Supplementary Information for

Lysosomal cathepsin D mediates endogenous mucin glycodomain catabolism in mammals

Kayvon Pedram, Nouf N. Laqtom, D. Judy Shon, Alessandro Di Spiezio, Nicholas M. Riley, Paul Saftig, Monther Abu-Remaileh, Carolyn R. Bertozzi*

*Correspondence should be addressed to C.R.B., email: bertozzi@stanford.edu

This PDF file includes:

Other supplementary materials for this manuscript include the following:

Supplementary Methods

MUC16 purification

The purification was adapted from the procedure by Wong *et al*.[1] OVCAR-3 cells (ATCC) were cultured at 37 ºC, 5% CO2 in RPMI-1640 (no phenol red) supplemented with 20% fetal bovine serum (FBS) and 1% penicillin/streptomycin until confluent. The culture medium was removed, and the cells were washed twice with serum-free RPMI-1640 (no phenol red). The cells were cultured for 5 days in this medium, after which the medium was removed and spun down at 300 x g for 5 mins. The clarified supernatant was stored at -80 ºC. This process was repeated multiple times. Clarified OVCAR-3 supernatant (250 mL) containing MUC16 was concentrated to less than 5 mL using Amicon Ultra 100,000 NMWL filters (Millipore Sigma UFC910024) and filtered through a 5.0 µm Acrodisc syringe filter (VWR). MUC16 was purified by size exclusion chromatography using a HiPrep Sephacryl 16/60 S-500 HR column (Cytiva Life Sciences) in PBS, pH 7.4. Purified MUC16 concentration in U/mL was measured using a Human CA125 ELISA Kit (Abcam) according to manufacturer recommendations. Fractions containing MUC16 in PBS were stored at -80 ºC.

MUC16 labeling

IRdye® 800CW or IRDye® 680RD NHS Ester (LI-COR Biosciences) was dissolved at 10 mM in DMSO. IRdye® 800CW or IRDye® 680RD NHS Ester (50 µM) was incubated with MUC16 (50,000 U) in PBS supplemented with 100 mM K_2HPO_4 for 2 hours at room temperature, rotating end-over-end in the dark. Free dye was removed using 40K 0.5 mL Zeba Spin Desalting Columns (Thermo Fisher Scientific). Labeled MUC16 was then concentrated to 100,000 U/mL, assuming no loss of material. Aliquots were stored at -20 ºC.

Biodistribution of MUC16-680

Animal care was monitored by the Veterinary Service Center at Stanford University and procedures were approved under APLAC protocol no. 31511. 10-week old NU/J mice (Jackson) were injected retro-orbitally with 2000 U MUC16-680 in 100 µl sterile PBS. Live mice were anesthetized with isoflurane and imaged at various timepoints using an IVIS Lumina instrument. At 4.5 hours, one mouse was sacrificed and organs were re-imaged by IVIS *ex vivo*.

MUC2 purification

The MUC2 purification protocol was adapted from Recktenwald and Hansson[2]. Briefly, LS174T cells (ATCC) were maintained at 37 ºC, 5% CO_2 in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin. At confluency, cells were washed once with 5 mL of PBS, resuspended in at least 5x the volume of cold 6 M guanidinium hydrochloride (GuHCl), 5 mM EDTA, 10 mM NaH_2PO_4, pH 6.5, and solubilized by rotating at 4 ºC for 24 h. Insoluble mucins were precipitated by centrifugation at 18,000 rpm for 30 min at 4 ºC. The wash step was repeated 5 times in the same GuHCl buffer with incubation times of at least 3 h. The pellet was solubilized in 6 M GuHCl, 100 mM Tris, 5 mM EDTA, 25 mM DTT, pH 8.0 for 5 h at 37 ºC. Iodoacetamide was added to 62.5 mM and the solution was rotated overnight at room temperature in the dark. The sample was spun down at 10,000 rpm for 30 min at 4 ºC to remove insoluble material, supernatants were dialyzed against water for 36 h using 10K MWCO Slide-A-Lyzer Dialysis Cassettes, and aliquots were stored at -80 ºC.

MUC2 labeling

IRDye® 800CW NHS Ester (LI-COR Biosciences) was dissolved at 10 mM in DMSO. 1 mM of IRDye® 800CW NHS Ester was incubated with 2 mg of MUC2 (measured via lyophilization) in PBS supplemented with 100 mM K_2HPO_4 for 3 h at room temperature, rotating end-over-end in the dark. Free dye was removed via dialysis against water using a 100KD Spectra/Por Float-A-Lyzer G2 Dialysis Device (Fisher Scientific) and aliquots were stored at -20 ºC.

Mass spectrometry of tissue fractions

Sample preparation

Size exclusion chromatography fractions from Balb/cJ liver and brain lysate prepared as described above were brought to a total volume of 25 µl with 100 mM tris, pH 8. Total protein in each fraction was in the range of 0.1 to 10 µg. Dithiothreitol was added to a final concentration of 5 mM and the tubes were incubated at 55 °C for 15 mins. Samples were allowed to return to room temperature, then iodoacetamide was added a final concentration of 15 mM and allowed to react in the dark at room temperature for 30 mins. Digestion was completed by adding 0.8 µg trypsin (Promega) for 8 hours at 37 °C and was quenched by dilution with 500 µl 0.2% formic acid. Peptides were desalted on a 10 mg Strata-X cartridge (Phenomenex) by conditioning the cartridge with 1 mL ACN followed by 1 mL 0.2% formic acid (FA) in water. Acidified peptides were loaded onto the cartridge, followed by a 1 mL wash with 0.2% FA in water. Peptides were eluted with 400 µl of 0.2% FA in 80% ACN, dried via lyophilization, then resuspended in 10 µl of 0.2% FA.

LC-MS/MS

Equal volumes of each sample (1 µl) were injected on column. Peptides were separated over a 50 cm EasySpray reversed phase LC column (75 µm inner diameter packed with 2 µm, 100 Å, PepMap C18 particles, Thermo Fisher Scientific). The mobile phases (A: water with 0.2% formic acid and B: acetonitrile with 0.2% formic acid) were driven and controlled by a Dionex Ultimate 3000 RPLC nano system (Thermo Fisher Scientific). An integrated loading pump was used to load peptides onto a trap column (Acclaim PepMap 100 C18, 5 um particles, 20 mm length, Thermo Fisher Scientific) at 5 µl/min, which was put in line with the analytical column 6 mins into the gradient. Gradient elution was performed at 300 nL/min. The gradient increased from 0% to 5% B over the first 6 mins of the analysis, followed by an increase from 5% to 25% B from 6 to 86 mins, an increase from 25% to 90% B from 86 to 94 mins, isocratic flow at 90% B from 94 to 102 mins, and a re-equilibration at 0% for 18 mins for a total analysis time of 120 mins. Precursors were ionized using an EASY-Spray ionization source (Thermo Fisher Scientific) held at +2.2 kV compared to ground, and the column was held at 45 °C. The inlet capillary temperature on the quadrupole-Orbitrap-linear ion trap Tribrid mass spectrometer (Orbitrap Fusion, Thermo Fisher Scientific) used for analysis was held at 275 °C. Survey scans of peptide precursors were collected in the Orbitrap from 350-1350 m/z with an AGC target of 1,000,000, a maximum injection time of 50 ms, and a resolution of 60,000 at 200 m/z. Monoisotopic precursor selection was enabled for peptide isotopic distributions, precursors of z = 2-5 were selected for data-dependent MS/MS scans for 2 seconds of cycle time, and dynamic exclusion was set to 45 seconds with a ±10 ppm window set around the precursor monoisotope. An isolation window of 1 m/z was used to select precursor ions with the quadrupole. MS/MS scans were collected using HCD at 30 normalized collision energy (NCE) with an AGC target of 50,000 and a maximum injection time of 54 ms. Mass analysis was performed in the Orbitrap with a resolution of 30,000 at 200 m/z and an automatically determined mass range.

Proteomics data analysis

Raw data were processed using MaxQuant[3] version 1.6.10.43 and tandem mass spectra were searched with the Andromeda search algorithm[4]. Oxidation of methionine and protein N-terminal acetylation were specified as variable modifications, while carbamidomethylation of cysteine was set as a fixed modification. 20 ppm, 4.5 ppm, and 20 ppm were used for first search MS1 tolerance, main search MS1 tolerance, and MS2 product ion tolerance, respectively. Cleavage specificity was set to Trypsin/P with 2 missed cleavage allowed. Peptide spectral matches (PSMs) were made against a target-decoy mouse reference proteome database downloaded from Uniprot (17,030 entries). Peptides were filtered to a 1% false discovery rate (FDR), two peptides were required for a protein identification, and a 1% protein FDR was applied. Proteins were quantified and normalized using MaxLFQ[5] with a label-free quantification (LFQ) minimum ratio count of 2. The match between runs feature was enabled.

