generated using Proteome Discoverer version 1.4 (Orbitrap Velos) or 2.2 (Q-Exactive HF and Lumos) and 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, respectively.

**Database search** - A local implementation of the Global Proteome Machine version 2.2.1 (21, 22) (GPM) Cyclone XE (Beavis Informatics Ltd., Winnipeg, Canada) with X!Tandem version ALANINE (2017.02.01) was used to search mass spectrometry data against a unique mouse genome assembly based on GRCh38.79 (Jan 2012; 41,214 total genes, 22592 protein coding genes) and a custom database containing laboratory contaminants and other non-murine proteins. GPM peptide and proteins assignments are included in MASSIVE repository submissions (see below).

The following parameters were used to search data acquired using the Orbitrap Velos: fragment mass error, 0.4 Da; parent mass error, 10 ppm; maximum charge, +4; minimum 15 peaks assigned; one missed cleavage allowed; cysteine carbamidomethylation was a constant modification and methionine oxidation was a variable modification throughout the search; tryptophan oxidation, asparagine deamidation, and glutamine deamidation were variable modifications during model refinement. Peptide false positive rate (FPR) generated as GPM output was 0.65%. Models are presented using peptide log(e) of -1.

The following parameters were used to search data acquired using the QExactive HF: fragment mass error, 20 ppm; parent mass error, 10 ppm; maximum charge, +4; minimum 15 peaks
assigned; one missed cleavage allowed; cysteine carbamidomethylation; TMT10 at lysine and N-terminus were constant modifications; methionine oxidation was a variable modification throughout the search. Tryptophan oxidation, asparagine deamidation, glutamine deamidation, TMT10 at tyrosine and minus TMT10 at lysine and N-terminus were variable modifications during model refinement. MS data acquired using the Lumos was analyzed using identical parameters except the MS2 fragment mass error was 0.4 Da. Models are presented using peptide log(e) of -1. Peptide FPR generated as GPM output was 0.08% and 0.69% for brain and CSF samples, respectively, analyzed on the QExactive and 1.02% for brain samples analyzed on the Lumos. Protein identification data for all experiments including percentage coverage and number of unique and total peptides are shown in Tables S2-5.

Data normalization and filtering - Where possible, protein accession numbers were mapped to associated gene names using Ensembl Biomart (23) lookup tables: if no corresponding gene name could be identified, the Ensembl mouse protein identifier (ENSMUSP) was retained. Isobaric reporter ion intensities were extracted and corrected for isotopic purity with custom in-house scripts (https://github.com/cgermain/IDEAA) using vendor-supplied channel cross-over data, then merged with the X!Tandem output. Data for protein quantitation were obtained by filtering the raw data in the following sequential steps: 1) Gene products were filtered for protein assignments with at least two unique assigned peptides; 2) Data were normalized to total reporter ion intensity per channel to account for differences in protein quantity or efficiency of labeling; 3) Spectra were filtered for a minimum acceptable total label intensity (10,000 for Lumos, 20,000 for QE-HF); 4) Peptide-spectrum matches were filtered for fully tryptic cleavage with no missed sites, complete isobaric labeling of lysines and amino termini, and a lack of posttranslational modifications deemed to increase variability (asparagine or glutamine deamidation, methionine dioxidation, tryptophan mono- and dioxidation, and isobaric labeling of tyrosine at positions other than the amino terminus).

Analysis of brain samples from 27 individual animals using TMT-10plex was conducted in three separate experiments (details are provided in Table S1). Each study was comprised of 9 animals,
reserving the final channel for a pooled bridging sample derived from all 27 individual samples. Thus, after normalization to the total reporter ion signal per channel for each experiment (Step 2 above), spectra were filtered for a minimum intensity of 0.05 in the normalized reference channel and then normalized to the bridging sample for each experiment.

Statistics - Proteins analyzed were filtered for a minimum average spectral count of 5 per biological replicate. Spectral count data was analyzed using an in-house R program that modifies the package Quasiseq (24) for analysis of mass spectrometry data. Instead of independently estimating a variance term for each protein as before (5) this method (QuasiSpectral version 0.22, https://github.com/mooredf22/quasispectral) uses a Bayesian procedure to shrink these variance term estimates towards the common mean. This improves the estimate of the variance term by drawing on information from the estimates of all of the proteins. Output is provided in terms of false-discovery based q-values calculated using the original method of Benjamini and Hochberg (25).

Relative protein expression in NCL brain and CSF compared to control animals was measured by isobaric labeling. Data were filtered as described above and the mean normalized intensity of all spectra assigned to a given protein was used to obtain a point estimate for the relative abundance of that protein in each individual animal. Statistical analyses were conducted in GraphPad Prism Version 8 with multiple t-tests using the two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli (26) to calculate significance using indicated false discovery rates as thresholds. Note that missing values were present when a given protein was not found in all three experiments, thus t-tests could only be conducted when spectra corresponding to given proteins were identified in at least two of both control and NCL samples. Statistical analyses were conducted as described for brain except normalized data from all spectra assigned to each protein were used to calculate significance of point estimates.

Deposition of MS data - Raw files, mgf files, GPM result files, Excel workbooks listing protein and peptide assignments, and keys for data interpretation are archived in the MassIVE
Biomarker discovery in NCL mouse models

(http://massive.ucsd.edu) and ProteomeXchange (http://www.proteomexchange.org/) repositories: MSV000083797 (analysis of NCL mouse brain extracts using the QExactive HF), MSV000083827 (analysis of NCL mouse brain extracts using the Lumos), MSV000083619 (analysis of purified mouse brain M6P glycoproteins on the QExactive HF) and MSV000083828 (analysis of mouse CSF on the QExactive HF).

**Human data** – A goal of this study was to cross-validate data obtained with mouse NCL models with data from human patients with the rationale that similar changes in both species would add confidence to conclusions. Data for isobaric label analysis of human brain and CSF was published previously (5) but new statistical analyses were conducted as described above.
RESULTS

Quantitative proteomic analysis of mouse brain - To identify differentially expressed proteins, we conducted untargeted analysis of total mouse brain extracts from NCL models and controls. While brain is not an accessible sample, our rationale was that changes in brain-derived proteins might also be detected in CSF or plasma. NCL animals and controls were analyzed at a presymptomatic early time-point (4 weeks) and at late-stage disease (CLN1, 26 weeks, CLN2, 19 weeks and CLN3, 52 weeks). Phenotype of the CLN3 mice is subtle but animals at 52 weeks age are reported to have behavioral deficits and detectable accumulation of storage material (27). Protein quantitation was conducted by both MS2 and MS3 measurement of reporter ions using Thermo Scientific QExactive HF or Lumos instruments, respectively (Table S6) and protein identification data are summarized in Table 1. MS3 analysis using the Lumos provided ~15% more confident protein assignments (see EXPERIMENTAL PROCEDURES) than the MS2 analysis on the QExactive HF, with ~11% more quantifiable proteins.

We initially compared wild-type groups to determine whether these could be combined to increase statistical power. Preliminary analyses (Fig. S1) indicated significant changes when data obtained using MS3 measurement of reporter ions from 4-week animals were compared with the 19-25 and 44 week groups. However, there were no significant differences between the two older groups of wild-type animals therefore these were combined when analyzing data from the late-stage NCL animals. Thus, for the early time-point, we compared 4-week NCL animals with age matched wild-type controls (n=3 per group). No significant changes were detected for any NCL (Fig. S2). For the late-timepoints, we compared NCL animals (n=3) with the combined wt group of 16-25 and 44 week animals (n=6).

Disease-associated changes in protein expression are visualized as volcano plots in Figs. 1 and 2 with measurement of isobaric reporter ions measured at the MS2 and MS3 level, respectively. When reporter ion quantitation methods are compared, we find more significant changes and greater effect sizes with MS3 analysis, consistent with a reduction in the
background signal and subsequent ratio compression associated with measurement of MS2 reporter ions (28). We compared fold-changes obtained by MS2 and MS3 reporter ion measurements for proteins in late stage CLN1 compared to wild-type controls given that the most dramatic changes in protein expression were found in this group. Significant changes detected by MS2 were a subset of those detected by MS3 but data were generally well correlated for proteins that were found to be significantly altered with, as expected, the MS3 approach having a greater dynamic range (Fig. 3). Significant changes detected in brain using MS3 reporter ion measurement in late-stage disease for the three NCLs are discussed below.

Numerous (445) significant changes in protein expression were detected in CLN1 brain at late-stage disease. Most proteins were elevated although palmitoyl protein thioesterase 1 (PPT1) was decreased as expected - mutations in the gene encoding this lysosomal protein underlie CLN1 (29) and the Ppt1 gene is targeted in the CLN1 mouse model (8).

