CLN8 is an endoplasmic reticulum cargo receptor that regulates lysosome biogenesis

Alberto di Ronza¹, Lakshya Bajaj¹, Jaiprakash Sharma¹, Deepthi Sanagasetti¹, Parisa Lotfi¹, Carolyn Joy Adamski¹, John Collette², Michela Palmieri³, Abdallah Amawi¹, Lauren Popp³, Kevin Tommy Chang¹, Maria Chiara Meschini³, Hon-Chiu Eastwood Leung⁶, Laura Segatori³, Alessandro Simonati³, Richard Norman Sifers², Filippo Maria Santorelli⁷ and Marco Sardiello⁸ ¹*

Organelle biogenesis requires proper transport of proteins from their site of synthesis to their target subcellular compartment. Lysosomal enzymes are synthesized in the endoplasmic reticulum (ER) and traffic through the Golgi complex before being transferred to the endolysosomal system, but how they are transferred from the ER to the Golgi is unknown. Here, we show that ER-to-Golgi transfer of lysosomal enzymes requires CLN8, an ER-associated membrane protein whose loss of function leads to the lysosomal storage disorder, neuronal ceroid lipofuscinosis 8 (a type of Batten disease). ER-to-Golgi trafficking of CLN8 requires interaction with the COPII and COPI machineries via specific export and retrieval signals localized in the cytosolic carboxy terminus of CLN8. CLN8 deficiency leads to depletion of soluble enzymes in the lysosome, thus impairing lysosome biogenesis. Binding to lysosomal enzymes requires the second luminal loop of CLN8 and is abolished by some disease-causing mutations within this region. Our data establish an unanticipated example of an ER receptor serving the biogenesis of an organelle and indicate that impaired transport of lysosomal enzymes underlies Batten disease caused by mutations in CLN8.

Lysosomes play critical roles in the maintenance of cellular homeostasis by degrading and recycling the majority of cellular macromolecules via the autophagic, endocytic and phagocytic programmes. Lysosomal digestive functions rely on more than 50 hydrolytic enzymes whose synthesis is coordinated by a genetic programme that oversees the cell’s catabolic needs. Lysosomal enzymes are trafficked to the lysosome in two stages: transport of the newly synthesized proteins from the endoplasmic reticulum (ER) to the Golgi complex, and their subsequent receptor-assisted transfer from the Golgi to endolysosomal compartments. How lysosomal enzymes are transported from the ER to the Golgi complex is unknown, and whether this process is aided by specific receptors has not been investigated.

Mounting evidence supports the idea that protein cargos require specific cargo receptor systems for proper sorting along the secretory route. Thus, we hypothesized that ER exit of lysosomal enzymes is assisted by a dedicated receptor system. To identify and prioritize candidate ER receptors, we analysed the relationship between the expression of genes encoding human ER-resident proteins (hereafter referred to as ER genes) and that of genes encoding lysosomal enzymes (lysosomal genes). We used sets of expression microarray data from a wide variety of dynamic states following chemical, biological or genetic perturbation, an established approach for studying lysosomal regulation. We found that the expression of ~10% of ER genes correlates with that of lysosomal genes (Fig. 1a), notably including ER genes participating in the maturation of lysosomal proteins, such as oligosaccharyltransferase genes, γ-secretase complex genes and sulfatase-modifying factor genes (Supplementary Table 2). Four ER genes with expression correlated to lysosomal genes encode candidate cargo receptors or proteins that are possibly involved in vesicular transport: CLN8, TMED4, TMED9 and LMF1 (Supplementary Table 2). To determine whether any of these four candidate receptors interacts with lysosomal enzymes, we employed a bimolecular fluorescence complementation (BiFC) system based on a split yellow fluorescent protein (YFP) variant. We generated two libraries of plasmids encoding lysosomal enzymes (n=53) fused either to YFP amino-terminal fragment (Y1) or to YFP carboxy-terminal fragment (Y2), respectively, and two libraries of plasmids encoding the four candidate cargo receptors with Y1 or Y2 tags, which we confirmed localize to the ER (Supplementary Fig. 1a). A DQ-Red BSA assay to measure the proteolytic activity of the lysosome showed that the over-expression of candidate ER cargo receptors did not alter lysosomal degradation capability (Fig. 1b). Confocal microscopy of HeLa cells co-transfected with CLN8, TMED4, TMED9 or LMF1 plasmids and pools of plasmids expressing lysosomal enzymes showed that only the Y2-CLN8 construct consistently interacted with all of the pools (Fig. 1c and Supplementary Fig. 1b). Pairwise co-transfection of Y2-CLN8 with lysosomal enzymes followed by quantification through flow cytometry showed that CLN8 interacted with two-thirds of the lysosomal enzymes, but not with non-lysosomal proteins that we tested as a control (Fig. 1d and Supplementary Fig. 1c). Co-immunoprecipitation (co-IP) of CLN8-Myc followed by immunoblotting confirmed the interactions detected by the BiFC assay (Fig. 1e).

CLN8 is a ubiquitously expressed ER membrane protein of unknown function that forms homodimers. CLN8 deficiency

¹Department of Molecular and Human Genetics, Baylor College of Medicine, Jan and Dan Duncan Neurological Research Institute, Texas Children’s Hospital, Houston, TX, USA. ²Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX, USA. ³Department of Bioengineering, Chemical and Biomolecular Engineering, and Biochemistry and Cell Biology, Rice University, Houston, TX, USA. ⁴Department of Biosciences, Rice University, Houston, TX, USA. ⁵Department of Neurological and Movement Sciences, University of Verona, Verona, Italy. ⁶Departments of Medicine, Pediatrics, and Molecular and Cellular Biology, Dan Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA. ⁷IRCCS Stella Maris, Pisa, Italy. ⁸e-mail: sardiell@bcm.edu

*Corresponding author.
**Fig. 1** | CLN8 interacts with lysosomal enzymes.  