Cathepsin knockouts in 8988T cells

Human CTSB, CTSD, and CTSL were depleted using the pLentiCRISPRv1 system. The following sense (S) and antisense (AS) oligo-nucleotides were cloned into pLentiCRISPRv1:

Primer	*Sequence*
sgCTSD4 (S)	caccgGGAGAGGCAGGTCTTTGGGG
sgCTSD4 (AS)	aaacCCCCAAAGACCTGCCTCTCCc
sgCTSD7 (S)	caccgCGTCTCAAAGTACTCCCAGG
sgCTSD7 (AS)	aaacCCTGGGAGTACTTTGAGACGc
sgCTSB3 (S)	caccgGTGGTGCTCACAGGGAGGGA
sgCTSB3 (AS)	aaacTCCCTCCCTGTGAGCACCACc
sgCTSB7 (S)	caccgGTCACCGGAGAGATGATGGG
sgCTSB7 (AS)	aaacCCCATCATCTCTCCGGTGACc
sgCTSL10 (S)	caccgAAGGCAGCAAGGATGAGTGT
sgCTSL10 (AS)	aaacACACTCATCCTTGCTGCCTTc
sgAAVS (S)	caccgTCCCCTCCACCCCACAGTG
sgAAVS (AS)	aaacCACTGTGGGGTGGAGGGGAc

Lentiviruses were produced by transfecting HEK-293T cells with one of the pLentiCRISPRv1 plasmids, in combination with the packaging plasmids, VSV-G envelope and the ΔVPR. The culture medium was changed to DMEM supplemented with 30% inactivated fetal calf serum 16 hours post transfection. The virus-containing supernatant was collected 48 hours post transfection and spun for 5 mins at 400 x g to remove cells and then frozen at -80 °C. To generate knockout cells, 8988T cells were transduced as described below and after puromycin selection, cells were single-cell sorted into 96-well plates containing 200 µl of DMEM supplemented with 30% inactivated fetal calf serum. Cell clones with the desired knockouts were identified by western blot. Control cells were generated by targeting the non-coding AAVS1 locus as described previously[6].

Rescue of 8988T cells with cathepsin D lentivirus

Q5 Site-Directed Mutagenesis (New England Biolabs) was used to mutate the guide PAM sequences in pMXs-CTSD-FLAG using primers listed below.

Primer	*Sequence*

sg4_PAM_for	TCTTTGGGGAAGCCACCAAGC
sg4_PAM_rev	CCTGCCTCTCCACTTTGAC
sg7_PAM_for	TACTCCCAGGCAGTGCCAGCCGT
sg7_PAM_rev	CTTTGAGACGGGGCCTTTGGC

HEK-293Ts ($1.5x10^6$ cells) were seeded into 6 cm dishes in 5 mL of DMEM supplemented with 10% heat-inactivated fetal bovine serum (HI FBS) and 1% penicillin/streptomycin (P/S). 24 hours later, the plasmid of interest (1 µg) was mixed with retrovirus pol/gag (900 ng), VSVg DNA (150 ng), DMEM (130 µl), and polyethylenimine (PEI) at 1 mg/mL (6 µl). The mixture was incubated for 20 mins at room temperature and added to HEK-293T cells dropwise. 18 hours later, the culture media was replaced with 5 mL of DMEM supplemented with 30% HI FBS and 1% P/S. 30 hours later, the media was collected and centrifuged at 1000 rpm for 5 mins. The clarified supernatant was stored at -80 ºC prior to infection. To establish stably expressing cell lines, 1.5 million cells were seeded in 6-well plates in 2.75 mL DMEM with 10% HI FBS and 1% penicillin/streptomycin. Polybrene was added at 10 µg/mL and cells were infected with 250 µl of virus-containing media. Cells were lifted with trypsin and plated in 10 cm dishes with 10 µg/mL blasticidin S (Thermo Fisher Scientific) 16 hours later. Stable cell lines were tested for expression of fully processed lysosomal CTSD by western blot, as described above.

O-glycoproteomics and cleavage specificity

Sample Preparation

Bovine serum albumin (BSA) was purchased from Sigma-Aldrich (A7906-1KG). Fetuin was purchased from Promega (V4961). Recombinantly expressed MUC16, podocalyxin, CD43, and PSGL-1 were purchased from R&D Systems (5609-MU, 1658-PD, 9680-CD, 3345-PS, respectively). Each protein was digested with cathepsin D (CTSD) or trypsin (Tryp) separately, or a sequential digestion was performed with CTSD proteolysis followed by trypsin. These samples were digested in solution, similar to previous methods to investigate glycoproteases[7–10]. Cathepsin D digestions were set up with 31.2 µl 100 mM NaOAc, pH 5, 3 µl of 1 mg/mL protein substrate in water, and 1.8 µl of 10 µM cathepsin D purified from human liver (Molecular Innovations, HCD) in PBS, then incubated at 37 °C overnight. One third (12 µl) of the reactions was removed and quenched via addition of 4 µl NuPAGE™ LDS Sample Buffer (4X) (Thermo Fisher Scientific) supplemented with 250 mM DTT. These samples were boiled for 5 mins at 95 °C then loaded along with minus cathepsin D control samples into an 18-well 4-12% Criterion™ XT Bis-Tris gels (Bio-Rad) and run in MOPS. Total protein was detected using AcquaStain Coomassie stain (Bulldog Bio) and visualized with a Li-Cor Odyssey Blot Imager. The remaining two-thirds of the cathepsin D reactions (24 µl) was diluted to 99 µl with 100 mM tris, pH 8, then de-N-glycosylated via addition of 1 µl PNGase-F (NEB, diluted to 5000 U/mL in 100 mM tris, pH 8) and incubated at room temperature for 8 hours at 37 °C. Dithiothreitol was added to a final concentration of 5 mM and the tubes were incubated at 55 °C for 15 mins. Samples were allowed to return to room temperature, then iodoacetamide was added to a final concentration of 15 mM and allowed to react in the dark at room temperature for 30 mins. One half of the reaction was removed and trypsin (Promega) was added at a 20:1 protein:protease ratio for 8 hours at 37 °C. All samples were quenched by dilution with 500 µl 0.2% formic acid. Peptides were desalted on a 10 mg Strata-X cartridge (Phenomenex) by conditioning the cartridge with 1 mL ACN followed by 1 mL 0.2% formic acid (FA) in water. Acidified peptides were loaded onto the cartridge, followed by a 1 mL wash with 0.2% FA in water. Peptides were eluted with 400 µl of 0.2% FA in 80% ACN, dried via lyophilization, then resuspended in 10 µl of 0.2% FA.

LC-MS/MS

Loading and gradient elution were performed using the same conditions as described above, and an Orbitrap Fusion was used for MS analysis. Again, precursors were ionized using an EASY-Spray ionization source (Thermo Fisher Scientific) held at +2.2 kV compared to ground, the column was held at 40 °C, and the inlet capillary temperature was held at 275 °C. Survey scans of peptide precursors were collected in the Orbitrap from 500-1800 m/z with an AGC target of 400,000, a maximum injection time of 50 ms, RF lens at 60%, and a resolution of 120,000 at 200 m/z. Monoisotopic precursor selection was enabled for peptide isotopic distributions, precursors of z = 2-10 were selected for data-dependent MS/MS scans for 3 seconds of cycle time, and dynamic exclusion was set to 45 seconds with a ±10 ppm window set around the precursor monoisotope. Precursor priorities were set to favor highest charge state and lowest m/z precursor ions, and an isolation window of 2 m/z was used to select precursor ions with the quadrupole.

All methods used were product-dependent, where the presence of at least two out of nine oxonium ions (126.055, 138.0549, 144.0655, 168.0654, 186.076, 204.0865, 274.0921, 292.1027, and 366.1395) in a "scouting" higher-energy collisional dissociation (HCD) MS/MS scan triggered acquisition of a second MS/MS scan[11–13], which was an EThcD scan in these experiments. Requiring at least two oxonium ions can help improve the specificity of triggering on glycopeptide precursor ions. The "scout HCD" scan had an automated scan range determination and a first mass of 100 Th, an NCE of 36, an AGG target value of 50,000, a maximum injection time of 60 ms, and a resolution of 30,000 at 200 m/z. If at least two of the nine listed oxonium ions were present in the scout HCD scan within a ±10 ppm tolerance and were among the 20 most intense peaks, a second MS/MS scan was triggered. All triggered ETHcD MS/MS scans used a fixed scan range of 120-4,000 m/z (this extended mass range has been shown to improve glycopeptide characterization[14]), an AGC target of 100,000 ions, a maximum injection time of 200 ms, and a resolution of 30,000 at 200 m/z. Calibrated charge dependent parameters for calculating reagent AGC targets and ion-ion reaction times[15] were enabled for EThcD scans, and supplemental activation of 25 NCE was used.