Fewer proteins (91) were significantly altered in late-stage CLN2 although tripeptidyl peptidase 1 (TPP1) was decreased as expected - mutations in the gene encoding this lysosomal protease underlie CLN2 (30) and the Tpp1 gene is targeted in the CLN2 mouse model (6). In CLN3, only 4 proteins were significantly elevated. In Fig. 4, we determined the subcellular distribution of proteins found to be significantly elevated in each NCL disease using the predominant compartment assigned to each protein in Expt A of the PROLOCATE database (http://prolocate.cabm.rutgers.edu/index) of rat liver protein subcellular localizations (31). When all proteins are considered, those that are assigned to the lysosome are a relatively minor fraction of the total assigned proteins (~5%). However, when considering proteins elevated in NCL, lysosomal proteins are over represented in CLN1 (~17% lysosomal, p=1.3 x 10^{-15}), CLN2 (~31%, p=1.3 x 10^{-11}) and CLN3 (75%, p=5.6 x 10^{-4}). Lysosomal storage diseases are frequently associated with multiple changes in lysosomal enzyme activities that are secondary to the primary defect. The cause of such changes are generally unclear but they may be compensatory cellular responses to the accumulation of multiple substrates within the lysosome. There are
examples of such secondary alterations in the NCLs. For example, PSAP is a major component of the storage material in INCL (32), β-glucuronidase (GUSB) is elevated in LINCL brain (33), and TPP1 (34), lysosomal acid phosphatase (35) and other lysosomal enzymes (33) are reported to be elevated in JNCL brain. Given that correction of the primary lysosomal defect by enzyme replacement or gene therapy is likely to diminish such responses, secondary lysosomal changes may represent useful biomarkers. In CLN1, CD63 protein, TPP1, CTSD and PSAP were elevated ~4-fold or higher and 14 other lysosomal proteins were significantly elevated more than 2-fold including scavenger receptor class B member 2 (SCARB2) and cathepsins S and Z (CTSS and CTSZ). In CLN2, lysosomal proteins CTSZ, CTSA, beta-hexosaminidase subunit beta (HEXB), serine carboxypeptidase 1 (SCPEP1) and SCARB2 were modestly (< 2.2-fold) elevated. In CLN3, only 4 proteins were significantly elevated and these included lysosomal proteins SCARB2, HEXB and TPP1.

Nicotinamide nucleotide transhydrogenase (NNT) was elevated >4-fold in CLN2. The latter is not a reflection of TPP1 deficiency - investigation subsequent to this analysis revealed that elevated NNT results from a difference in substrain between the CLN2 mice and wild-type controls. While both are in C57BL6 genetic background, the controls were of the C57BL6/J substrain which has reduced NNT levels due to mutations within the Nnt gene (36) but we anticipate that the effect of NNT on CLN2 phenotype should be minimal. However, it is worth noting that the detection of elevated NNT in the CLN2 mouse, as well as decreased PPT1 and TPP1 in the CLN2 and CLN3 mice respectively, help validate the quantitative mass spectrometry approach. Comparison of TMT-MS3 data obtained in this study for specific proteins that are reported previously to be altered in NCL samples are shown in Table S7.

Table 2 lists 71 proteins that are significantly elevated in both CLN1 and CLN2 and two striking observations can be made from this comparison. As noted above, a significant proportion of these proteins are of lysosomal origin. However, a striking observation was that 32 out of 71 proteins significantly elevated in both CLN1 and CLN2 brain play roles in immune system function.
including inflammatory response and glial activation. Alterations in expression of proteins involved in inflammatory response have been implicated in CLN1, 2 and 3 (37-40).

In our earlier study (5), we conducted quantitative mass spectrometry analysis of extracts from brain autopsy samples from patients with NCL diseases and we compared data obtained from animal models with NCL patient autopsy samples. While there will be species-dependent differences and post-mortem sample collection will likely influence data obtained from human samples, our rationale is that conservation of proteomic changes between human and mouse in a disease-specific manner will provide additional confidence in potential biomarker candidates. For both CLN1 and CLN2, we detect good correlation between changes in protein expression in human and mouse brain (Fig. 5). Correlations between human and mouse are strongest when lysosomal proteins are considered alone ($r^2=0.5957$ and 0.2226 for CLN1 and CLN2, respectively).

Visualizing M6P glycoproteins in mouse NCL models - Given the changes in expression of lysosomal proteins detected by data-dependent analysis of whole brain extracts in the NCL mouse models, we conducted a focused analysis of this group of proteins. M6P is a carbohydrate modification that is primarily associated with newly-synthesized lumenal lysosomal proteins (also referred to as soluble or matrix lysosomal proteins) that is recognized by two receptors that direct intracellular targeting from the Golgi to an acidified, prelysosomal compartment (41). Note that M6P is not found on lysosomal membrane proteins. While this modification is rapidly removed in most cell types by acid phosphatase 5 (42), it is retained in neurons (13, 43). Purified M6P receptors (MPRs) can be used for affinity isolation (44, 45) and visualization (14) of lysosomal proteins. In human NCL cases (5, 46), we found characteristic changes in the expression of M6P glycoproteins in brain samples, presumably reflecting lysosomal perturbation.

A radiolabeled cation-independent M6P receptor derivative was used to broadly visualize M6P glycoproteins in brain extracts from NCL mice and controls (Fig. 6) to determine whether there were any gross changes in the expression of lumenal lysosomal proteins. In CLN1, there was a
generalized increase in M6P glycoproteins at late-stage disease while a M6P glycoprotein of \(\sim 45-50\) kDa was highly elevated at both early and late stage disease. This indicates that even at the presymptomatic 4-week time-point, lysosomal changes due to PPT1 deficiency were already present in CLN1. Levels of the corresponding \(\sim 45-50\) kDa protein band were reduced in the CLN2 samples, suggesting that it likely corresponds to TPP1. As with CLN1, there was a generalized increase in M6P glycoproteins detected in late-stage CLN2. M6P glycoprotein expression in early-stage CLN3 appeared similar to wild-type while at late-stage disease, the \(\sim 45-50\) kDa band was highly elevated, as were bands corresponding to other M6P glycoproteins.

**Quantitative analysis of M6P glycoproteins** - M6P glycoproteins were purified from the brains of CLN1 (26-week old), CLN2 (19-week old) and CLN3 (52-week old) and age-matched wild-type controls (n=5-6 animals per experimental group). Extracts were applied to columns of immobilized soluble cation-independent M6P receptor (sCl-MPR) and M6P glycoproteins specifically eluted by free M6P sugar prior to analysis by data-dependent mass spectrometry. Preparations of purified M6P glycoproteins primarily comprise well-characterized luminal lysosomal proteins. However, there are other proteins not assigned to the lysosome – these can represent previously unrecognized lysosomal proteins, non-lysosomal proteins that interact with and copurify with lysosomal M6P glycoproteins, or contaminants that are not fully removed during column washing (13, 17). The lysosomal M6P glycoproteome has been extensively characterized (reviewed in (44)) and the presence of M6P directly demonstrated for most lysosomal proteins (47, 48).

Individual proteins were quantified by spectral counting. Note that equal amounts of each sample were analyzed by LC-MS/MS thus this analysis does not account for changes in overall levels of M6P glycoproteins *in vivo*, providing information about relative rather than absolute expression levels. NCL samples were compared to controls and volcano plots representing the expression of lysosomal and other proteins in each NCL at early and late stage disease are
shown in Fig. 7 and spectral count analyses are presented in Tables S8-10 (mouse) and Tables S11-13 (human).

Consistent with blotting experiments (Fig. 6), there were numerous disease-associated alterations in the relative expression of M6P glycoproteins in the late-stage (26-week) CLN1 samples. M6P containing forms of TPP1, acid lipase (LIPA), CTSF and CTSD were elevated ~2-fold (Fig. 7). A low level of peptides corresponding to PPT1 were detected in the CLN1 brain extracts, possibly reflecting expression of some inactive protein from the targeted gene. Note that a protein of unknown function that has not been assigned to the lysosome, Von Willebrand factor A domain-containing protein 5A (VWA5A) was highly elevated (~20-fold) in the M6P glycoprotein preparation from CLN1 brain. This protein was significantly elevated in CLN1 and CLN2 whole-brain extracts (Table 2) and its presence in the purified M6P glycoprotein preparation may indicate association with other M6P glycoproteins during the purification process.

In CLN2, TPP1 was not detected while cathepsins H (CTSH) and C (CTSC) were both elevated ~ 2-fold. Few significant changes were observed in CLN3 although TPP1 was elevated modestly (~1.4 fold). This is consistent with the elevated levels in CLN1 of a 45-50 kDa glycoprotein that presumably corresponds to TPP1 (Fig. 6) and earlier studies reporting an increase in TPP1 activity in juvenile neuronal ceroid lipofuscinosis (34). Sphingomyelin phosphodiesterase 1 (SMPD1) was significantly decreased in CLN3.

Expression of mouse brain M6P glycoproteins at the late-stage time-point for each NCL was compared with the human brain samples (5) and we identified a number of conserved changes between both species (Fig. 8). For CLN1, there was correlation between relative changes in lysosomal M6P glycoprotein levels between human and mouse ($r^2=0.3391$ with PPT1 censored in linear regression analysis). A number of lysosomal proteases (CTSC, CTSH, CTSS and TPP1) were elevated as well as proteins involved in lipid metabolism (LIPA and NPC2, a small cholesterol-binding protein involved in lysosomal cholesterol transport (49)). Protease inhibitor
cystatin B (CSTB), which is not lysosomal and likely copurifies with cathepsins, was also elevated. Relative expression of inhibitor of serine proteases (SERPINB8) was decreased – this protein is not lysosomal and it is not clear whether it contains M6P (a related protein, neuroserpin, has been experimentally shown to do so (47)). GDP-fucose protein O-fucosyltransferase 2 (POFUT2), a M6P glycoprotein that likely localizes to the endoplasmic reticulum (47), was also decreased in mouse and human brain M6P glycoproteins preparations. For CLN2 and CLN3, correlation between human and mouse M6P glycoprotein expression was less marked, likely reflecting the relatively modest magnitude of changes measured in both species. However, the M6P form of GUSB was relatively elevated in mouse and human brain in both diseases. SMPD1 was consistently lower in both mouse and human CLN3 samples.