**a.** Co-expression analysis of ER and lysosomal (Lyso) genes. Shown is a heatmap representing the extent of pairwise co-expression between 620 ER genes (x axis) and 60 lysosomal genes (y axis). Among ER genes that are significantly co-expressed with lysosomal genes (P < 10^-4, two-tailed Kolmogorov-Smirnov test; vertical dotted line), CLN8, TMED4, LMFI and TMED9 encode candidate cargo receptors.  

**b.** DQ-Red BSA degradation assay to measure the proteolytic activity of lysosomes upon transfection of CLN8, TMED4, TMED9 and LMF1 plasmids. CTCF, corrected total cell fluorescence. Data are means ± s.e.m. (n = 3 independent experiments, n = 10 independent images quantified). Scale bar, 20 μm.  

**c.** Representative live imaging of reconstituted BiFC fluorescence between Y2-tagged, full-length candidate cargo receptors and pools of Y1-tagged lysosomal enzymes. Green fluorescence shows reconstitution of YFP as an indicator of protein–protein interaction. Control experiments (CTRL) were performed by co-transfecting Y2-tagged candidates with pools of Y2-tagged lysosomal enzymes. Scale bar, 200 μm.  

**d.** Pairwise interaction between Y2-CLN8 and lysosomal enzymes evaluated by BiFC followed by flow cytometry. The non-lysosomal proteins, angiogenin (AGN), insulin-like growth factor 1 (IGF1) and transforming growth factor β-1 (TGFβ1), and Y2-tagged CLN8 were used as negative controls. Data are means ± s.e.m. (n = 3 independent experiments, *P < 0.05, **P < 0.01, two-tailed Student’s t-test). Data are corrected for multiple comparison by using the Bayesian sequential goodness of fit (SGoF) procedure.  

**e.** Co-IP analysis of CLN8 and lysosomal enzymes. Proteins were transiently expressed in HeLa cells, and immunoprecipitates were analysed by immunoblotting (IB) with the indicated antibodies. β-hexosaminidase (HEXB) was used as a negative control. Input represents 10% of the total cell extract used for immunoprecipitation (IP). Molecular weights are in kDa. Images shown in **c** and **e** are representative of n = 3 independent experiments.
causes a form of neuronal ceroid lipofuscinosis (NCL) or Batten disease (CNL8; OMIM 600143), a fatal neurodegenerative disorder in which lysosomes accumulate ceroid lipopigments\textsuperscript{7,14}. It is worth noting that, although most lysosomal storage disorders are caused by deficiency of lysosomal enzymes, CLN8 is not in this category.

To determine whether CLN8 is required for proper delivery of lysosomal enzymes, we analysed the \textit{Cln8\textsuperscript{mnd}} mouse strain, which bears an early frameshift mutation in \textit{Cln8} that causes CLN8 protein deficiency and features resembling those seen in Batten disease\textsuperscript{7}.

To focus on the primary molecular insult caused by CLN8 deficiency, we analysed \textit{Cln8\textsuperscript{mnd}} mice at the pre-symptomatic age of 2 months\textsuperscript{15}. We obtained subcellular fractions enriched for lysosomes from the livers of \textit{Cln8\textsuperscript{mnd}} mice (knockout) and age-matched wild-type (WT) mice using a Nycodenz gradient (Supplementary Fig. 2a–c). Comparison of lysosome-enriched fractions by liquid chromatography–tandem mass spectrometry (LC–MS/MS) showed that \textit{Cln8\textsuperscript{mnd}} samples had lower levels of lysosomal enzymes than WT samples (Fig. 2a and Supplementary Fig. 2d).
set enrichment analysis (GSEA) of proteomic data showed that lysosomal enzymes were specifically depleted in Cln8mnd samples (Fig. 2b and Supplementary Table 3), whereas the sets of lysosomal membrane proteins—which are transported to lysosomes via a different route than lysosomal soluble proteins—and mitochondrial matrix proteins (used as a control) showed no overall variation (Supplementary Fig. 2e and Supplementary Tables 4 and 5). Immunoblot analysis confirmed lower levels of tested lysosomal enzymes in the lysosome-enriched fraction from Cln8mnd mice (Fig. 2c). Accordingly, the lysosome-enriched fraction from Cln8mnd mice showed reduced activity of several tested lysosomal enzymes (Supplementary Fig. 2f). However, the amounts of transcripts