Data analysis

All raw data were searched using O-Pair Search implemented in MetaMorpheus (0.0.308), which is available at https://github.com/smith-chem-wisc/MetaMorpheus.[16] All files were searched in one batch so that the FDR calculations were performed on the entire dataset. The "Glyco Search" option was selected, where the O-glycopeptide search feature was enabled and the Oglycan.gdb glycan database was selected, representing 12 common O-glycans: HexNAc(1), HexNAc(1)Hex(1), HexNAc(1)NeuAc(1), HexNAc(2)Hex(1), HexNAc(1)Hex(1)NeuAc(1), HexNAc(1)Hex(1)NeuAc(2), HexNAc(2)Hex(2) NeuAc(1), HexNAc(2)Hex(2)NeuAc(2), HexNAc(2)Hex(2), HexNAc(2)Hex(1)NeuAc(1), HexNAc(2)Hex(2)Fuc(1)NeuAc(1), HexNAc(2)Hex(2)Fuc(1)NeuAc(2). The "Keep top N candidates" feature was set to 50, and Data Type was set as HCD with Child Scan Dissociation set as EThcD. The "Maximum OGlycan Allowed" setting was set to 5, where this number represents both the maximum number of O-glycan modifications that could occur on a glycopeptide candidate and the number of times each O-glycan could occur per peptide. Under Search Parameters, both "Use Provided Precursor" and "Deconvolute Precursors" were checked. Peak trimming was not enabled and Top N peaks and minimum ratio were set to 1000 and 0.01, respectively. In-Silico Digestion Parameters were set to generate decoy proteins using reversed sequences, and the initiator methionine feature was set to "Variable". The maximum modification isoforms allowed was 1024, and non-specific digestion was enabled for peptides ranging from 5 to 25 residues. Precursor and product mass tolerances were 10 and 20

ppm, respectively, and the minimum score allowed was 3. Modifications were set as Carbamidomethyl on C as fixed, and Oxidation on M and Deamidation on N as variable. Note, O-Pair Search returns a single identification representing two spectra, both an HCD and EThcD spectrum. Identifications are made using the HCD spectrum, and the associated EThcD spectrum is used to localize O-glycosites.

The oglyco.psmtsv and single.psmtsv results file was used for all data processing as the O-glycopeptide and standard peptide identifications, respectively. For both O-glycopeptide and standard peptides, identifications were filtered to include only target matches (T) and identifications with a q-value < 0.01. O-glycopeptide identifications were filtered to include only Level 1 or Level 1b identifications, as described in the O-Pair Search reference[16], which include only identifications with confident and unambiguous O-glycosite localization. To get cleavage motifs, identifications that were in common between the "trypsin only" files were removed to retain identifications that resulted only from CTSD-derived cleavage. Note, this means that some semi-tryptic peptides remain from the sequential digestion data, where the non-tryptic terminus is attributed to CTSD cleavage. The percent of O-glycosylated serine and threonine residues was determined by counting the number of modified residues at a given position relative to the total number of residues. Serine and threonine data were summed, so this is an aggregate value for both residues. Calculations were performed in Microsoft Excel or in Visual Studio 2017 using C#, and data was graphed using OriginPro 2018. WebLogo was used for Logo plot generation (https://weblogo.berkeley.edu/logo.cgi)[17]. To make this plot, residues +5 and -5 residues from the cleavage site were used. Spectral annotation was aided by the Interactive Peptide Spectral Annotator (IPSA, http://www.interactivepeptide-spectralannotator.com)[18]. Raw data files and result files from in this study have been deposited to the ProteomeXchange via the PRIDE[19] partner repository with the dataset identifier PXD027616 (access with username "reviewer_pxd027616@ebi.ac.uk" and password "TiP4qF2r").

StcE^{E447D} labeling

EZ-Link sulfo-NHS-biotin (Thermo Fisher Scientific) was dissolved at 20 mM in ddH_2O and incubated with StcE^{E447D} at a 10:1 molar ratio in PBS for 2 hours at room temperature, rotating end-over-end. Free biotin was removed using 7K 5 mL Zeba Spin Desalting Columns. Biotin labeling was quantified using the Pierce Biotin Quantitation Kit (Thermo Fisher Scientific) and aliquots were stored at -20 °C. Alexa Fluor 647 (AF647) NHS Ester (Thermo Fisher Scientific) was dissolved at 10 mM in DMSO and incubated with StcE^{E447D} at a 10:1 molar ratio in PBS supplemented with 100 mM K_2HPO_4 for 1.5 hours at room temperature, rotating end-over-end in the dark. Free dye was removed using 7K 5 mL Zeba Spin Desalting Columns and aliquots were stored at -20 °C.

Quantification of mucin aggregates

Hematoxylin and DAB staining was separated by color deconvolution in Fiji. The DAB channel was altered to grayscale and aggregates were selected via the Difference of Gaussian approach. For liver mucin aggregates, the settings were as follows: minimum Gaussian kernel standard deviation (min_sigma) – 20; maximum Gaussian kernel standard deviation (max_sigma) – 30; threshold – 1.4. For spleen mucin aggregates, the settings were as follows: min_sigma – 10; max_sigma – 30; threshold – 0.2.

Confocal microscopy

Alexa Fluor® 488 anti-mouse CD107a (LAMP-1) [1D4B] (BioLegend), Alexa Fluor® 594 anti-mouse F4/80 [BM8] (BioLegend), MUC1 Recombinant Rabbit Monoclonal Antibody (ARC0352) (Thermo Fisher Scientific #MA5-35250), Endomucin Monoclonal Antibody (eBioV.7C7 (V.7C7)), eBioscience™ (Thermo Fisher Scientific #14-5851-82), and CA125 Polyclonal Antibody (Thermo Fisher Scientific #PA5-87558) were used according to manufacturer recommendations. Tissue slides were dried in an oven at 62 °C for 1 hour. Slides were dewaxed in xylene (3 × 5 mins) and then hydrated in a 100%, 95%, 75% ethanol gradient (2 × 3 mins each). Antigen retrieval was performed in Antigen Unmasking Solution, Citric Acid Based (Vector Laboratories) for 15 mins. Tissues were permeabilized with 0.2% Triton X-100 (Sigma-Aldrich) in PBS for 10 mins at room temperature. Slides were washed in PBS-T (0.05% Tween-20) (2 x 5 mins) and incubated in Carbo-Free Blocking Solution (Vector Laboratories) for 1 hour at room temperature. AF647-StcE^{E447D} (5 µg/mL) was applied to slides in PBS at room temperature for 30 mins. Following two washes with PBS-T (2 x 5 mins), tissues were incubated overnight with fluorophore-conjugated antibodies at 4 ºC in the dark. Slides were washed three times with PBS-T (3 x 5 mins) and mounted with ProLong Glass Antifade Mountant with NucBlue Stain (Thermo Fisher Scientific) at room temperature prior to imaging with a Nikon A1R confocal microscope.