In Fig. 9, we compare relative levels of lysosomal proteins compared to wild-type controls in purified M6P glycoprotein preparations and in whole brain extracts. In general, they are well correlated, indicating the levels of lysosomal proteins containing M6P reflect overall levels.

Quantitative proteomic analysis of mouse cerebrospinal fluid (CSF) - In order to identify potential biomarkers in CSF, deep proteome analysis using two-dimensional LC fractionation was required given that proteins of interest (e.g., lysosomal proteins) are extremely low-abundance components of the proteome. As typically only ~5 μl of CSF (containing <1 μg protein) could be collected from an individual animal, we used a pooling strategy to obtain sufficient material for the analysis (see EXPERIMENTAL PROCEDURES). While not affording estimates of animal to animal variability, this allowed measurement of point estimates to cross-reference against data obtained from mouse brain and human samples (5) to help corroborate possible biomarkers.

In terms of proteome coverage, this approach was successful, with 4,688 mouse CSF proteins (Table 1) assigned with 2 or more unique peptide assignments. When compared to mouse brain analyzed using MS3 (Table S14), 3,814/3,312 proteins were identified/quantified in both sources, 3,973/3,862 were identified/quantified in brain alone while 874/699 proteins were
identified/quantified in CSF alone.

Expression of CSF proteins at different stages of disease was visualized by Volcano plots (Fig. 10) and results summarized in Tables S14 and S15. Note that in this case, significance of change was calculated by multiple t-tests at the peptide level for each protein point estimate, a similar approach to that used with human samples (5). There are individual proteins that can be used to benchmark the point estimates. For example, in both early and late-stage CLN1, PPT1 was significantly reduced. In early and late stage CLN2, TPP1 was reduced and while not achieving significance, it was the most decreased lysosomal protein. NNT was increased by ~8- and 3.5-fold in early and late CLN2, respectively.

In CLN1, there were numerous changes in protein expression identified in CSF. At early stage disease, there was a generalized but modest (<2-fold) increase in levels of lysosomal proteins. In late-stage CLN1 CSF, a number of lysosomal proteins were elevated 1.6- to 2.5-fold including HEXA, CTSZ, CTSD and CTSS. Fewer changes were observed in early stage CLN2 although a number of proteins were significantly elevated by more than 2-fold, including GFAP. At late-stage CLN2, numerous proteins were elevated compared to wild-type including ~20 lysosomal proteins e.g., CTSZ, CTSS, and CTSD. In Table S15, we measured the numbers of proteins that were significantly increased or reduced in the NCL CSF samples. It is worth noting that in CLN1 and CLN2 CSF, more non-lysosomal proteins were significantly decreased than increased in the NCL samples compared to controls. In contrast, lysosomal proteins were elevated in the NCL samples (with the exception of PPT1 and TPP1 in CLN1 and CLN2, respectively). Overall, this correlates well with brain, where lysosomal proteins tended to be increased in the disease samples. In CLN3, very few changes were detected at either early or late-stage disease in CSF.

CSF Biomarker candidates - Our overall strategy was to cross-reference results obtained from brain and CSF from the mouse NCL models. Our rationale was that proteins that show significant changes in CSF that reflect similar changes in brain would potentially provide the most
informative biomarkers for brain disease. A number of promising candidates emerging from this analysis that were altered in both brain and CSF are discussed below. Proteins that are significantly altered in late-stage CLN1 and CLN2 are shown in Tables 3 and 4, respectively. No proteins were significantly altered in both CLN3 brain and CSF. In Fig. 11, NCL-associated changes in protein expression in brain and CSF are compared. Overall, we do detect some proteins with changes in expression that are well correlated between brain and CSF, e.g., SERPINA3N, LYZ2, C4B and CTSZ are elevated in both brain and CSF. However, we also detect a number of proteins that are highly-elevated in CSF (e.g., NEFH, NEFM, CHIL1 and CALB1) but unchanged in brain. It is possible that expression of these proteins is altered in a disease-specific manner in localized regions of the brain that contribute significantly to the CSF proteome (e.g., cells of the choroid plexus) but not in the brain overall. It is worth noting that the most striking changes in CSF were detected in both CLN1 and CLN2. Conservation of these changes between CLN1 and CLN2 not only suggests that they may represent non-specific) responses to neurodegeneration (or common responses to lysosomal dysfunction) but also adds confidence in their selection as potential candidates.
DISCUSSION

The goal of this study was to conduct quantitative analyses of the brain and CSF proteomes of murine NCL models to provide potential biomarker candidates for further validation. A major rationale was that analysis of genetically homogeneous NCL mouse models would reduce some of the inherent data variability encountered in using human autopsy samples. The study focused on total proteins from brain and CSF analyzed at a presymptomatic time-point and in late-stage disease while we also conducted targeted analysis of lysosomal proteins in brain. For global surveys, the use of isobaric labeling with extensive prefractionation allowed us to achieve extremely broad proteome coverage. For example, two previous extensive proteomic studies of mouse CSF (50, 51) identified 566 and 715 proteins, respectively while in this study, we confidently (see EXPERIMENTAL PROCEDURES) identified ~4,700 proteins in CSF samples, ~4,000 of which were quantifiable.

Reflecting similar observations in human brain and CSF samples (5), a wider range of proteomic changes were found in the samples derived from CLN1 mice compared to CLN2 and changes in expression tended to be of higher magnitude. In mice, this does not correlate with disease severity in terms of lifespan: median survival of the CLN2 mouse (~ 4 months (6)) is shorter than the CLN1 mouse (~7 months (8)). It is possible that pathology is more widely distributed throughout the CLN1 mouse brain compared to the CLN2 brain, accounting for the broader proteomic changes that are detected when whole brain is analyzed. In contrast, the shorter lifespan of the CLN2 mouse may be a consequence of localized but highly damaging pathology. In the human study, changes in CLN2 and CLN3 were similar. This discrepancy with the mouse likely reflects that the human samples were obtained at autopsy from severely affected patients while the CLN3 mice have highly attenuated disease. Lifespan of the CLN3 in our colony is essentially normal (data not shown) and there is no overt end-stage phenotype.

Several other studies have conducted proteomic analyses of NCL mouse models. An analysis was conducted of samples from newborn CLN1 animals of brain membrane proteins
that were enriched for acylated proteins with the rationale that palmitoylation would be increased in the absence of PPT1 (52). This study found that the acylated forms of 88 proteins were significantly altered in CLN1 compared to wild-type controls. Of these, we identified 83 in our analysis of whole brain homogenates. With the exception of TLN1 and ANXA2 which were significantly but modestly elevated in late-stage CLN1, we detected no significant changes in aggregate expression levels (i.e., both acylated and not acylated) of these proteins in this study.

In another study (53), thalamus of CLN1 mice was analyzed using label-free methods, identifying 139 proteins that were significantly altered at one or more time points (presymptomatic, symptomatic and late-stage). We identified 106 of these proteins in this study, 7 of which were significantly (FDR ≤2%) altered at late-stage CLN1, including APOE and GFAP. A detailed comparison of this dataset with the current study is shown in Table S16. In a recent study (54), proteomic analysis of lysosomal fractions from CLN3 mouse cerebellar cells revealed a number of significant lysosomal changes. We detected few significant changes associated with CLN3 in lysosomal or other proteins in whole brain extracts or M6P glycoprotein preparations from whole brain. This may reflect cell-type specific responses that may be diluted when whole brain is analyzed. However, elevated TPP1 is a noteworthy and consistent observation in CLN3 in the current and other studies (34, 54).

Several proteins were highly elevated in both CLN1 and CLN2 (Table 2). CD63, is a transmembrane protein located on the plasma membrane and intracellularly within the lysosomal-endosomal system. Elevated ~4- and 2-fold in CLN1 and CLN2 respectively, CD63 may play a role in leukocyte motility (55) and activation of mast cells and basophils (56). Glycoprotein non-metastatic melanoma B (GPNMB) was elevated 15- and 3-fold in CLN1 and CLN2, respectively and is thought to attenuate inflammatory responses in microglia and astrocytes via interaction with CD44 (57), which was also significantly elevated in CLN1 and CLN2. GPNMB is also elevated in CSF from patients with the lysosomal disorder Gaucher disease (58) and has been implicated in Parkinson’s disease. Lysozyme 2 (LYZ2) was
significantly elevated ~7-fold and ~2-fold in CLN1 and CLN2 respectively. LYZ2 expression is highly-elevated in neurons from specific regions of the brain in a model of lysosomal storage disease Sanfilippo syndrome type B (59) and cortex from mouse models of mucopolysaccharidoses I and IIIB (60). Serpin family A3 (SERPINA3), a neuroprotective inhibitor of granzyme B (61), was elevated ~11-fold and ~5-fold in CLN1 and CLN2, respectively. Other proteins elevated in CLN1 and CLN2 associated with neuroinflammation and reactive gliosis include GFAP, C4B, C1QA, C1QB, C1QC, CD44 and S100A6. Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma (PIK3CG) was significantly decreased in both CLN1 and CLN2 brain. PIK3CG has been implicated in attenuating neuroinflammatory responses in ischemia (62).