Fig. 3 | Interaction with COPI and COPII complexes mediates CLN8 trafficking. a, Confocal microscopy showing ER-to-Golgi shift of CLN8 localization upon treatment with CBM. Trace outline is used for line-scan (white dashed line) analysis of relative fluorescence intensity of CLN8, GM130 (a Golgi marker) and KDEL signals. Signal overlap is quantified by Pearson correlation analysis of % independent experiments, n = 10 independent images quantified. b, Confocal microscopy of HeLa cells showing ER-to-Golgi shift of CLN8 localization upon KXXX signal mutagenesis (CLN8dK). c, BifC assay of α-COP with CLN8 showing disruption of interaction upon mutagenesis of CLN8’s KXXX signal (CLN8dK). d, Co-IP assay of α-COP and CLN8. Input represents 10% of the total cell extract. e, Confocal microscopy of BifC complexes composed of TPP1-Y1 and Y2-CLN8 (with or without CBM) or TPP1-Y1 and Y2-CLN8dK. f, Immunoblot of lysates from HeLa cells transfected with Myc-tagged Sec24 proteins after pull down with the cytosolic C-terminus of CLN8 fused with GST showing disruption of interaction upon mutagenesis of CLN8’s COPPI signal. g, Confocal microscopy of HeLa cells showing decreased Golgi localization of CLN8 upon mutagenesis of CLN8 ER export signal (CLN8dK) using the Golgi-localizing CLN8dK backbone. In a, b, e and g, trace outline, relative fluorescence intensity and Pearson correlation analyses are performed as in a. Images shown in c, d and f are representative of n = 3 independent experiments. In a, b, e and g, data are means ± s.e.m. (n = 3 independent experiments, n = 10 independent images quantified. ***P < 0.001, **P < 0.01, ***P < 0.001, two-tailed Student’s t-test). In a, b, e and g, scale bars, 20 µm. In c, scale bars, 200 µm. In d and f, molecular weights are in kDa.
Defective maturation of lysosomal enzymes upon CLN8 deficiency. a–c. Metabolic radiolabelling of CLN8<sup>−/−</sup> cells and their parental HeLa cells showing defective maturation of CTSD and palmitoyl-protein thioesterase 1 (PPT1; a), TPP1 (b) and GALNS (c) in the absence of CLN8. CTSD and PPT1 were immunoprecipitated by using antibodies against the endogenous proteins. 3xFlag-tagged (3xF) TPP1 and GALNS were expressed by using a doxycycline-inducible vector used to transduce CLN8<sup>−/−</sup> and control cells. Arrows indicate the mature, lysosome-associated enzyme. The asterisk in c indicates a nonspecific signal. d. Quantification of the mature enzymes for the metabolic radiolabelling experiment in a–c. e–f. Decreased enzyme stability in CLN8<sup>−/−</sup> cells compared to parental HeLa cells. CTSD and PPT1 proteins were monitored at the indicated time points following cycloheximide (CHX)-mediated blockage of protein synthesis (e). 3xFlag-tagged TPP1 and GALNS were expressed in cells transduced with a doxycycline-inducible vector and monitored at 0, 4, 8 and 12 h upon doxycycline removal to turn off their synthesis (f). In all experiments, GAPDH was used to normalize the residual protein for quantifications. g. Quantification of the immunoblot experiments in e and f. In d and g, data are means ± s.e.m. (n = 3 independent experiments, *P < 0.05, **P < 0.01, two-tailed Student’s t-test). For a–c, e and f, molecular weights are in kDa.

for lysosomal enzymes in the samples from Cln8<sup>null</sup> mice were unchanged or slightly increased (Supplementary Fig. 2g), indicating that the reduction of lysosomal enzymes occurs post-transcriptionally. Immunohistochemical analysis of cortical and cerebellar sections from Cln8<sup>null</sup> mice confirmed a reduction of enzyme signals (Supplementary Fig. 3a). Confocal microscopy revealed a reduction of enzyme signals at lysosomes (Fig. 2d and Supplementary Fig. 3b) and sparse spots of residual enzymes overlapping with the ER (Supplementary Fig. 4a), indicating a maturation defect. Confocal microscopy and immunoblot analyses of fibroblasts derived from patients with mutations in CLN8 (ref. 16) showed impaired colocalization of tripeptidyl-peptidase 1 (TPP1) with the lysosomal marker LAMP1 (Fig. 2e and Supplementary Fig. 4b) and depletion of tested lysosomal enzymes (Supplementary Fig. 4c).

Cargo receptors that traffic from the ER to the Golgi complex can be retrieved to the ER through interaction with COPI, a protein complex that serves as the coatomer of vesicles involved in retrograde transport<sup>17,18</sup>. COPI recognizes a KXXX retrieval signal at the C terminus of its target cargo receptors<sup>19,20</sup> and mediates a very
efficient retrieval that localizes the cargo receptors predominantly to the ER (and only minimally to the Golgi complex)\(^2\). Several lines of evidence indicated that CLN8 is retrieved from the Golgi complex to the ER in a COPI-dependent manner. First, CLN8 carries a terminal KKXX signal\(^2\) and localizes primarily to the ER (Supplementary Fig. 4d). Second, cell treatment with CBM, a drug that inhibits COPI-mediated vesicular transport and thereby abolishes the retrieval of cargo receptors\(^2\), increased localization of CLN8 to the Golgi complex (Fig. 3a), as did mutating the KKXX sequence of CLN8 (ref. \(^2\)) (Fig. 3b). Third, BiFC and co-IP experiments showed that CLN8 interacts with the COPI complex and this interaction could be disrupted either by mutating the KKXX signal of CLN8 or by CBM

---

**Fig. 5** CLN8 interaction with lysosomal enzymes requires the second luminal loop. a,b, Schematic representations of the CLN8 protein (a) and the CLN8ΔL construct (b). Amino acids are represented by coloured circles. Amino acids in yellow indicate the position of clinical mutations. The ER export signal and the ER retrieval signal at the protein C terminus are indicated. ERGIC, ER-Golgi intermediate compartment. c, A multi-alignment of CLN8 protein sequences along with a plot of local evolutionary rates. Transmembrane domains (TM), luminal domains (ER) and cytosolic domains (Cyt) are reported. Clinical mutations that fall in the second luminal loop are reported. ECRs, evolutionary-constrained regions. d, Confocal microscopy analysis showing that Y2-CLN8ΔL colocalizes with both CLN8-Myc and the ER marker KDEL. Scale bar, 20 μm. e, Confocal microscopy analysis showing that mutagenesis of the KKXX signal of the Y2-CLN8ΔL protein determines its localization to the Golgi, indicating that the second luminal loop of CLN8 is not required for CLN8 ER export. f, BiFC assay of Y1-CLN8 with Y2-CLN8ΔL showing a reconstituted GFP signal indistinguishable from that of Y1-CLN8 with Y2-CLN8. g, Comparative BiFC assay of Y1-tagged lysosomal enzymes with either Y2-CLN8 (top panels) or Y2-CLN8ΔL (bottom panels), showing disruption of interaction upon removal of the CLN8 second luminal loop. h, Comparative BiFC/flow cytometry analysis of Y1-tagged lysosomal enzymes with Y2-CLN8 or Y2-CLN8ΔL constructs. Values are expressed as a percentage of the enzyme–Y1/Y2-CLN8 interaction. i, j, Comparative BiFC/flow cytometry assay of CLN8 constructs harbouring clinical mutations in the second luminal loop. Images shown in d–g are representative of \(n = 3\) independent experiments. In h–j, data are means ± s.e.m. (\(n = 3\) independent experiments; NS, not significant, **\(P < 0.01\), two-tailed Student’s t-test). In i and j, the group-level \(P\) values were estimated from the mean \(z\)-scores from each individual test. In d and e, scale bars, 20 μm. In f and g, scale bars, 200 μm.