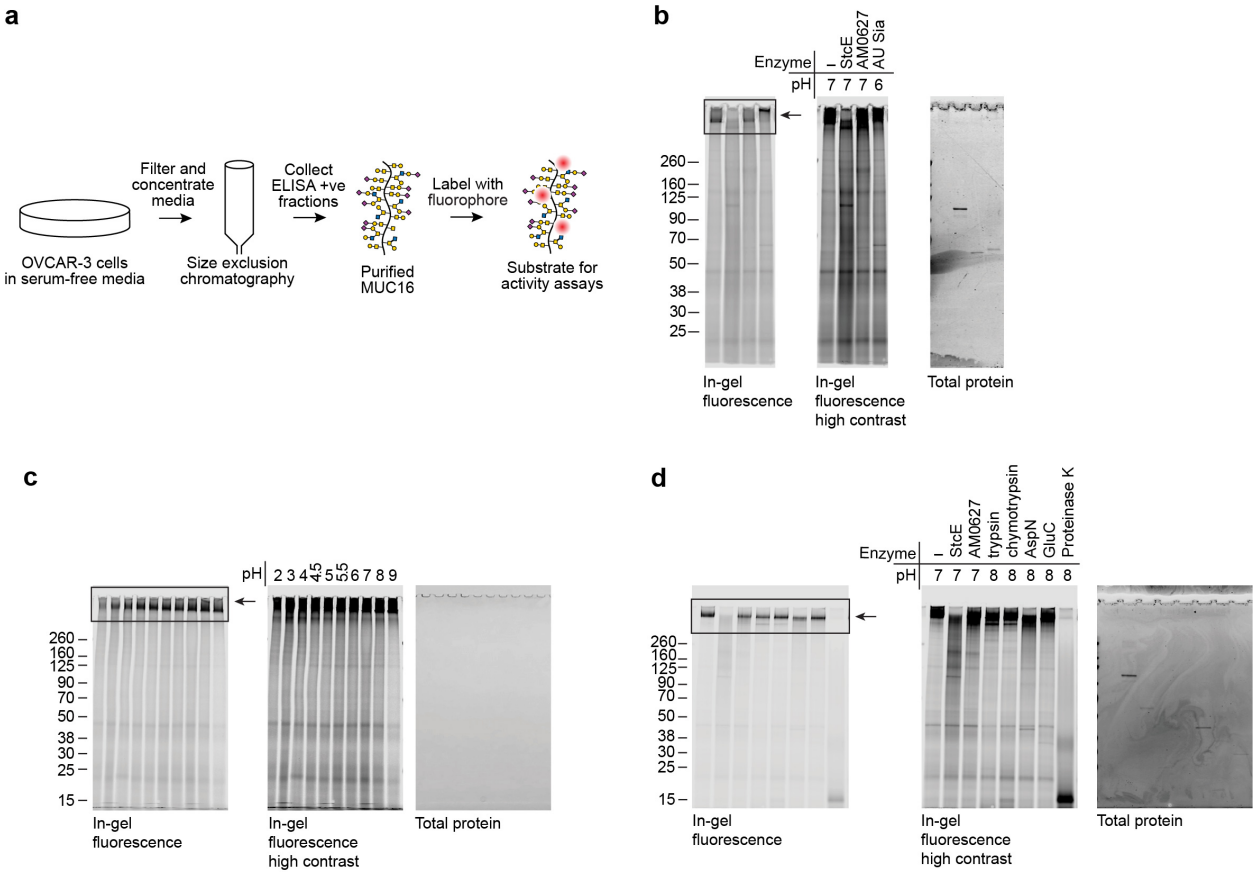


Figure S1. MUC16 purified from OVCAR-3 cells is sialylated, stable in buffers of various pH, and resistant to proteolysis.
(a) Schematic of MUC16 purification and fluorophore conjugation.
(b) MUC16-800 was incubated overnight at 37 °C with 50 nM buffer alone, the bacterial mucinase StcE, the bacterial mucinase AM0627, or the *Arthrobacter ureafaciens* (AU) sialidase. Reaction pH was chosen based on optimal activity of each enzyme. Observed gel shifts in the high molecular weight region of the gel are consistent with a sialylated mucin structure[20].
(c) MUC16-800 was incubated alone in buffers of the indicated pH overnight at 37 °C. Note that mucins such as MUC16 do not stain readily with Coomassie (total protein). The absence of Coomassie-positive bands highlights the purity of the preparation.
(d) MUC16-800 was incubated overnight at 37 °C with 50 nM of the bacterial mucinase StcE, the bacterial mucinase AM0627, trypsin, chymotrypsin, AspN, GluC, or proteinase K. Reaction pH was chosen based on optimal activity of each enzyme.

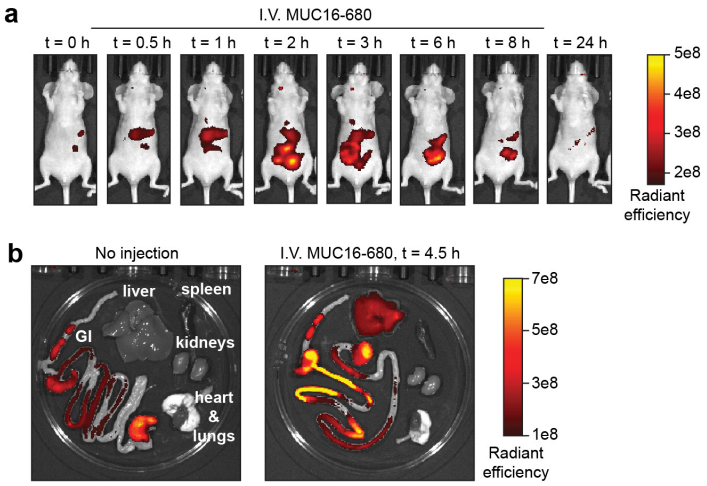
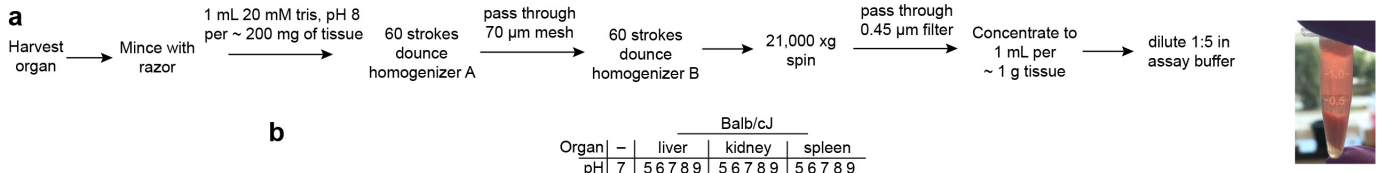
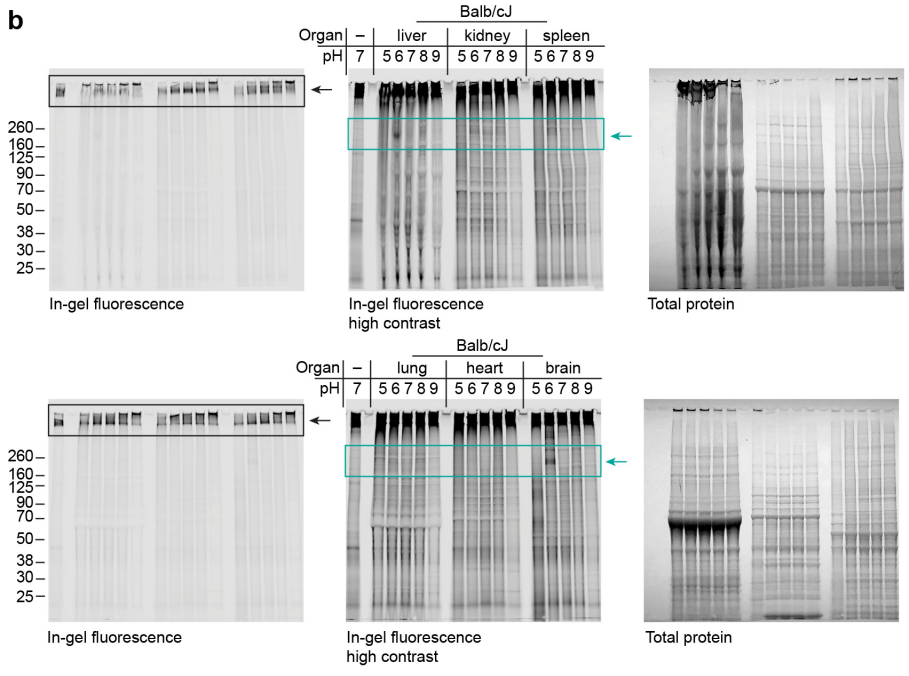
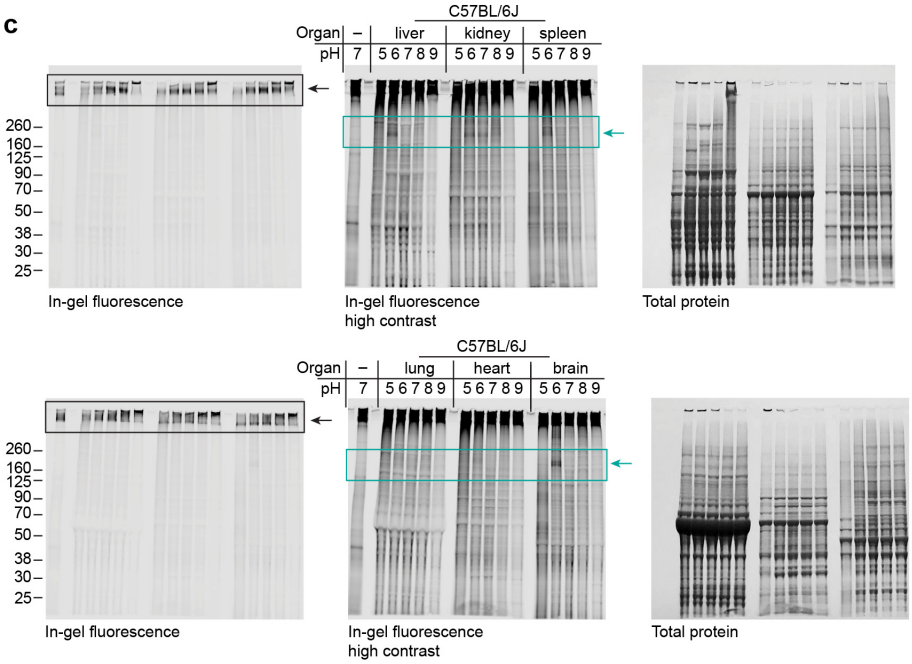