Transglutaminase 1 (TGM1) is not directly implicated in immune function but was elevated ~19- and 5-fold in CLN1 and CLN2, respectively. TGM1 mutations cause rare hereditary skin disorders in humans (63) but this protein has not previously been implicated in neurodegenerative disease. However, the related protein transglutaminase 2 (TG2) has been proposed play a role in neurodegeneration, although whether it promotes or protects against neuronal death is not clear (64).

The majority of these proteomic changes detected were only identified at late stage disease, suggesting that they were a consequence of progressive neurodegeneration or other pathological processes rather than and immediate response to the metabolic defect caused by mutation of the disease gene. To examine this further, we investigated expression of these proteins in related diseases. For example, in a microarray analysis of cortex from mouse models of mucopolysaccharidoses I and IIIB (60), which are also lysosomal storage diseases characterized by severe neurodegeneration, transcripts corresponding to various proteins were found to be elevated with 13 proteins that were increased in both MPSI and IIIB mouse models. Most of this common set are proteins that are expressed by and associated with macrophage function. Nine proteins of the MPSI/IIIB common set were detected in the NCL models (Table
5). Remarkably, nearly all were significantly elevated in CLN1 and CLN2 brain at late stage disease and most were significantly elevated in CLN1 and CLN2 CSF (Table 3). There was less concordance between the NCLs and proteins encoded by transcripts that were elevated in MPSI alone although there was some agreement – PSMB8 and SERPINA3N were both elevated in late-stage CLN1 and CLN2 brain while LAG3 and SERPING1 were elevated in late-stage CLN2 CSF. None of the MPSI/IIIIB common set were significantly elevated in CLN3.

Another transcriptomic study has examined gene expression in the fore-midbrain regions of the CLN2 mouse model at 12 weeks (65) and in Figure 12, we compare protein expression changes measured by transcriptomic and proteomic analyses for the intersecting dataset of 437 proteins. Overall, correlation was not high, reflecting differences frequently seen when measuring protein expression at the protein and transcript level, although it is interesting to note that changes in cerebellum more closely correlated with proteomic measurements in whole brain ($r^2=0.140$) than changes in mid/forebrain ($r^2=0.031$). More importantly, there is extremely good concordance between transcriptomic and proteomic measurements in brain of several proteins that were markedly altered in the current proteomic study including LYZ2, TGM1, GPNMB, C4B, SERPINA3N, GFAP and C1QC.

The overall goal of the study was to identify proteomic changes in brain that are reflected by parallel responses in CSF and a number of promising candidates have emerged from this study including CTSZ, SERPINA3N, LYZ2, TGM1 and C4B. In addition, we detected changes in expression of neurofilaments NEFH, NEFL and NEFM which may also be promising candidates as biomarkers for NCL diseases given that elevated levels of these proteins have been well established in CSF and other samples from patients with various other neurodegenerative diseases (66). Many of these protein changes were found in both CLN1 and CLN2 and other LSDs suggesting that they may reflect neurodegeneration in general but there were some changes that may be disease–specific. For example, envoplakin (EPVL), which is suggested to interact with TGM1, is highly elevated in CLN2 CSF. The origin of these potential biomarkers in
CSF remains unclear but one possibility is that these intracellular proteins are released into the CSF as a consequence of neurodegeneration. If this were the case, we would predict that abundant intracellular proteins would be elevated in CSF in an NCL-dependent manner. We investigated this possibility by specifically examining levels of abundant intracellular proteins in NCL CSF. In Fig. S3, we compared protein abundance in brain extracts as determined by GPM log intensity scores (i.e., log base 10 of the sum of intensities of fragment ions, Table S2), with the corresponding ratio between NCL samples and control in CSF. There is no clear correlation between abundance in brain and elevated levels in NCL CSF samples, suggesting that biomarkers of interest in CSF do not appear to be present simply due to generalized release of contents of brain cells due to cell death.

Finally, it should be borne in mind that while CSF biomarkers for NCL diseases could potentially be extremely useful, responsive biomarkers may not actually exist. Thus, subsequent studies are needed to validate candidates identified in this study and use of mouse models will facilitate long-term studies to evaluate the response of biomarkers to disease progression. In addition, therapeutic approaches have now been developed that extend the lifespan of the CLN1 and CLN2 (67-71) mouse models thus the response of potential biomarkers to effective treatments could also be evaluated.

Acknowledgements - This project was supported by a grant from the Batten Disease Support and Research Association, Beyond Batten Disease Foundation and Hope 4 Bridget (D.E.S.) and NIH grants R21NS088786 (D.E.S.), R01NS37918, S10RR024584 and S10OD016400 (P.L.), and P30CA072720 (D.F.M.). Mass spectrometry was conducted by the Biological Mass Spectrometry Facility of Robert Wood Johnson Medical School and Rutgers University which is supported in part by NIH P30NS046593.
REFERENCES
1. Mole, S. E., Williams, R. E., and Goebel, H. H. (2011) The neuronal ceroid lipofuscinoses (Batten disease), 2nd Ed., Oxford University Press, Oxford
2. Schulz, A., Ajayi, T., Specchio, N., de Los Reyes, E., Gissen, P., Ballon, D., Dyke, J. P., Cahan, H., Slasor, P., Jacoby, D., Kohlschutter, A., and Group, C. L. N. S. (2018) Study of Intraventricular Cerliponase Alfa for CLN2 Disease. *N Engl J Med* 378, 1898-1907
3. Steinfeld, R., Heim, P., von Gregory, H., Meyer, K., Ullrich, K., Goebel, H. H., and Kohlschutter, A. (2002) Late infantile neuronal ceroid lipofuscinosis: quantitative description of the clinical course in patients with CLN2 mutations. *Am J Med Genet* 112, 347-354
4. Marshall, F. J., de Blieck, E. A., Mink, J. W., Dure, L., Adams, H., Messing, S., Rothberg, P. G., Levy, E., McDonough, T., DeYoung, J., Wang, M., Ramirez-Montealegre, D., Kwon, J. M., and Pearce, D. A. (2005) A clinical rating scale for Batten disease: reliable and relevant for clinical trials. *Neurology* 65, 275-279
5. Sleat, D. E., Tannous, A., Sohar, I., Wiseman, J. A., Zheng, H., Qian, M., Zhao, C., Xin, W., Barone, R., Sims, K. B., Moore, D. F., and Lobel, P. (2017) Proteomic Analysis of Brain and Cerebrospinal Fluid from the Three Major Forms of Neuronal Cereoid Lipofuscinosis Reveals Potential Biomarkers. *J Proteome Res* 16, 3787-3804
6. Sleat, D. E., Wiseman, J. A., El-Banna, M., Kim, K. H., Mao, Q., Price, S., Macauley, S. L., Sidman, R. L., Shen, M. M., Zhao, Q., Passini, M. A., Davidson, B. L., Stewart, G. R., and Lobel, P. (2004) A mouse model of classical late-infantile neuronal ceroid lipofuscinosis based on targeted disruption of the CLN2 gene results in a loss of tripeptidyl-peptidase I activity and progressive neurodegeneration. *J Neurosci* 24, 9117-9126
7. Sleat, D. E., El-Banna, M., Sohar, I., Kim, K. H., Dobrenis, K., Walkley, S. U., and Lobel, P. (2008) Residual levels of tripeptidyl-peptidase I activity dramatically ameliorate disease in late-infantile neuronal ceroid lipofuscinosis. *Mol Genet Metab* 94, 222-233

8. Gupta, P., Soyombo, A. A., Atashband, A., Wisniewski, K. E., Shelton, J. M., Richardson, J. A., Hammer, R. E., and Hofmann, S. L. (2001) Disruption of PPT1 or PPT2 causes neuronal ceroid lipofuscinosis in knockout mice. *Proc Natl Acad Sci U S A* 98, 13566-13571

9. Greene, N. D., Bernard, D. L., Taschner, P. E., Lake, B. D., de Vos, N., Breuning, M. H., Gardiner, R. M., Mole, S. E., Nussbaum, R. L., and Mitchison, H. M. (1999) A murine model for juvenile NCL: gene targeting of mouse Cln3. *Mol Genet Metab* 66, 309-313

10. Lim, N. K., Moestrup, V., Zhang, X., Wang, W. A., Moller, A., and Huang, F. D. (2018) An Improved Method for Collection of Cerebrospinal Fluid from Anesthetized Mice. *J Vis Exp*

11. Shacka, J. J. (2012) Mouse models of neuronal ceroid lipofuscinoses: useful pre-clinical tools to delineate disease pathophysiology and validate therapeutics. *Brain Res Bull* 88, 43-57

12. Cooper, J. D., Russell, C., and Mitchison, H. M. (2006) Progress towards understanding disease mechanisms in small vertebrate models of neuronal ceroid lipofuscinosis. *Biochimica et biophysica acta* 1762, 873-889