---

To further investigate the role of the second luminal loop in the localization of CLN8 to the Golgi complex, we performed a multi-alignment of CLN8 protein sequences from various species (Fig. 5c). This analysis revealed evolutionary-constrained regions (ECRs) that are conserved across different species. We then performed a comparative BiFC/flow cytometry analysis of Y1-tagged lysosomal enzymes with Y2-CLN8 or Y2-CLN8ΔL constructs (Fig. 5d–g). This analysis showed that the second luminal loop of CLN8 is not required for CLN8 ER export, as demonstrated by the reconstitution of the GFP signal in the BiFC assay (Fig. 5f).

Additionally, we performed comparative BiFC/flow cytometry assays of CLN8 constructs harbouring clinical mutations in the second luminal loop (Fig. 5i–j). These assays showed that the group-level \(P\) values were not significant, indicating that the second luminal loop is not essential for the localization of CLN8 to the Golgi complex.

In conclusion, our findings suggest that the second luminal loop of CLN8 is not required for its localization to the Golgi complex, and that the KKXX signal is a critical determinant for the retrieval of CLN8 from the Golgi complex to the ER. These results provide new insights into the mechanisms of cargo retrieval in the secretory pathway and may have implications for the treatment of CLN8-related diseases.
(Fig. 3c–d and Supplementary Fig. 4e). To test whether CLN8 determines the localization of newly synthesized lysosomal enzymes, we took advantage of the fact that, in the BiFC assay, interacting proteins form a stable complex\(^{12}\) whose localization can be monitored by confocal microscopy. Consistent with the above results, mutagenesis of CLN8’s KXXX retrieval signal and CBM inhibition of COPI-mediated retrieval each led to increased localization of the enzyme–Y1/Y2-CLN8 complex to the Golgi (Fig. 3e).

We next investigated the association of CLN8 with COPI, the coatamer of vesicles involved in ER-to-Golgi anterograde traffic\(^{29}\). COPI interacts with cargo receptors via its Sec24 subunits, which recognize specific cytosolic ER export signals within the cargo receptor protein sequence. The cytosolic tail of CLN8 contains a \(\Phi\Phi\Phi\Phi\Phi\Phi\Phi\) motif that overlaps with the \(\Phi\Phi\Phi\Phi\Phi X\) ER export signal\(^{29}\). Lysates derived from HeLa cells expressing Myc-tagged Sec24A, Sec24B, Sec24C or Sec24D were subjected to pull down using the cytosolic tail of CLN8 fused with glutathione S-transferase (GST). Immunoblotting revealed interaction of CLN8 with Sec24a and Sec24c; this interaction was abolished by mutations in the \(\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Phi\Ph...
21. Dominguez, M. et al. gp25L/emp24/p24 protein family members of the cis-Golgi network bind both COP I and II coatamer. *J. Cell Biol.* **140**, 751–765 (1998).

22. Emery, G., Parton, R. G., Rojo, M. & Grünberg, J. The trans-membrane protein p25 forms highly specialized domains that regulate membrane composition and dynamics. *J. Cell Sci.* **116**, 4821–4832 (2003).

23. Lonka, L., Kyttala, A., Ranta, S., Jalanko, A. & Lehesjoki, A. E. The neuronal ceroid lipofuscinosis CLN8 membrane protein is a resident of the endoplasmic reticulum. *Hum. Mol. Genet.* **9**, 1691–1697 (2000).

24. Hu, T., Kao, C. Y., Hudson, R. T., Chen, A. & Draper, R. K. Inhibition of secretion by 1,3-cyclohexanebis(methylamine), a dibasic compound that interferes with coatamer function. *Mol. Biol. Cell* **10**, 921–933 (1999).

25. Kuehn, M. J., Herrmann, J. M. & Schekman, R. COPII–cargo interactions direct protein sorting into ER-derived transport vesicles. *Nature* **391**, 187–190 (1998).

26. Otsu, W., Kurooka, T., Otsuka, Y., Sato, K. & Inaba, M. A new class of endoplasmic reticulum export signal ΦXΦΦΦ for transmembrane proteins and its selective interaction with Sec24C. *J. Biol. Chem.* **288**, 18521–18532 (2013).

27. Kousi, M., Lehesjoki, A. E. & Mole, S. E. Update of the mutation spectrum and clinical correlations of over 360 mutations in eight genes that underlie the neuronal ceroid lipofuscinoses. *Hum. Mutat.* **33**, 42–63 (2012).

28. Hamalisto, S. & Jaattela, M. Lysosomes in cancer-living on the edge (of the cell). *Curr. Opin. Cell Biol.* **39**, 69–76 (2016).

29. Perera, R. M. et al. Transcriptional control of autophagy–lysosome function drives pancreatic cancer metabolism. *Nature* **524**, 361–365 (2015).

Acknowledgements

We thank P. Lobel, D. Pearce, J. Cooper, T. Diersk, M. Damme, Z. Liu and S. Elsea for helpful discussion; A. Schiano and M. Rousseaux for technical assistance; H. Zoghbi, V. Brandt, A. Ballabio, H. Jafar-Nejad, K. Venkatachalam and M. Wang for critical reading of the manuscript; and B. Turk (J. Stefan Institute, Slovenia), K. Yamada (Institute for Developmental Research, Japan), B. Blazar (University of Minnesota, USA), D. Kohn (University of California, Los Angeles, USA), T. Beccari (Università degli Studi di Perugia, Italy) and M. Peterfy (University of California, Los Angeles, USA) for providing plasmids encoding lysosomal proteins or other tested proteins. This work was supported by the NIH grant NS079618 (to M.S.) and grants from the Beyond Batten Disease Foundation (to M.S.) and the NCL-Stiftung (to M.S.). This project was supported in part by the Hamill Foundation and by IDDRC grant number 1U54 HD083092 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