Figure S2. Fluorophore labeled MUC16 localizes to liver and gastrointestinal tissues at 4.5 hours post retro-orbital injection.
(a) Purified, fluorophore-labeled MUC16 in PBS (MUC16-680) was injected at 2000 U retro-orbitally into a 10-week old NU/J mouse and fluorescence images were acquired with an IVIS Illumina instrument at various time points.
(b) A 10-week old NU/J mice was injected with MUC16-680 as in (a). At 4.5 hours, the mouse was sacrificed. Organs were removed and imaged using an IVIS Illumina instrument.


a
Harvest organ
Mince with razor
1 mL 20 mM tris, pH 8 per ~ 200 mg of tissue
60 strokes dounce homogenizer A
pass through 70 µm mesh
60 strokes dounce homogenizer B
21,000 xg spin
pass through 0.45 µm filter
Concentrate to 1 mL per ~ 1 g tissue
dilute 1:5 in assay buffer


b
Balb/cJ
Organ – liver kidney spleen
pH 7 5 6 7 8 9 5 6 7 8 9 5 6 7 8 9
260
160
125
90
70
50
38
30
25
In-gel fluorescence
In-gel fluorescence high contrast
Total protein
Balb/cJ
Organ – lung heart brain
pH 7 5 6 7 8 9 5 6 7 8 9 5 6 7 8 9
In-gel fluorescence
In-gel fluorescence high contrast
Total protein


c
C57BL/6J
Organ – liver kidney spleen
pH 7 5 6 7 8 9 5 6 7 8 9 5 6 7 8 9
260
160
125
90
70
50
38
30
25
In-gel fluorescence
In-gel fluorescence high contrast
Total protein
C57BL/6J
Organ – lung heart brain
pH 7 5 6 7 8 9 5 6 7 8 9 5 6 7 8 9
In-gel fluorescence
In-gel fluorescence high contrast
Total protein

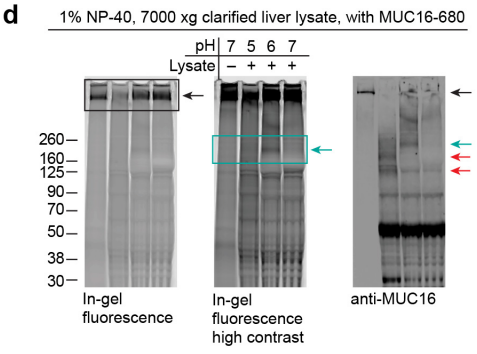

d
1% NP-40, 7000 xg clarified liver lysate, with MUC16-680
pH 7 5 6 7
Lysate – + + +
260
160
125
90
70
50
38
30
In-gel fluorescence
In-gel fluorescence high contrast
anti-MUC16

Figure S3. MUC16-degrading activity is observed across tissues in Balb/c and C57BL/6 mice.
(a) *Left*, Schematic of lysis protocol. Liver, kidney, spleen, lung, heart, and brain were dissected from a Balb/c mouse post sacrifice. Tissues were Dounce homogenized in 20 mM tris, pH 8, clarified at 21,000 x g, then filtered and concentrated to ~10 mg/mL. *Right*, image of liver lysate following 21,000 x g centrifugation showing clarified material.
(b) Balb/cJ liver, kidney, spleen, lung, heart, and brain lysate prepared as in (a) and 2 µl was mixed with 1 µl of 100,000 U/mL MUC16-680 and 9 µl of buffer of the indicated pH supplemented with common biological ions (see *Methods*). Reactions were allowed to proceed overnight at 37 °C. Proteins were separated by SDS-PAGE and imaged by in-gel fluorescence. Coomassie was used for total protein stain. Black boxes highlight the substrate band; cyan boxes highlight appearance of a product band between 160 and 260 kDa at pH 6. The expected product band between 8 and 15 kDa at pH 5 is not visible in these gels, which were run for a longer time.
(c) C57BL/6J tissues were treated as in (b).
(d) Liver lysate prepared in detergent-free 20 mM tris, pH 8 was spin clarified at 7000 x g and incubated with MUC16-680 as in (b). After visualization by in-gel fluorescence, proteins were transferred to nitrocellulose and immunoblotted for MUC16. High molecular weight mucins transfer poorly to nitrocellulose, likely accounting for apparent loss of the substrate band in lane 4. Black boxes and arrows highlight the substrate band; cyan boxes and arrows highlight appearance of the pH 6 product band. Red arrows highlight lower molecular weight product bands seen following incubation at pH 5.

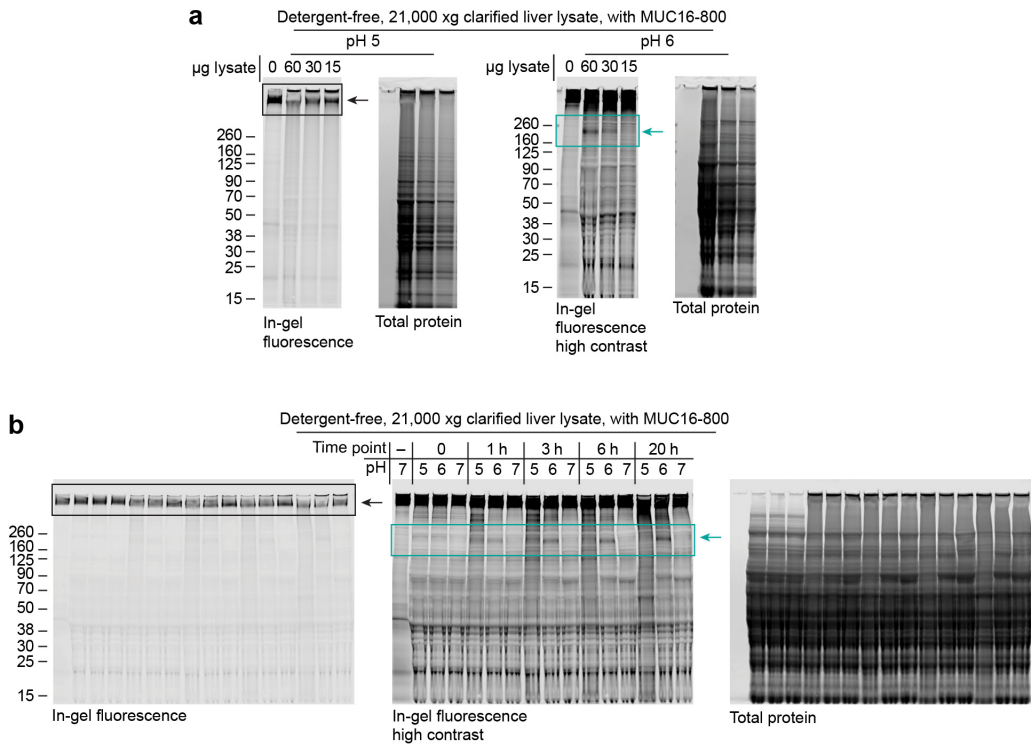


Figure S4. MUC16 cleavage by liver lysate is concentration- and time-dependent at both pH 5 and 6.
(a) Liver lysate was prepared in detergent-free 20 mM tris, pH 8, spin clarified at 21,000 x g, and diluted 2-fold serially in 20 mM tris, pH 8. It was then incubated at pH 5 and 6 with MUC16-800. Use of MUC16-800 allowed in-gel detection of substrate fluorescence in a separate channel (800 ch) from total protein by Coomassie (700 ch). Black boxes and arrows highlight the substrate band; cyan boxes and arrows highlight appearance of the pH 6 product band.
(b) Liver lysate (60 µg) prepared as in (a) was incubated with MUC16-800 for various times at the indicated pHs. Black boxes and arrows highlight the substrate band; cyan boxes and arrows highlight appearance of the pH 6 product band.