13. Sleat, D. E., Sohar, I., Lackland, H., Majercak, J., and Lobel, P. (1996) Rat brain contains high levels of mannose-6-phosphorylated glycoproteins including lysosomal enzymes and palmitoyl-protein thioesterase, an enzyme implicated in infantile neuronal lipofuscinosis. *J Biol Chem* 271, 19191-19198

14. Valenzano, K. J., Kallay, L. M., and Lobel, P. (1993) An assay to detect glycoproteins that contain mannose 6-phosphate. *Anal Biochem* 209, 156-162

15. Sleat, D. E., Della Valle, M. C., Zheng, H., Moore, D. F., and Lobel, P. (2008) The mannose 6-phosphate glycoprotein proteome. *J Proteome Res* 7, 3010-3021
16. Sleat, D. E., Lackland, H., Wang, Y., Sohar, I., Xiao, G., Li, H., and Lobel, P. (2005) The human brain mannose 6-phosphate glycoproteome: a complex mixture composed of multiple isoforms of many soluble lysosomal proteins. *Proteomics* 5, 1520-1532

17. Sleat, D. E., Sun, P., Wiseman, J. A., Huang, L., El-Banna, M., Zheng, H., Moore, D. F., and Lobel, P. (2013) Extending the mannose 6-phosphate glycoproteome by high resolution/accuracy mass spectrometry analysis of control and acid phosphatase 5-deficient mice. *Molecular & cellular proteomics : MCP* 12, 1806-1817

18. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72, 248-254

19. Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat Methods* 6, 359-362

20. McAlister, G. C., Nusinow, D. P., Jedrychowski, M. P., Wuhr, M., Huttlin, E. L., Erickson, B. K., Rad, R., Haas, W., and Gygi, S. P. (2014) MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal Chem* 86, 7150-7158

21. Craig, R., Cortens, J. P., and Beavis, R. C. (2004) Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res* 3, 1234-1242

22. Beavis, R. C. (2006) Using the global proteome machine for protein identification. *Methods Mol Biol* 328, 217-228

23. Kinsella, R. J., Kahari, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P., and Flicek, P. (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* 2011, bar030

24. Lund, S. P., Nettleton, D., McCarthy, D. J., and Smyth, G. K. (2012) Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat Appl Genet Mol Biol* 11
25. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001) Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 125, 279-284

26. Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006) Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93, 491-507

27. Katz, M. L., Johnson, G. S., Tullis, G. E., and Lei, B. (2008) Phenotypic characterization of a mouse model of juvenile neuronal ceroid lipofuscinosis. *Neurobiol Dis* 29, 242-253

28. Ting, L., Rad, R., Gygi, S. P., and Haas, W. (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods* 8, 937-940

29. Vesa, J., Hellsten, E., Verkruyse, L. A., Camp, L. A., Rapola, J., Santavuori, P., Hofmann, S. L., and Peltonen, L. (1995) Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature* 376, 584-587

30. Sleat, D. E., Donnelly, R. J., Lackland, H., Liu, C. G., Sohar, I., Pullarkat, R. K., and Lobel, P. (1997) Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis. *Science* 277, 1802-1805

31. Jadot, M., Boonen, M., Thirion, J., Wang, N., Xing, J., Zhao, C., Tannous, A., Qian, M., Zheng, H., Everett, J. K., Moore, D. F., Sleat, D. E., and Lobel, P. (2016) Accounting for protein subcellular localization: A compartmental map of the rat liver proteome. *Molecular & cellular proteomics : MCP*

32. Tyynela, J., Palmer, D. N., Baumann, M., and Haltia, M. (1993) Storage of saposins A and D in infantile neuronal ceroid-lipofuscinosis. *FEBS letters* 330, 8-12

33. Sleat, D. E., Sohar, I., Pullarkat, P. S., Lobel, P., and Pullarkat, R. K. (1998) Specific alterations in levels of mannose 6-phosphorylated glycoproteins in different neuronal ceroid lipofuscinoses. *Biochem J* 334 ( Pt 3), 547-551

34. Mitchison, H. M., Bernard, D. J., Greene, N. D., Cooper, J. D., Junaid, M. A., Pullarkat, R. K., de Vos, N., Breuning, M. H., Owens, J. W., Mobley, W. C., Gardiner, R. M., Lake, B. D., Taschner, P. E., and Nussbaum, R. L. (1999) Targeted disruption of the Cln3 gene provides a
mouse model for Batten disease. The Batten Mouse Model Consortium [corrected].

_Neurobiology of disease_ 6, 321-334

35. Pohl, S., Mitchison, H. M., Kohlschutter, A., van Diggelen, O., Braulke, T., and Storch, S. (2007) Increased expression of lysosomal acid phosphatase in CLN3-defective cells and mouse brain tissue. _J Neurochem_ 103, 2177-2188

36. Freeman, H. C., Hugill, A., Dear, N. T., Ashcroft, F. M., and Cox, R. D. (2006) Deletion of nicotinamide nucleotide transhydrogenase: a new quantitative trait locus accounting for glucose intolerance in C57BL/6J mice. _Diabetes_ 55, 2153-2156

37. Qiao, X., Lu, J. Y., and Hofmann, S. L. (2007) Gene expression profiling in a mouse model of infantile neuronal ceroid lipofuscinosis reveals upregulation of immediate early genes and mediators of the inflammatory response. _BMC Neurosci_ 8, 95

38. Xiong, J., and Kielian, T. (2013) Microglia in juvenile neuronal ceroid lipofuscinosis are primed toward a pro-inflammatory phenotype. _J Neurochem_ 127, 245-258

39. Jalanko, A., Vesa, J., Manninen, T., von Schantz, C., Minye, H., Fabritius, A. L., Salonen, T., Rapola, J., Gentile, M., Kopra, O., and Peltonen, L. (2005) Mice with Ppt1Deltaex4 mutation replicate the INCL phenotype and show an inflammation-associated loss of interneurons. _Neurobiol Dis_ 18, 226-241

40. Vuillemenot, B. R., Kennedy, D., Cooper, J. D., Wong, A. M., Sri, S., Doeleman, T., Katz, M. L., Coates, J. R., Johnson, G. C., Reed, R. P., Adams, E. L., Butt, M. T., Musson, D. G., Henshaw, J., Keve, S., Cahayag, R., Tsuruda, L. S., and O'Neill, C. A. (2015) Nonclinical evaluation of CNS-administered TPP1 enzyme replacement in canine CLN2 neuronal ceroid lipofuscinosis. _Mol Genet Metab_ 114, 281-293

41. Dahms, N. M., Lobel, P., and Kornfeld, S. (1989) Mannose 6-phosphate receptors and lysosomal enzyme targeting. _J Biol Chem_ 264, 12115-12118
42. Sun, P., Sleat, D. E., Lecocq, M., Hayman, A. R., Jadot, M., and Lobel, P. (2008) Acid phosphatase 5 is responsible for removing the mannose 6-phosphate recognition marker from lysosomal proteins. *Proc Natl Acad Sci U S A* 105, 16590-16595

43. Jadot, M., Lin, L., Sleat, D. E., Sohar, I., Hsu, M. S., Pintar, J., Dubois, F., Wattiaux-De Coninck, S., Wattiaux, R., and Lobel, P. (1999) Subcellular localization of mannose 6-phosphate glycoproteins in rat brain. *J Biol Chem* 274, 21104-21113

44. Lubke, T., Lobel, P., and Sleat, D. E. (2009) Proteomics of the lysosome. *Biochimica et biophysica acta* 1793, 625-635

45. Sleat, D. E., Jadot, M., and Lobel, P. (2007) Lysosomal proteomics and disease. *Proteomics. Clinical applications* 1, 1134-1146

46. Sleat, D. E., Sohar, I., Pullarkat, P. S., Lobel, P., and Pullarkat, R. K. (1998) Specific alterations in levels of mannose 6-phosphorylated glycoproteins in different neuronal ceroid lipofuscinoses. *Biochem J* 334 (Pt 3), 547-551

47. Sleat, D. E., Zheng, H., Qian, M., and Lobel, P. (2006) Identification of sites of mannose 6-phosphorylation on lysosomal proteins. *Mol Cell Proteomics* 5, 686-701

48. Caval, T., Zhu, J., Tian, W., Remmelzwaal, S., Yang, Z., Clausen, H., and Heck, A. J. R. (2019) Targeted Analysis of Lysosomal Directed Proteins and Their Sites of Mannose-6-phosphate Modification. *Molecular & cellular proteomics : MCP* 18, 16-27

49. Naureckiene, S., Sleat, D. E., Lackland, H., Fensom, A., Vanier, M. T., Wattiaux, R., Jadot, M., and Lobel, P. (2000) Identification of HE1 as the second gene of Niemann-Pick C disease. *Science* 290, 2298-2301

50. Dislich, B., Wohlrab, F., Bachhuber, T., Muller, S. A., Kuhn, P. H., Hogl, S., Meyer-Luehmann, M., and Lichtenthaler, S. F. (2015) Label-free Quantitative Proteomics of Mouse Cerebrospinal Fluid Detects beta-Site APP Cleaving Enzyme (BACE1) Protease Substrates In Vivo. *Molecular & cellular proteomics : MCP* 14, 2550-2563
51. Smith, J. S., Angel, T. E., Chavkin, C., Orton, D. J., Moore, R. J., and Smith, R. D. (2014) Characterization of individual mouse cerebrospinal fluid proteomes. *Proteomics* 14, 1102-1106