**Author contributions**

M.S. conceived and supervised the study. A.d.R. and M.S. designed the experiments and analysed the data with the contribution of R.N.S., L.S., A.S. and F.M.S. A.d.R. performed the confocal analysis, flow cytometry, the cycloheximide assay, qPCR and enzyme assays. A.d.R. and L.B. performed subcellular fractionation. A.d.R. and J.S. performed the Co-IP experiments. A.d.R., L.B., J.S., D.S., P.L., C.J.A., J.C., M.P., A.A., L.P., K.T.C. and M.C.M. performed cloning, cell culture and transfection experiments. A.d.R. and P.L. performed immunohistochemistry. H.-C.E.L. performed LC–MS/MS. M.S. performed the co-expression and evolutionary analyses. A.d.R. and M.S. wrote the manuscript with help from H.-C.E.L., L.S., A.S. and F.M.S. All authors reviewed and edited the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41556-018-0228-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.S.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2018
Methods
Molecular biology, cell culture and transfection. Complementary DNA libraries generated by retrotranscription of RNAs from HeLa and HEK293 cells using QuantiTect Reverse Transcription kit were used to PCR-ampify cDNAs for lysosomal proteins and candidate receptors. cDNAs were then inserted in pcDNA3.1[EF1], pcDNA3.1[EF2], pPDND0-inFusion[3xFlag] or pGEX-2T vectors using the in-Fusion cloning kit (Clontech). CLN8 cDNA was also cloned into the pcDNA3.1/Myc vector (Invitrogen). The oligonucleotides used for InFusion cloning are reported in Supplementary Table 6. pcDNA3.1[EF1]-CLN8 and pcDNA3.1[EF2]-CLN8 constructs were generated by inserting the signal peptide sequence of cathepsin U into the YFP sequence to assure ER import of the constructs and a lumen-facing YFP tag. Cells were grown at 37 °C in 5% CO2 in DMEM (Euroclone), supplemented with 1% glutamine and Pen-Strep. HeLa cells and biopsied fibroblasts were grown with 10% and 20% heat-inactivated FBS (HyClone), respectively. Transfection of full-length cDNAs in HeLa cells was performed using Lipofectamine LTX Transfection Reagent (Invitrogen), according to the manufacturer’s directions. Short interfering RNA (siRNA) oligonucleotides were purchased from Thermo Scientific. CLN8 expression was knocked down using ON-TARGETplus SMARTpool (L-013004-00-00), Non-silencing control siRNA analysis was performed using ON-TARGETplus Non-targeting Pool (D-001810-10-20). Transfection of siRNAs (50 nM, 24 h after cell plating) in HeLa cells was performed by using Lipofectamine RNAiMAX (Invitrogen) according to the manufacturer’s instructions. CLN8dK constructs, CLN8 constructs harbouring NCL1-related mutations, and glycosylation mutations of cathepsin D (CTSD), N-acetylglactosamine-6-sulfatase (GALNS) and ganglioside GM2 activator (GM2A) were obtained by using the QuikChange XLII Site-directed Mutagenesis kit (Agilent Technologies) and followed by Sanger sequencing of the insert to confirm the correct sequence. HeLa cells were selected for more than 1 week in geneticin-containing medium (G418, 200 µg ml−1) and incubated at 37 °C for 48 h. Cells were collected for more than 1 week in genenicin-containing medium (G418, 200 µg ml−1) and incubated at 37 °C for 48 h. Cells were collected for more than 1 week in genenicin-containing medium (G418, 200 µg ml−1) and incubated at 37 °C for 48 h.

Lentivirus generation, infection and expression. CLN8, GALNS and TPP1 full- coding sequences were PCR amplified and cloned into the pPDND0-20 lentiviral vector by InFusion procedure after removal of the ccdB site, and insertion of a 3’ XhoI cassette and a SaI restriction site into the vectors’ backbone, obtaining the pPDND0-CLN8-3xFlag, pPDND0-GALNS-3xFlag and pPDND0-TPP1-3xFlag constructs. The oligonucleotides used for CLN8, TPP1 and GALNS cloning into the pPDND0-20 vector are shown in Supplementary Table 6. Lentiviral vectors and their respective packaging vectors (psPAX2 and pMD2G) were transfected into HEK293T cells in a 4:1 molar ratio, respectively. Media were changed 16 h following transfection to low-volume media (5 ml for a 10-cm dish). Media were collected at 48 h following transfection, replaced with fresh media (5 ml) and collected again at 72 h. Viral supernatant was cleared from debris via centrifugation (10 min at 4,000 r.p.m.) as well as filtration through a 0.45-µm polyethersulfone membrane (VWR). Viruses were added to the receiving cells in complete media with polybrene (8μg ml−1) and incubated at 37 °C for 48 h. Cells were selected for more than 1 week in genenicin-containing medium (G418, 400 µg ml−1). Expression of the transgenes was induced with 2 µg ml−1 doxycycline.

Co-IP. Cells were lysed in cold NP-40 lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% NP-40, 0.1% SDS, protease inhibitors cocktail (Sigma)). Lysates were pre-cleared by incubation with pre-immune serum for 1 h. Protein G-agarose (Roche) was then added to the mixtures and gently rocked for 3 h at 4 °C. The beads were washed three times with NP-40 lysis buffer, and the resulting immunoprecipitates were analysed by western blotting.

Western blot analysis. Cells were lysed in cold lysis buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl and 1% Triton X-100) in the presence of protease inhibitors (Sigma) for 30 min on ice. Total protein concentrations were determined by BCA Protein Assay (Pierce). Protein samples were separated on 4–12% SDS-PAGE gels, transferred to nitrocellulose membranes using the iBlot transfer system (Invitrogen) according to manufacturer’s instructions. Primary and horseradish peroxidase-conjugated secondary antibodies (Supplementary Table 7) were diluted in 5% milk in TBST. Detection was carried out with ECL Western Blotting detection reagent (GE Healthcare). Images were detected with ImageQuant LAS 4000 (GE Healthcare) and quantified by Fiji analysis software.