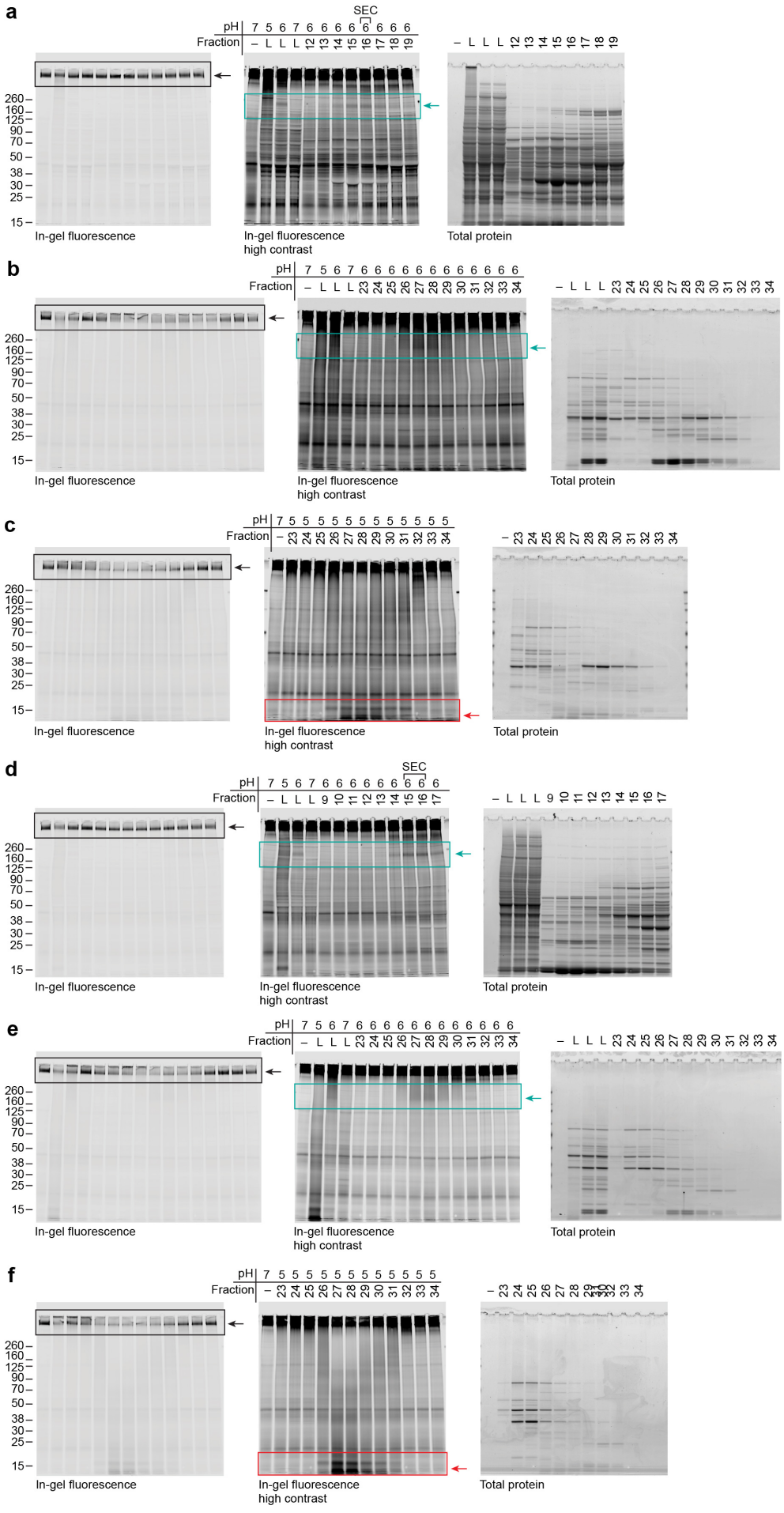

a
SEC
pH 7 5 6 7 6 6 6 6 6 6 6 6 6
Fraction – L L L 12 13 14 15 16 17 18 19
– L L L 12 13 14 15 16 17 18 19
260
160
125
90
70
50
38
30
25
15
In-gel fluorescence
In-gel fluorescence high contrast
Total protein
b
pH 7 5 6 7 6 6 6 6 6 6 6 6 6 6 6 6
Fraction – L L L 23 24 25 26 27 28 29 30 31 32 33 34
– L L L 23 24 25 26 27 28 29 30 31 32 33 34
In-gel fluorescence
In-gel fluorescence high contrast
Total protein
c
pH 7 5 5 5 5 5 5 5 5 5 5 5 5
Fraction – 23 24 25 26 27 28 29 30 31 32 33 34
– 23 24 25 26 27 28 29 30 31 32 33 34
In-gel fluorescence
In-gel fluorescence high contrast
Total protein
d
SEC
pH 7 5 6 7 6 6 6 6 6 6 6 6 6
Fraction – L L L 9 10 11 12 13 14 15 16 17
– L L L 9 10 11 12 13 14 15 16 17
In-gel fluorescence
In-gel fluorescence high contrast
Total protein
e
pH 7 5 6 7 6 6 6 6 6 6 6 6 6 6 6 6
Fraction – L L L 23 24 25 26 27 28 29 30 31 32 33 34
– L L L 23 24 25 26 27 28 29 30 31 32 33 34
In-gel fluorescence
In-gel fluorescence high contrast
Total protein
f
pH 7 5 5 5 5 5 5 5 5 5 5 5 5
Fraction – 23 24 25 26 27 28 29 30 31 32 33 34
– 23 24 25 26 27 28 29 30 31 32 33 34
In-gel fluorescence
In-gel fluorescence high contrast
Total protein

Figure S5. Uncropped gels, lower contrast images, and total protein stain corresponding to Figure 1.
(a) Corresponding to Figure 1e.
(b-c) Corresponding to Figure 1f.
(d) Corresponding to Figure 1g.
(e-f) Corresponding to Figure 1h.
Black boxes highlight the substrate band; cyan boxes highlight appearance of a product band between 160 and 260 kDa at pH 6; red boxes highlight appearance of the product band between 8 and 15 kDa at pH 5. L = input lysate. Dash = MUC16 only.