52. Segal-Salto, M., Hansson, K., Sapir, T., Kaplan, A., Levy, T., Schweizer, M., Frotscher, M., James, P., and Reiner, O. (2017) Proteomics insights into infantile neuronal ceroid lipofuscinosis (CLN1) point to the involvement of cilia pathology in the disease. *Hum Mol Genet* 26, 1678

53. Tikka, S., Monogioudi, E., Gotsopoulos, A., Soliymani, R., Pezzini, F., Scifo, E., Uusi-Rauva, K., Tyynela, J., Baumann, M., Jalanko, A., Simonati, A., and Lalowski, M. (2016) Proteomic Profiling in the Brain of CLN1 Disease Model Reveals Affected Functional Modules. *Neuromolecular Med* 18, 109-133

54. Schmidtke, C., Tiede, S., Thelen, M., Kakela, R., Jabs, S., Makrypidi, G., Sylvester, M., Schweizer, M., Braren, I., Brocke-Ahmadinejad, N., Cotman, S. L., Schulz, A., Gieselmann, V., and Braulke, T. (2019) Lysosomal proteome analysis reveals that CLN3-defective cells have multiple enzyme deficiencies associated with changes in intracellular trafficking. *J Biol Chem*

55. Yeung, L., Hickey, M. J., and Wright, M. D. (2018) The Many and Varied Roles of Tetraspanins in Immune Cell Recruitment and Migration. *Front Immunol* 9, 1644

56. Kabashima, K., Nakashima, C., Nonomura, Y., Otsuka, A., Cardamone, C., Parente, R., De Feo, G., and Triggiani, M. (2018) Biomarkers for evaluation of mast cell and basophil activation. *Immunol Rev* 282, 114-120

57. Neal, M. L., Boyle, A. M., Budge, K. M., Safadi, F. F., and Richardson, J. R. (2018) The glycoprotein GPNMB attenuates astrocyte inflammatory responses through the CD44 receptor. *J Neuroinflammation* 15, 73

58. van der Lienden, M. J. C., Gaspar, P., Boot, R., Aerts, J., and van Eijk, M. (2018) Glycoprotein Non-Metastatic Protein B: An Emerging Biomarker for Lysosomal Dysfunction in Macrophages. *Int J Mol Sci* 20
59. Ohmi, K., Kudo, L. C., Ryazantsev, S., Zhao, H. Z., Karsten, S. L., and Neufeld, E. F. (2009) Sanfilippo syndrome type B, a lysosomal storage disease, is also a tauopathy. *Proc Natl Acad Sci U S A* 106, 8332-8337

60. Ohmi, K., Greenberg, D. S., Rajavel, K. S., Ryazantsev, S., Li, H. H., and Neufeld, E. F. (2003) Activated microglia in cortex of mouse models of mucopolysaccharidoses I and IIIB. *Proc Natl Acad Sci U S A* 100, 1902-1907

61. Haile, Y., Carmine-Simmen, K., Olechowski, C., Kerr, B., Bleackley, R. C., and Giuliani, F. (2015) Granzyme B-inhibitor serpina3n induces neuroprotection in vitro and in vivo. *J Neuroinflammation* 12, 157

62. Schmidt, C., Frahm, C., Schneble, N., Muller, J. P., Brodhun, M., Franco, I., Witte, O. W., Hirsch, E., Wetzker, R., and Bauer, R. (2016) Phosphoinositide 3-Kinase gamma Restrains Neurotoxic Effects of Microglia After Focal Brain Ischemia. *Mol Neurobiol* 53, 5468-5479

63. Russell, L. J., DiGiovanna, J. J., Rogers, G. R., Steinert, P. M., Hashem, N., Compton, J. G., and Bale, S. J. (1995) Mutations in the gene for transglutaminase 1 in autosomal recessive lamellar ichthyosis. *Nat Genet* 9, 279-283

64. Tatsukawa, H., Furutani, Y., Hitomi, K., and Kojima, S. (2016) Transglutaminase 2 has opposing roles in the regulation of cellular functions as well as cell growth and death. *Cell Death Dis* 7, e2244

65. Domowicz, M. S., Chan, W. C., Claudio-Vazquez, P., Henry, J. G., Ware, C. B., Andrade, J., Dawson, G., and Schwartz, N. B. (2019) Global Brain Transcriptome Analysis of a Tpp1 Neuronal Ceroid Lipofuscinoses Mouse Model. *ASN Neuro* 11, 1759091419843393

66. Khalil, M., Teunissen, C. E., Otto, M., Piehl, F., Sormani, M. P., Gattringer, T., Barro, C., Kappos, L., Comabella, M., Fazekas, F., Petzold, A., Blennow, K., Zetterberg, H., and Kuhle, J. (2018) Neurofilaments as biomarkers in neurological disorders. *Nat Rev Neurol* 14, 577-589

67. Lu, J. Y., Nelvagal, H. R., Wang, L., Birnbaum, S. G., Cooper, J. D., and Hofmann, S. L. (2015) Intrathecal enzyme replacement therapy improves motor function and survival in a
Biomarker discovery in NCL mouse models

preclinical mouse model of infantile neuronal ceroid lipofuscinosis. *Mol Genet Metab* 116, 98-105

68. Meng, Y., Sohar, I., Sleat, D. E., Richardson, J. R., Reuhl, K. R., Jenkins, R. B., Sarkar, G., and Lobel, P. (2014) Effective intravenous therapy for neurodegenerative disease with a therapeutic enzyme and a peptide that mediates delivery to the brain. *Mol Ther* 22, 547-553

69. Meng, Y., Wiseman, J. A., Nemtsova, Y., Moore, D. F., Guevarra, J., Reuhl, K., Banks, W. A., Daneman, R., Sleat, D. E., and Lobel, P. (2017) A Basic ApoE-Based Peptide Mediator to Deliver Proteins across the Blood-Brain Barrier: Long-Term Efficacy, Toxicity, and Mechanism. *Mol Ther* 25, 1531-1543

70. Wiseman, J. A., Meng, Y., Nemtsova, Y., Matteson, P. G., Millonig, J. H., Moore, D. F., Sleat, D. E., and Lobel, P. (2017) Chronic Enzyme Replacement to the Brain of a Late Infantile Neuronal Ceroid Lipofuscinosis Mouse Has Differential Effects on Phenotypes of Disease. *Mol Ther Methods Clin Dev* 4, 204-212

71. Vuillemenot, B. R., Katz, M. L., Coates, J. R., Kennedy, D., Tiger, P., Kanazono, S., Lobel, P., Sohar, I., Xu, S., Cahayag, R., Keve, S., Koren, E., Bunting, S., Tsuruda, L. S., and O’Neill, C. A. (2011) Intrathecal tripeptidyl-peptidase 1 reduces lysosomal storage in a canine model of late infantile neuronal ceroid lipofuscinosis. *Mol Genet Metab* 104, 325-337
Figure 1. Relative protein expression in late-stage NCL mouse model brain determined by MS3 quantitation of isobaric peptides. Axes are truncated at $-\log_{10}q$ value of 6 and $\log_2$ NCL/control ratio of -4 to 4. Q values were calculated by unpaired t tests of data from each respective mouse NCL time-point (n=3 animals) against adult wild-type animals (n=6 animals) using the False Discovery Rate two-stage step up method of Benjamini, Krieger and Yekutieli. Dashed lines indicate an FDR of 2% (y axis) and fold-change of 2-fold (x axis). Red and black symbols indicate known lysosomal (based on (17)) and other proteins, respectively. Names of select proteins of interest are shown in red (lysosomal) or black (other).
Figure 2. Relative protein expression in NCL mouse model brain determined by MS2 quantitation of isobaric labeled peptides. Data are displayed as in Fig. 1.
Figure 3. Correlation of changes in protein expression in late-stage CLN1 brain detected by MS2 and MS3 measurement of isobaric label reporter ions.