Binding assay for the interaction of CLN8 with Sec24 proteins. The cytosolic tail of CLN8 (nucleotides 739–861 of CLN8’s open reading frame) with or without the WDK mutation was amplified from the CLN8-Y2 or CLN8-WDK-Y2 construct, respectively, and cloned in the pGEX-4T vector. Cells were selected for more than 1 week in geneticin-containing medium (G418, 200 µg ml−1) and incubated at 37 °C for 48 h.

Flow cytometry analysis. HeLa cells were plated in 24-well plates and, after 24 h, transfected with 200 ng YFP1-tagged and 200 ng YFP2-tagged constructs, in combination with 200 ng Ruby plasmid that was used as a reference for transfection efficiency. After 48 h, the fluorescence of 10,000 cells per sample was determined by flow cytometry using the BD LSRLFortessa Cell Analyzer (BD Biosciences) with the HTS autosampler device. Only Ruby-positive cells were considered for the subsequent analysis of reconstituted YFP signal quantification.

Generation of CLN8 knockout cells. CRISPR-Cas9 genome editing was used to introduce a deletion in the CLN8 gene in HeLa cells, specifically, to delete exon 2, which contains the translation start codon (Supplementary Fig. 5a). To this end, two complementary oligonucleotide couples (Supplementary Table 6), coding for a guide RNA upstream of a protospeck adapter motif site in exon 2 of CLN8, were designed using the online CRISPR design tool (http://crispr.mit.edu/). The two oligos were annealed and subsequently cloned into the px4sgr plasmid13, followed by Sanger sequencing of the insert to confirm the correct sequence. HeLa cells were transfected with 2 µg plasmid and split in a 96-well plate by single-cell deposition. After 3–4 days, DNA was isolated from the expanded single colonies and used in PCRs using oligos to amplify exon 2 of CLN8 with CloneAmp HiFi PCR Premix (Clontech) according to the manufacturer’s instructions. cDNAs inserted in vectors that were unable to produce an ampiclon were subsequently Sanger sequenced to confirm deletion of exon 2. CLN8−/− cells were analysed by quantitative real-time PCR (qRTPCR) to confirm the absence of CLN8 RNA expression and used for further experiments.

Nature Cell Biology | www.nature.com/naturecellbiology
study were males and age matched (2-month old). Mouse studies were approved by the Institutional Animal Care and Use Committee of Baylor College of Medicine (Houston, TX, USA) and are compliant with all relevant ethical regulations regarding animal research.

Subcellular fractionation. For each sample and replicate, a pool of three mouse livers was centrifuged in a discontinuous Nycodenz (Progen Biotechnik) density gradient as previously described\(^{17}\), with modifications. Briefly, tissues were homogenized in an assay buffer (20 mM sucrose, pH 7.2) and centrifuged first at 4,800g for 5 min and then at 17,000g for 10 min. The sediment of the second centrifugation was washed at 17,000g for 10 min, resuspended 1:1 v/v in 84.5% Nycodenz and placed on the bottom of an Ultraclear (Beckman) tube. On top, a discontinuous gradient of Nycodenz was constructed (the layers from bottom to top were: 32.8%, 26.3% and 19.8% Nycodenz). Samples were then centrifuged for 1h at an SW 40 rotor (Beckman) at 141,000g. Lysosome-enriched fractions were collected from the 26.3/19.8 interface, diluted in 5–10 volumes of assay buffer and centrifuged at 37,000g for 15 min. Pellets were resuspended in 500μl assay buffer.

LC–MS/MS. An aliquot of 2μl of each sample was quantified using NanoOrange protein quantitation kit (Invitrogen). All samples were normalized to 1μg per 100μl with 50 mM ammonium bicarbonate buffer (pH 7.9). The samples were reduced with 100μM dithiothreitol (Bio–Rad) in a rotary shaker at 800 rpm at room temperature for 30 min. The reduced cysteine residues were further treated with 400 mM iodoacetamide (Sigma) in the dark at room temperature for 30 min in the same solution. Sequences of the tetracycline-resistance gene 400 μM iodoacetamide was added to the digests. The digests were added with acetonitrile to a 10% final volume. The digestion was allowed to proceed at 37°C for 16h with mild shaking. The digestion was stopped by adding formic acid to the 5% final volume. The tubes were Speedvac dried. The digests were resuspended in 0.1% formic acid and 5% acetonitrile solution. The concentration of digest was measured using NanoOrange. Each sample was run in a triplicate on the Eksigent nanoLC system. The Eksigent nanoLC system and the ABICEXI TripleTOF 5600 mass spectrometer were controlled by the Analyst software version 1.6 (ABICEXI). The Eksigent nanoLC system had a cHPLC system. The chips contained a trap column (200μm × 0.5 mm, ChromXP C18-CL, 3 μm, 120 Å) that trapped the injected peptides in a flow rate of 3μl/min\(^{−1}\) for 50% 0.1% formic acid and 50% acetonitrile for 5 min at 23°C. The second chip was an analytical column (75μm × 15 cm, ChromXP C18-CL, 3 μm, 120 Å) with the organic mobile phase gradient set at a 90-min gradient starting from 5% to 35% acetonitrile in 0.1% formic acid in 90 min with a flow rate of 300 nll/min\(^{−1}\). The acetonitrile concentration was increased to 80% in 5 min. The 80% acetonitrile was transferred to new tubes. Cells were lysed on ice for 30 min in lysis buffer (50 mM Tris–HCl, pH 7.4, 150 mM NaCl, 0.5% NP-40, 1 mM phenylmethylsulfonyl fluoride, 1 mM NaF and 1X protease inhibitor cocktail) and centrifuged at 13,000g for 15 min. Pellets were resuspended in 500 μl assay buffer.