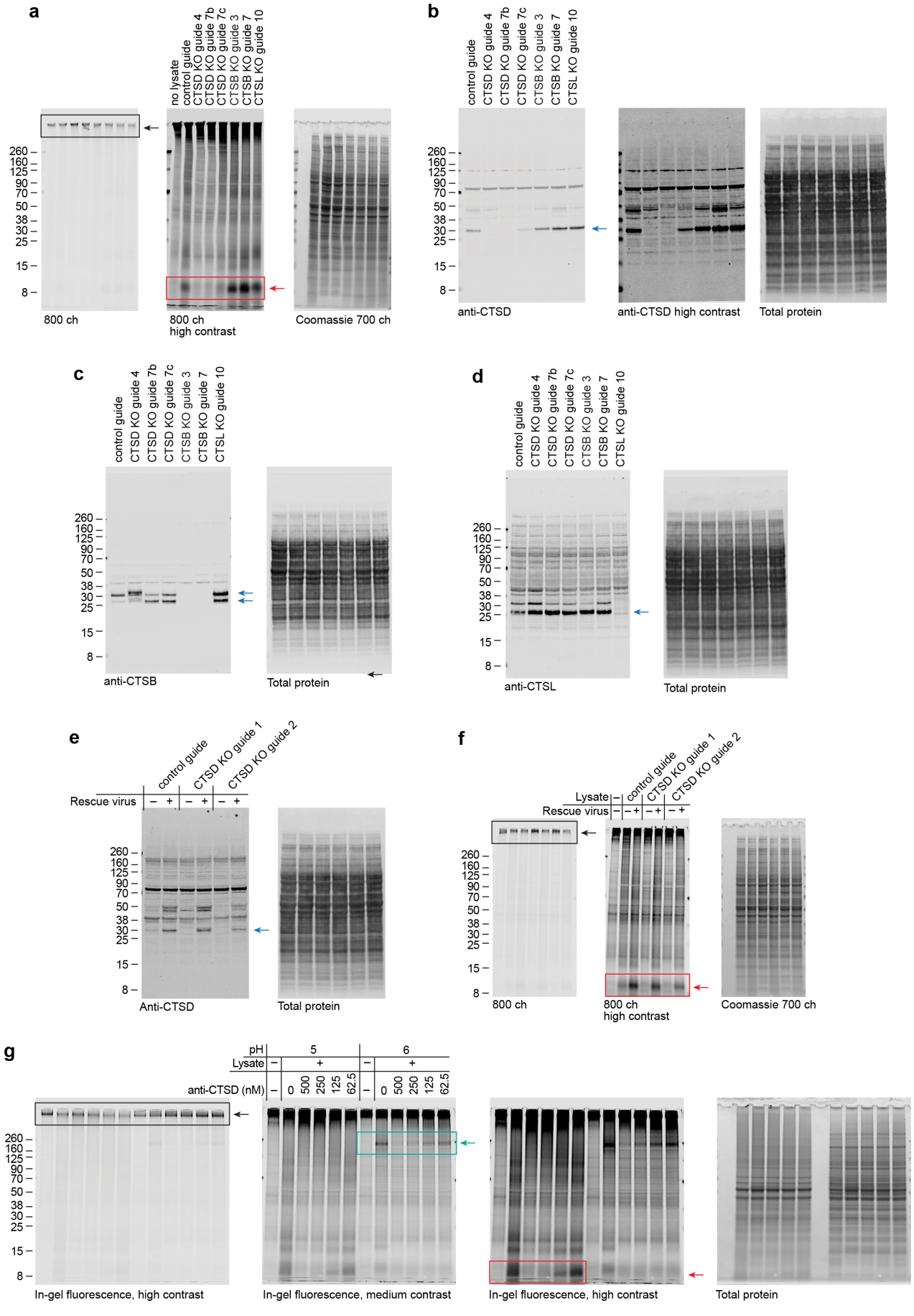

a
no lysate
control guide
CTSD KO guide 4
CTSD KO guide 7b
CTSD KO guide 7c
CTSB KO guide 3
CTSB KO guide 7
CTSL KO guide 10
260
160
125
90
70
50
38
30
25
15
8
800 ch
800 ch
high contrast
Coomassie 700 ch
b
control guide
CTSD KO guide 4
CTSD KO guide 7b
CTSD KO guide 7c
CTSB KO guide 3
CTSB KO guide 7
CTSL KO guide 10
anti-CTSD
anti-CTSD high contrast
Total protein
c
control guide
CTSD KO guide 4
CTSD KO guide 7b
CTSD KO guide 7c
CTSB KO guide 3
CTSB KO guide 7
CTSL KO guide 10
anti-CTSB
Total protein
d
control guide
CTSD KO guide 4
CTSD KO guide 7b
CTSD KO guide 7c
CTSB KO guide 3
CTSB KO guide 7
CTSL KO guide 10
anti-CTSL
Total protein
e
control guide
CTSD KO guide 1
CTSD KO guide 2
Rescue virus
Anti-CTSD
Total protein
f
Lysate
control guide
CTSD KO guide 1
CTSD KO guide 2
Rescue virus
800 ch
800 ch
high contrast
Coomassie 700 ch
g
pH
5
6
Lysate
anti-CTSD (nM)
0
500
250
125
62.5
In-gel fluorescence, high contrast
In-gel fluorescence, medium contrast
In-gel fluorescence, high contrast
Total protein

Figure S6. Uncropped gels, lower contrast images, and total protein stain corresponding to Figure 2b-d.
(a-d) Corresponding to Figure 2b.
(e-f) Corresponding to Figure 2c.
(g) Corresponding to Figure 2d.
Black boxes highlight the substrate band; cyan boxes highlight appearance of a product band between 160 and 260 kDa at pH 6; red boxes highlight appearance of the product band between 8 and 15 kDa at pH 5. Blue arrows highlight the expected MW of the immunoblot target.

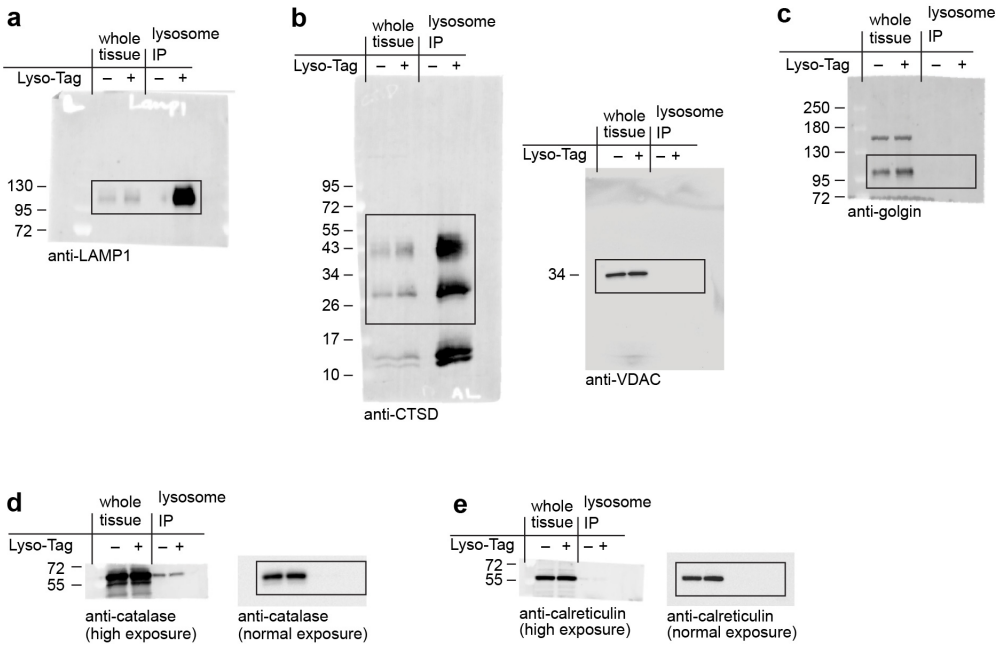


Figure S7. Uncropped gels and lower contrast images corresponding to Figure 2f. Black boxes indicate the region displayed in the main text figure.

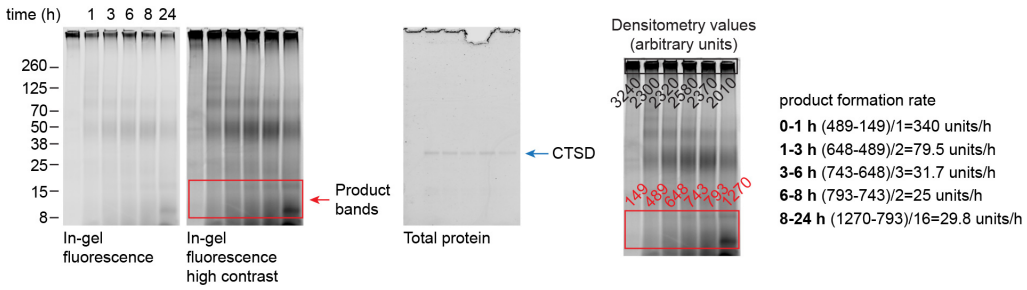


Figure S8. Quantification of MUC16 cleavage by human lysosomal cathepsin D. MUC16-800 (8000 U/mL) and purified cathepsin D (500 nM) were incubated in pH 5 buffer at 37 °C for the times shown, quenched with protein loading buffer, then loaded on an SDS-PAGE gel. Background-subtracted densitometry of the high molecular weight substrate band and 8-15 kDa product band indicates ~40% substrate loss over 24 hours.

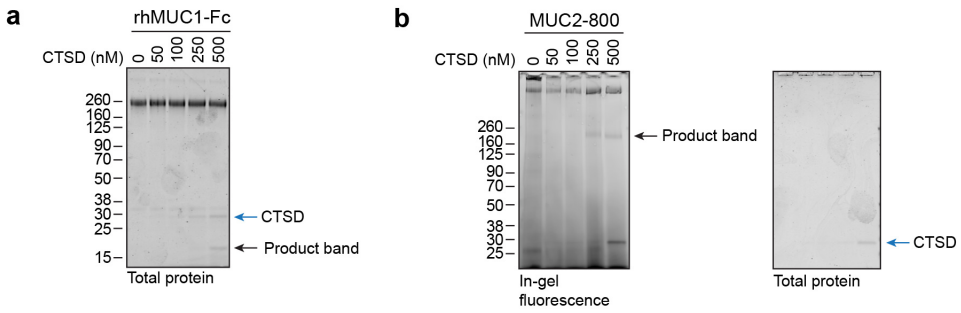


Figure S9. Cleavage of additional mucin substrates by human lysosomal cathepsin D, related to Figure 3c. Recombinant human MUC1-Fc chimera (R&D Systems) (a) and MUC2-800 (prepared as described in *Methods*) (b) were incubated with purified cathepsin D at the concentrations shown in pH 5 buffer for 20 hours at 37 °C, quenched with protein loading buffer, then loaded an SDS-PAGE gels. MUC1-Fc cleavage was assessed by Coomassie total protein stain; MUC2-800 cleavage was assessed by in-gel fluorescence.