- lysosomal, FDR ≤2% MS2 & MS3
- other, FDR ≤2% MS2 & MS3
- lysosomal, FDR ≤2% MS3 only
- other, FDR ≤2% MS3 only
- other, FDR ≤2% MS2 only

![Graph showing correlation of changes in protein expression](https://www.mcponline.org)
Figure 4. Subcellular distribution of quantified proteins and proteins that are significantly (FDR ≤ 2%) altered in late stage NCL mouse brain. Note that this represents the set of proteins whose subcellular distribution is listed in Expt A of the Prolocate database. Based on Fisher’s Exact test, lysosomal proteins are significantly over represented in late stage CLN1, CLN2 and CLN3 (p<0.001).
Figure 5. Comparison of proteins that are significantly altered (FDR 2%) in either mouse and/or human brain. Samples were terminal for human or late-stage (19 and 26 weeks for CLN1 and CLN2 respectively) for mouse. Red symbols, lysosomal proteins; black symbols, proteins that are not assigned to the lysosome. Linear regression analyses excluded PPT1 and TPP1 for CLN1 and CLN2, respectively. For CLN3, there are insufficient datapoints for meaningful regression analysis.
Figure 6. Brain M6P glycoproteins in NCL mouse models. Indicated brain extracts were fractionated by SDS-PAGE, transferred to nitrocellulose and M6P glycoproteins detected using a radiiodinated M6P receptor. Early disease stage is 4 weeks for all genotypes while late stage corresponds to 26 weeks for CLN1, 19 weeks for CLN2 and 52 weeks for CLN3. All wild-type controls were age-matched to corresponding mutant animals.

| Disease stage | Early | Late |
|---------------|-------|------|
| Sample animal | wt A  | NCL A | wt A  | NCL A |
| µg extract   | 104   | 58   | 104   | 58 |

- CLN1
- CLN2
- CLN3
Figure 7. Relative expression of purified brain M6P glycoproteins in late-stage NCL mouse models. Axes are truncated at $-\log_{10} q$ value of 10 and $\log_2$ NCL/control ratio of -4 to 4 and analyses was restricted to proteins with a minimum sum of 50 spectral counts. Significance (q-values) was calculated by unpaired t tests using the False Discovery Rate (FDR) two-stage step up method of Benjamini, Krieger and Yekutieli. Dashed lines indicate an FDR of 2 % (y axis) and fold-change of 2-fold (x axis). Names of select proteins of interest are shown in red (lysosomal proteins) or black (non-lysosomal proteins). Animals were aged 26-weeks (CLN1 and age-matched controls, n=6 per group), 19 weeks (CLN2 and age-matched controls, n=5 per group) and 52-weeks (CLN3 and age-matched controls, n=6 per group).
**Figure 8.** Comparison of relative changes in expression of human and mouse purified brain M6P glycoproteins. Red symbols, known lysosomal proteins; black symbols, other proteins that were specifically purified on immobilized M6P receptor in a previous study (17). Filled symbols, proteins that are significantly elevated (FDR 2%) in human and/or mouse; open symbols, proteins that are not significantly elevated in either species. PPT1 and TPP1 were excluded from linear regression analyses of CLN1 and CLN2, respectively.
Figure 9. Correlation between changes in brain lysosomal proteins in late-stage disease from analysis of unpurified extracts (TMT-MS3) and purified M6P glycoproteins (spectral counting).
Figure 10. Relative expression of lysosomal and other proteins in NCL mouse model CSF. Axes are truncated at –log₁₀qvalue of 15 and log₂ NCL/control ratio of -4 to 4. Q values were calculated by unpaired t tests of peptide data corresponding to each protein in NCL versus age-matched wild-type control (10 animals per genotype pooled) using the False Discovery Rate two-stage step up method of Benjamini, Krieger and Yekutieli. Dashed lines indicate an FDR of 2% (y axis) and fold-change of 2-fold (x axis). Red and black symbols indicate known lysosomal and other proteins, respectively. Names of proteins of interest (see Results) are shown.
Figure 11. Proteins that are significantly elevated (FDR ≤ 2%) in both brain and CSF at late stage time-points. Red symbols, lysosomal proteins; black symbols, non-lysosomal proteins. Filled symbols, proteins that are significantly elevated in both brain and CSF; open symbols, proteins that are significantly elevated in brain or CSF.
Figure 12. Comparison of transcriptomic and proteomic changes in the CLN2 mouse model. Global brain transcriptome analyses were conducted using isolated fore/midbrain samples and cerebellum from 16-week CLN2 mice (65) and compared with corresponding changes measured by TMT-MS3 in whole brain extracts at 19 weeks. Log$_2$ transformed ratios for CLN2/control are truncated at 4 and -4.
**Table 1. Summary of protein identification and quantitation data.** TMT10-plex labeled mouse brain and CSF samples were analyzed by either MS3 and/or MS3 and proteins and peptides identified are shown. Data are filtered for quantifiable peptides as described in Methods. Nd, not determined.

| Sample | Peptides | Proteins (≥ 2 unique peptides) | QExactive HF MS2 | Lumos MS3 |
|--------|----------|-------------------------------|------------------|-----------|
|        |          |                               | total            | quantifiable | total | quantifiable |
| brain  | 550,320  | 7204                          | 415,831          | 293,358    |
| CSF    | 103,176  | 4688                          | nd               | nd         |
Table 2. Proteins significantly (FDR ≤2%) altered in both CLN1 and CLN2 brain.

| Protein     | CLN1 / wt ratio | CLN2 / wt ratio | inflammatory (pro or anti) pathways and/or immune function | lysosomal |
|-------------|-----------------|-----------------|-------------------------------------------------------------|----------|
| ABHD4       | 1.43            | 1.24            |                                                             |          |
| AIF1        | 1.77            | 1.86            | yes                                                        |          |
| ALDH1L2     | 1.82            | 1.45            |                                                             |          |
| ALDH7A1     | 1.29            | 1.19            |                                                             |          |
| ANXA5       | 1.51            | 1.19            |                                                             |          |
| APOE        | 1.94            | 1.66            | yes                                                        |          |
| ARPC1B      | 1.77            | 1.42            |                                                             |          |
| ARSA        | 1.45            | 1.45            |                                                             |          |
| ASPH        | 1.36            | 1.29            |                                                             |          |
| C1QA        | 2.94            | 2.32            | yes                                                        |          |
| C1QB        | 2.61            | 2.06            | yes                                                        |          |
| C1QC        | 3.01            | 2.45            | yes                                                        |          |
| C4B         | 7.21            | 3.40            | yes                                                        |          |
| CAPG        | 2.89            | 1.95            | yes                                                        |          |
| CD44        | 3.01            | 1.74            | yes                                                        |          |
| CD63        | 4.04            | 1.72            | yes                                                        |          |
| CPNE1       | 1.21            | 1.16            | yes                                                        |          |
| CPNE3       | 1.56            | 1.35            |                                                             |          |
| CTNNA1      | 1.41            | 1.17            | yes                                                        |          |
| CTSA        | 1.94            | 1.88            | yes                                                        |          |
| CTSD        | 4.02            | 1.50            | yes                                                        |          |
| CTSD2       | 2.87            | 2.21            | yes                                                        |          |
| CYBRD1      | 2.25            | 1.62            | yes                                                        |          |
| DNAJC13     | 1.12            | 1.16            |                                                             |          |
| EMC1        | 1.11            | 1.07            |                                                             |          |
| ENSMUSP00000087960 | 2.42 | 1.89 |          |          |
| FARP1       | 1.28            | 1.15            |                                                             |          |
| FUCA1       | 2.87            | 1.63            | yes                                                        |          |
| GANAB       | 1.16            | 1.13            |                                                             |          |
| GBP2        | 3.55            | 2.24            | yes                                                        |          |
| GFAP        | 3.85            | 2.45            | yes                                                        |          |
| GFPT1       | 1.23            | 1.18            |                                                             |          |
| GLIPR2      | 1.92            | 1.48            |                                                             |          |
| GM20425     | 1.35            | 1.19            |                                                             |          |
| GM28035     | 1.20            | 1.19            |                                                             |          |
| GNS         | 1.56            | 1.31            | yes                                                        |          |
| GPNMB       | 14.54           | 3.44            | yes                                                        |          |
| HADHA       | 1.31            | 1.15            |                                                             |          |
| HEXB        | 2.19            | 1.50            | yes                                                        |          |
| Genes      | Ratio | Value | Status |
|-----------|-------|-------|--------|
| HRSP12    | 1.79  | 1.41  |        |
| HSD17B4   | 1.28  | 1.16  |        |
| HSPB1     | 1.65  | 1.78  | yes    |
| IFIT3     | 2.64  | 2.16  | yes    |
| IRGM1     | 2.37  | 1.67  | yes    |
| ITGB2     | 2.00  | 1.62  | yes    |
| LCP1      | 1.91  | 1.50  | yes    |
| LYZ2      | 7.32  | 2.30  | yes    |
| MLEC      | 1.13  | 1.09  |        |
| MTHFD1    | 1.13  | 1.12  |        |
| NAPRT     | 1.86  | 1.52  |        |
| NPL       | 1.46  | 1.24  |        |
| PIK3CG    | 0.18  | 0.15  | yes    |
| PLA2G7    | 1.54  | 1.35  | yes    |
| PSMB8     | 1.87  | 1.39  | yes    |
| S100A4    | 3.42  | 1.97  | yes    |
| S100A6    | 4.21  | 2.19  | yes    |
| SCARB2    | 2.48  | 1.32  | yes    |
| SCPEP1    | 1.48  | 1.35  | yes    |
| SERPINA3N | 11.39 | 4.70  | yes    |
| SERPINE2  | 1.29  | 1.19  | yes    |
| SGTB      | 0.79  | 0.65  | yes    |
| SLC24A2   | 0.846 | 0.863 |        |
| SLC39A12  | 1.252 | 1.175 |        |
| STAT1     | 1.492 | 1.447 | yes    |
| STT3A     | 1.306 | 1.247 |        |
| STX7      | 1.160 | 1.210 | yes    |
| TGM1      | 18.518| 5.026 |        |
| TMED9     | 1.157 | 1.140 |        |
| TPP1      | 4.113 | 0.468 | yes    |
| TRAM1     | 1.360 | 1.209 | yes    |
| VWA5A     | 1.442 | 1.199 |        |
Table 3. Proteins significantly altered in mouse CLN1 brain and CSF at late-stage disease. Fold-changes compared to wild-type controls are filtered for proteins showing unidirectional changes and significant changes (FDR ≤2%) are indicated by bold numerals.