ERAD inhibition. For immunoblot experiments, the indicated cell lines were treated with 1μM MG132 and 1μM eycertatin I for 2h before protein extraction. For metabolic radiolabelling experiments, 10μM MG132 and 6μM eycertatin I were added during the starvation step and kept in the pulse and chase media.

Co-expression and sequence analysis. Expression correlation analysis was performed as previously described\(^{17}\), with minor modifications. Briefly, 620 genes encoding ER proteins were analysed by using the gSorter tool, which is part of the gOliGoSeq gene expression (GEO) data set. The analysis was performed using 104 heterogeneous microarray experiments, based on the HG-U133plus2 GeneChip array. Pairwise probe-to-probe correlation scores were computed as the cumulative occurrence of probes at the top 3% of most correlated gene probes in each data set. For statistical analysis, gene correlation profiles were compared using the Kolmogorov–Smirnov test. For graphical output, heatmaps were obtained by converting ranks into scores using the formula: score = 100 × (N_{top}−rank)/N_{top}\(^{−}\). CLN8 sequence analysis was performed using sequences retrieved from the UCSC Genome Browser (http://genome.ucsc.edu) after recursive BLAT searches in multiple species using available CLN8 protein sequences. The alignment of CLN8 sequences was performed with MULTALIN\(^{36}\). Local evolution rates of CLN8 amino acid sequences from mammals, birds, reptiles, amphibians and fishes were estimated using the evolution–structure–function method\(^{37,38}\). The accession numbers of the microarray data sets analysed for the co-expression analysis are provided in the Data availability statement.

Statistics and reproducibility. Sample size for in vivo and in vitro experiments was chosen based on previous studies to ensure adequate statistical power. All experiments using cells were performed at least in biological triplicate. Experimental analysis was performed in a blinded manner for quantification of the Pearson correlation in immunofluorescence analysis. For all other experiments, the analysis was performed in a blinded manner whenever possible. Randomization together with blinding was used for in vivo immunohistochemistry analysis. Pre-determined exclusion criteria were used in flow cytometry analysis by determining the efficiency of transfection by quantification of Ruby fluorescence. If efficiency of transfection was not adequate, the data were omitted from the study and sample analysis was repeated. Statistical significance was tested using two-sided t-test for simple comparisons, whereas group-level P values were estimated from the mean z-scores from each individual test. A value of P < 0.05 was considered significant. All experimental procedures were reviewed and approved by the Institutional Animal Care and Use Committee at Baylor College of Medicine.

Data availability

The authors declare that the main data supporting the findings of this study are available within the article and its Supplementary Information files. Source data for Figs. 1–5 and Supplementary Figs. 2–5 have been provided as Supplementary Table 1. All other data supporting the findings of this study are available from the corresponding author upon request. Mass spectrometry data have been deposited in ProteomeXchange with the primary access code PXD011066. The GEO accession numbers of the microarray data sets analysed for the co-expression analysis are the following: GDS1237, GDS1249, GDS1344, GDS1369, GDS1411, GDS1413, GDS1427, GDS1439, GDS1553, GDS1579, GDS1580, GDS1604, GDS1617, GDS1665, GDS1667, GDS1673, GDS1685, GDS1732, GDS1779, GDS1807, GDS1812, GDS1869, GDS1917, GDS1962, GDS1973, GDS1989, GDS2010, GDS2023, GDS2046, GDS2052, GDS2083, GDS2088, GDS2089, GDS2118, GDS2125, GDS2154, GDS2161, GDS2162, GDS2219, GDS2221, GDS2223, GDS2236, GDS2237, GDS2238, GDS2239, GDS2240, GDS2241, GDS2242, GDS2341, GDS2453, GDS2470, GDS2471, GDS2484, GDS2486, GDS2491, GDS2495, GDS2499, GDS2526, GDS2534, GDS2548, GDS2565, GDS2579, GDS2587, GDS2594, GDS2597, GDS2629, GDS2677, GDS2679, GDS2682, GDS2693, GDS2694, GDS2717, GDS2722, GDS2725, GDS2737, GDS2772, GDS2773, GDS2774, GDS2779, GDS2782, GDS2789, GDS2794, GDS2819, GDS2821, GDS2822, GDS2832, GDS2835, GDS2838, GDS2860, GDS2902, GDS2919, GDS2935, GDS2958, GDS2959, GDS3062, GDS3217, GDS3220, GDS3223 and GDS651.
References

30. Seehafer, S. S. & Pearce, D. A. Spectral properties and mechanisms that underlie autofluorescent accumulations in Batten disease. Biochem. Biophys. Res. Commun. 382, 247–251 (2009).

31. Marwaha, R. & Sharma, M. DQ-Red BSA trafficking assay in cultured cells to assess cargo delivery to lysosomes. Bio. Protoc. 7, e2571 (2017).

32. Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. Nat. Protoc. 8, 2281–2308 (2013).

33. Wattiaux, R., Wattiaux-De Coninck, S., Ronveaux-dupal, M. F. & Dubois, F. Isolation of rat liver lysosomes by isopycnic centrifugation in a metrizamide gradient. J. Cell Biol. 78, 349–368 (1978).

34. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl Acad. Sci. USA 102, 15545–15550 (2005).

35. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. 35, W193–W200 (2007).

36. Corpet, F. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 16, 10881–10890 (1988).

37. Simon, A. L., Stone, E. A. & Sidow, A. Inference of functional regions in proteins by quantification of evolutionary constraints. Proc. Natl Acad. Sci. USA 99, 2912–2917 (2002).

38. Spatuzza, C. et al. Physical and functional characterization of the genetic locus of IBtk, an inhibitor of Bruton’s tyrosine kinase: evidence for three protein isoforms of IBtk. Nucleic Acids Res. 36, 4402–4416 (2008).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
- Clearly defined error bars
- State explicitly what error bars represent (e.g. SD, SE, CI)

Software and code

Policy information about availability of computer code

Data collection: FACSDiva, Version 8.0.1
Data analysis: Microsoft Excel 2011, Version 14.7.7; GraphPad Prism, Version 6.01; ImageJ, Version 2.0.0-rc-43/1.51p (Fiji); Flow, Version X 10.0.7; Photoshop CS5; ProteinPilot, Version 4.5; GSEA, Version 2.0; Aminode, Version 1; Analyst, version 1.6 (ABCIEX Inc); g:Profiler, Version r1227_e72_eg19.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that the main data supporting the findings of this study are available within the article and its Supplementary Information files. Extra data are
available from the corresponding author upon request. Mass spectrometry data have been deposited in ProteomeXchange with the primary accession code PXD011066 [http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD011066].

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑️ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | Sample size for in-vivo and in-vitro experiments was chosen based on previous studies in order to ensure adequate statistical power. All findings were reliably reproduced. |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data exclusions | Experiments failed for technical reasons were excluded from data analysis. |
| Replication | All experiments using mice and cells were performed at least in biological triplicate. For mouse proteomics analysis, two replicates of three pooled mice for each group were used. For staining overlap, Pearson correlation (range -1 to +1) was calculated on 10 independent images per experiment. All attempts at replication were successful for all experiments |
| Randomization | Randomization together with blinding was used for in-vivo IHC analysis. |
| Blinding | For all other experiments, the analysis was performed in a blinded fashion whenever possible. |

Reporting for specific materials, systems and methods

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| n/a | n/a |

| Involved in the study | n/a |
|-----------------------|-----|
| ☑️ ☑️ Unique biological materials | |
| ☑️ Antibodies | |
| ☑️ ☑️ Eukaryotic cell lines | |
| ☑️ ☑️ Palaeontology | |
| ☑️ ☑️ Animals and other organisms | |
| ☑️ Human research participants | |

| Involved in the study | n/a |
|-----------------------|-----|
| ☐ ChIP-seq | |
| ☐ Flow cytometry | |
| ☑ MRI-based neuroimaging | |

Unique biological materials

Policy information about availability of materials

Obtaining unique materials

All unique materials used are readily available from the authors.

Antibodies

Antibodies used

Primary Antibodies:
- anti-C-myc Sigma-Aldrich C3956-2MG Rabbit 1:1000
- anti-CTSB (CA10) Millipore IM27L-100UG Mouse 1:1000
- anti-CTSD (R-20) Santacruz sc-6487 goat 1:1000
- anti-CTSD (CTD-19) Abcam ab6313 Mouse 1:1000
- anti-CTSF Abcam ab36161 goat 1:1000
- anti-Flag Sigma F7425 Rabbit 1:1000
- anti-Flag (M2) Sigma F3165 Mouse 1:1000
- anti-GALNS Abcam ab187516 goat 1:1000
- anti-GAPDH (6C5) Santacruz sc-32233 Rabbit 1:1000
- anti-GFP Abcam ab13970 Chicken 1:1000
- anti-GFP Cell Signaling 29565 Rabbit 1:1000
- anti-GM130 Abcam ab526449 Rabbit 1:200
Validation

All the antibodies used in the manuscript were bought from commercial companies and are widely used for similar experiments by other researchers worldwide. The utility of these antibodies is stated on the websites (Certificates of analysis) of the corresponding suppliers. Relevant citations are provided on the websites of the corresponding antibody. The anti-KDEL antibody (Rabbit) was a generous gift from Dr. Victor Hsu and was used as in the referenced paper.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) HeLa, HEK293, HEK293T, Skin Biopsied fibroblasts, Rosetta (DE3) cells

Authentication Morphological evaluation of the cells was used as authentication method.

Mycoplasma contamination All cells used in this study were not contaminated. The assessment of the contamination was performed by using LookOut® Mycoplasma PCR Detection Kit (Sigma).

Commonly misidentified lines No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals The CLN8mnd mouse line was purchased from the Jackson Laboratory (species: C57/B6, strain: B6.KB2-Cln8mnd/MsrJ). All mice used in this study were males and age-matched (2-month old).

Wild animals The study did not involve wild animals

Field-collected samples The study did not involve samples collected in the field

Human research participants

Policy information about studies involving human research participants

Population characteristics Skin biopsied fibroblasts were collected from two healthy subjects and from two CLN8 patients (genotype Q194R/66delG and A30P/A30P).

Recruitment Recruitment was performed upon receiving informed written consent from the patients or their legal guardians. The study protocol was approved by the ethics committees of the IRCCS Stella Maris, Pisa. All the procedures complied with the requirements of the Declaration of Helsinki.
Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation
HeLa cells were plated in 24-well plates and, after 24 h, transfected with 200 ng YFP1- and 200 ng YFP2-tagged constructs, in combination with 200 ng of Ruby plasmid that was used as a reference for transfection efficiency. After 48 h, the fluorescence of 10,000 cells per sample was determined by flow cytometry using the BD LSRFortessa™ Cell Analyzer (BD Biosciences) with the HTS autosampler device. Only Ruby-positive cells were considered for the subsequent analysis of reconstituted YFP signal quantification.

Instrument
BD LSFRFortessa™ Cell Analyzer (BD Biosciences) with the HTS autosampler device.

Software
FACSDiva, Version 8.0.1; FloJo, Version X 10.0.7r2

Cell population abundance
No flow based sorting was performed.

Gating strategy
Appropriate control samples were used to gate the samples. These were: untransfected HeLa cells, Ruby transfected HeLa cells, CLN8Y1/Y1 co-transfected HeLa cells, CLN8Y1/Y2 co-transfected HeLa cells, Ruby/CLN8Y1/Y1 co-transfected HeLa cells and Ruby/CLN8Y1/Y2 co-transfected HeLa cells. Cells were selected in the FSC-A/SSC-A dot plot to remove debris and they were gated to exclude cellular aggregates in the FSC-H/FSC-W and SSC-H/SSC-W dot plot. Positive cells were gated with PE+ and FITC+ lasers and Ruby/CLN8Y1/Y1 co-transfected HeLa cells were used as control sample.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.