| Protein | BRAIN 4-week | BRAIN 26-week | CSF 4-week | CSF 26-week |
|---------|--------------|---------------|------------|-------------|
| CTSA    | 1.226        | 1.936         | 1.294      | 1.543       |
| CTSD    | 1.633        | 4.023         | 1.151      | 1.574       |
| CTSS    | 1.029        | 2.148         | 1.193      | 1.763       |
| CTS2    | 1.131        | 2.869         | 1.239      | 2.450       |
| FUCA1   | 1.512        | 2.867         | 1.644      | 1.757       |
| GALNS   | 0.948        | 1.279         | 1.195      | 1.192       |
| GUSB    | 0.955        | 2.160         | 1.054      | 1.611       |
| HEXA    | 1.143        | 2.269         | 1.259      | 1.893       |
| HEXB    | 1.149        | 2.194         | 1.244      | 1.122       |
| LAMP2   | 1.172        | 2.063         | 1.538      | 1.512       |
| NAGLU   | 1.004        | 1.678         | 1.049      | 1.576       |
| PPT1    | 0.245        | 0.165         | 0.375      | 0.383       |
| ACSF2   | 1.026        | 1.370         | 1.463      | 0.641       |
| ADD3    | 1.022        | 1.187         | 1.245      | 0.653       |
| AK3     | 0.966        | 1.221         | 1.124      | 0.736       |
| ALDH1L1 | 0.996        | 1.335         | 1.353      | 0.610       |
| ALDH4A1 | 1.004        | 1.233         | 0.836      | 0.625       |
| ANXA2   | 0.942        | 1.380         | 0.701      | 0.701       |
| ANAX5   | 0.958        | 1.506         | 0.887      | 0.883       |
| APOD    | 1.192        | 1.823         | 1.244      | 1.591       |
| APOE    | 1.126        | 1.943         | 1.122      | 1.765       |
| ASS1    | 1.086        | 1.210         | 1.581      | 0.836       |
| C1QA    | 1.092        | 2.942         | 1.257      | 1.674       |
| C4B     | 1.149        | 7.213         | 1.045      | 2.796       |
| CLU     | 0.965        | 1.863         | 1.223      | 1.525       |
| CP      | 0.964        | 1.217         | 1.171      | 1.136       |
| CST3    | 0.961        | 1.396         | 1.264      | 1.322       |
| EPB41   | 0.955        | 1.188         | 0.947      | 1.260       |
| EPB41L2 | 0.978        | 1.142         | 1.076      | 0.700       |
| EPHX1   | 1.077        | 1.336         | 1.014      | 0.741       |
| F13A1   | 0.998        | 1.662         | 0.819      | 0.801       |
| FLNA    | 0.986        | 1.416         | 0.780      | 1.263       |
| GPLD1   | 1.113        | 1.228         | 1.089      | 1.356       |
| IQGAP1  | 0.932        | 1.515         | 0.645      | 0.795       |
| KIF5B   | 0.934        | 1.179         | 0.921      | 0.725       |
| LRP1    | 1.006        | 1.114         | 0.849      | 1.183       |
| Gene    | Fold Change 1 | Fold Change 2 | Fold Change 3 | Fold Change 4 |
|---------|---------------|---------------|---------------|---------------|
| LYZ2    | 0.856         | 7.321         | 1.159         | 3.954         |
| MARC2   | 1.098         | 1.160         | 1.011         | 0.692         |
| MPP6    | 1.025         | 1.112         | 1.241         | 0.644         |
| MTHFD1  | 0.955         | 1.135         | 0.914         | 0.726         |
| MYH9    | 1.028         | 1.151         | 0.727         | 1.295         |
| NMRL1   | 0.784         | 0.772         | 0.960         | 0.699         |
| OAT     | 0.977         | 1.523         | 1.039         | 0.687         |
| OPLAH   | 1.013         | 1.298         | 0.625         | 0.670         |
| PCX     | 1.060         | 1.065         | 1.122         | 0.628         |
| PDIA3   | 0.841         | 1.195         | 0.886         | 0.726         |
| PDIA4   | 0.864         | 1.383         | 0.872         | 0.777         |
| PDIA6   | 0.864         | 1.206         | 0.728         | 0.745         |
| PLA2G7  | 1.158         | 1.536         | 1.073         | 1.564         |
| PLTP    | 0.897         | 1.637         | 1.032         | 1.478         |
| PSAT1   | 0.923         | 1.166         | 1.245         | 0.708         |
| PYGB    | 1.063         | 1.146         | 1.366         | 0.550         |
| SERPINA3N | 0.902   | 11.394        | 0.991         | 4.654         |
| SUCLG2  | 1.012         | 1.265         | 1.126         | 0.530         |
| TLN1    | 0.984         | 1.262         | 0.727         | 1.234         |
| VCAM1   | 0.974         | 1.548         | 1.043         | 1.373         |
| VWA5A   | 0.981         | 1.442         | 0.752         | 0.705         |
Table 4. Proteins significantly altered in mouse CLN2 brain and CSF at late-stage disease. Fold-changes compared to wild-type controls that are significant (FDR≤2%) are indicated by **bold** numerals.

| Protein  | 4 weeks | 19 weeks | 4 weeks | 19 weeks |
|----------|---------|----------|---------|----------|
| **BRAIN** |         |          |         |          |
| CTSA     | 1.537   | 1.885    | 1.222   | 1.922    |
| CTSD     | 1.047   | 1.499    | 1.244   | 3.023    |
| CTSZ     | 1.117   | 2.208    | 1.248   | 3.131    |
| GNS      | 0.954   | 1.309    | 0.994   | 1.489    |
| HEXB     | 0.977   | 1.495    | 1.231   | 1.859    |
| LGMN     | 1.519   | 1.871    | 1.360   | 1.449    |
| **CSF**  |         |          |         |          |
| ALDH7A1  | 1.035   | 1.193    | 1.095   | 0.755    |
| ANXA5    | 0.968   | 1.195    | 1.069   | 0.848    |
| APOE     | 0.990   | 1.659    | 1.320   | 3.235    |
| C1QA     | 1.139   | 2.321    | 1.366   | 2.478    |
| C1QC     | 1.145   | 2.454    | 1.206   | 2.428    |
| C4B      | 1.134   | 3.404    | 1.228   | 3.383    |
| CTNNA1   | 0.989   | 1.171    | 1.088   | 1.998    |
| HRSP12   | 1.090   | 1.411    | 1.022   | 0.581    |
| LYZ2     | 1.176   | 2.300    | 1.284   | 3.963    |
| MTHFD1   | 0.953   | 1.116    | 0.906   | 0.436    |
| NAPRT    | 0.987   | 1.523    | 1.001   | 0.538    |
| NEFH     | 1.130   | 0.758    | 1.736   | 7.022    |
| NEFM     | 1.017   | 0.816    | 2.268   | 4.486    |
| NNT      | 14.245  | 21.164   | 8.033   | 3.484    |
| PLA2G7   | 1.103   | 1.349    | 1.035   | 1.989    |
| SERPINA3N| 1.161   | 4.695    | 1.119   | 7.150    |
| SERPINE2 | 1.039   | 1.193    | 1.165   | 1.391    |
| TGM1     | 0.997   | 5.026    | 1.695   | 2.585    |
| VWA5A    | 0.983   | 1.199    | 0.849   | 0.691    |
Table 5. Comparison with microarray data. + and – indicate direction of significant changes.

| Genbank Accession | Gene name | Elevated transcript | CLN1 brain | CSF | CLN2 brain | CSF | CLN3 brain | CSF |
|-------------------|-----------|---------------------|------------|-----|------------|-----|------------|-----|
| X58861            | C1qa      | MPSI & MPSIIB       | +          | +   | +          | +   |            |     |
| M22531            | C1qb      | MPSI & MPSIIB       | +          |     |            |     |            |     |
| X66295            | C1qc      | MPSI & MPSIIB       | +          |     |            |     |            |     |
| X06454            | C4b       | MPSI & MPSIIB       | +          | +   | +          | +   | +          |     |
| AJ223208          | Cts      | MPSI & MPSIIB       | +          |     |            |     |            |     |
| AJ242663          | Ctsz     | MPSI & MPSIIB       | +          |     | +          |     | +          |     |
| X02801            | Gfap     | MPSI & MPSIIB       | +          |     | +          |     | +          |     |
| D16195            | Grn       | MPSI & MPSIIB       | +          |     |            |     |            |     |
| M21050            | Lyz2     | MPSI & MPSIIB       | +          |     | +          |     |            |     |
| AF053943          | Aebp1    | MPSIIIB             | +          |     | +          |     | +          |     |
| U74683            | Cts      | MPSIIIB             | +          |     |            |     |            |     |
| X53929            | Dcn      | MPSIIIB             | +          |     | -          |     |            |     |
| AV00854           | Gapdh    | MPSIIIB             | -          |     | -          |     | -          |     |
| X98113            | Lag3     | MPSIIIB             | +          |     |            |     |            |     |
| U22033            | Psmb8    | MPSIIIB             | +          |     | +          |     |            |     |
| M64086            | Serpina3n| MPSIIIB             | +          |     | +          |     | +          |     |
| AF010254          | Serping1 | MPSIIIB             | +          |     |            |     | +          |     |