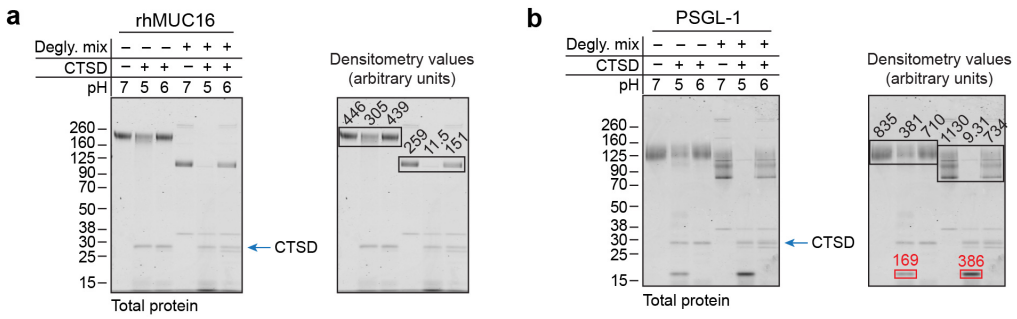


Figure S10. Activity of human lysosomal cathepsin D on enzymatically deglycosylated substrates. Recombinant human MUC16 (rhMUC16) (a) and PSGL-1 (b) substrates were deglycosylated using a commercial enzymatic deglycosylation cocktail (Degly. mix) as described in *Methods*, and as indicated by gel shifts to lower molecular weight. Equal portions of substrate and deglycosylated substrate were incubated with 500 nM purified cathepsin D for 6 hours at 37 °C. Background-subtracted densitometry values are shown, *right*.

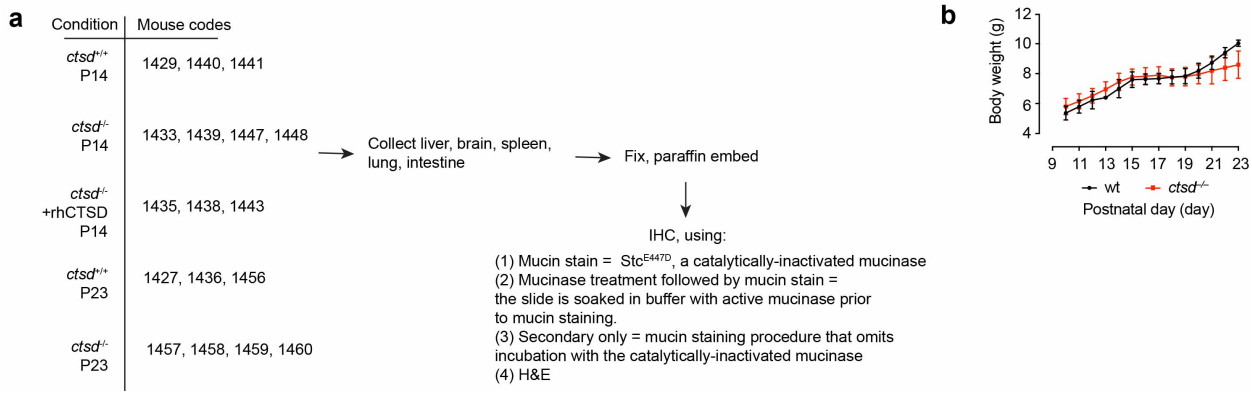


Figure S11. Mouse codes and weights for cathepsin D knockout animals and controls.
(a) Table of identifying codes for wild-type (*ctsd*$^{+/+}$), cathepsin D knockout (*ctsd*$^{-/-}$), and rescued (*ctsd*$^{-/-}$ + rhCTSD) animals, along with workflow for tissue processing.
(b) Weights of all animals through P23. Divergence in weight begins at P19. Weight loss correlates with disease progression in these animals. Error bars are standard deviation.

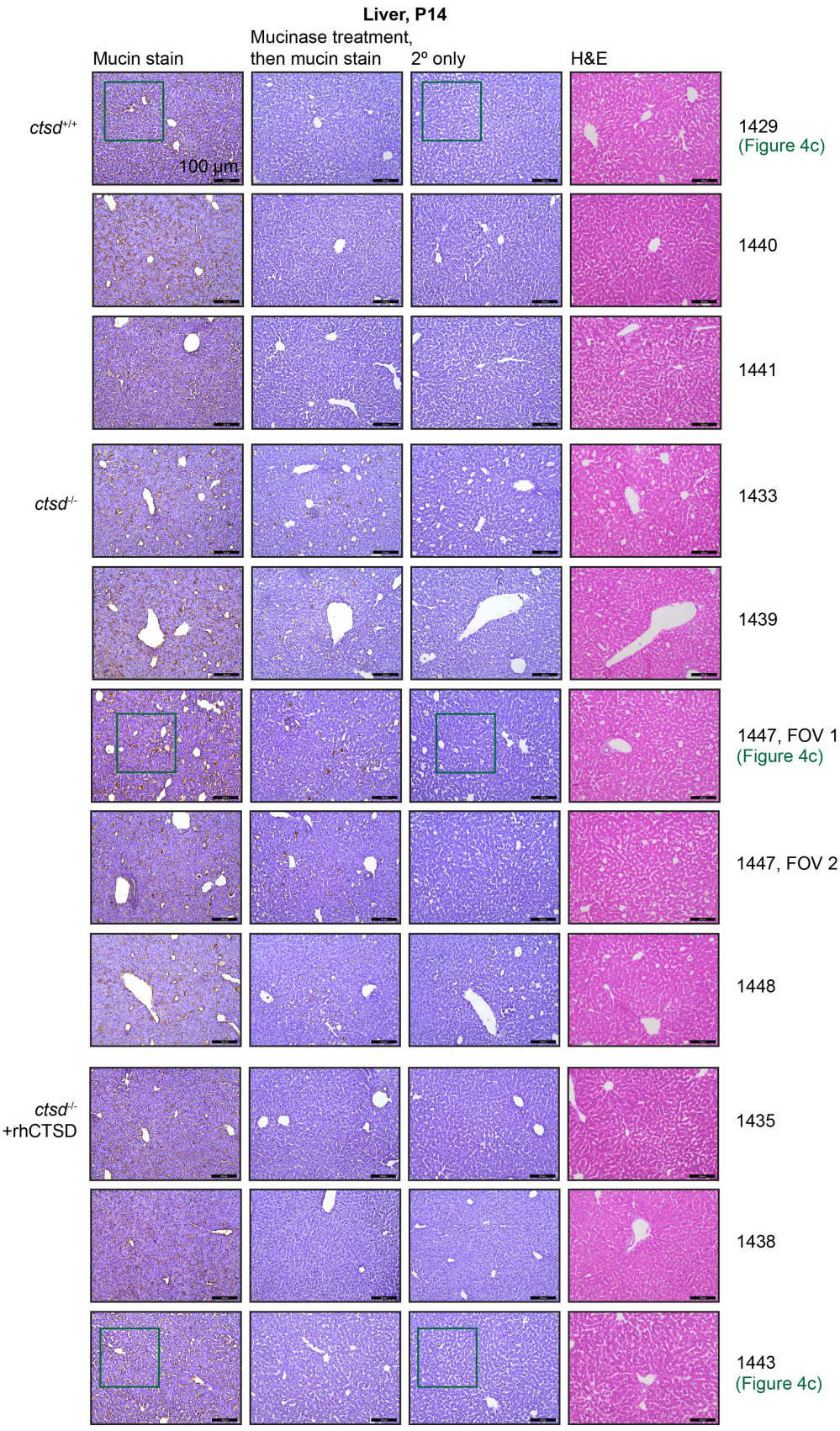

Liver, P14
Mucin stain
Mucinase treatment, then mucin stain
2º only
H&E
100 µm
ctsd+/+
ctsd-/-
ctsd-/- +rhCTSD
1429 (Figure 4c)
1440
1441
1433
1439
1447, FOV 1 (Figure 4c)
1447, FOV 2
1448
1435
1438
1443 (Figure 4c)

Figure S12. Liver P14 tissues. Fixed and paraffin-embedded liver from P14 wild-type (*ctsd*$^{+/+}$), cathepsin D knockout (*ctsd*$^{-/-}$), and rescued (*ctsd*$^{-/-}$ + rhCTSD) animals were stained for mucins using Biotin-StcE^{E447D}, with hematoxylin counterstaining ("Mucin stain"). As controls, sections from the same tissues were processed with omission of Biotin-StcE^{E447D} ("2° only"). For confirmation of the specificity of the mucin stain, tissues were pre-treated with 50 nM of the mucinase StcE for 20 hours in PBS, then stained for mucins ("Mucinase treatment, then mucin stain"). Green squares denote cropped regions shown in Figure 4d. Numbers at the right correspond to individual mice, as tabulated in Figure S11. Greater mucin staining is observed in the knockout animals relative to the wild-type and rescued animals. Mucin accumulation appears to be within enlarged cells.

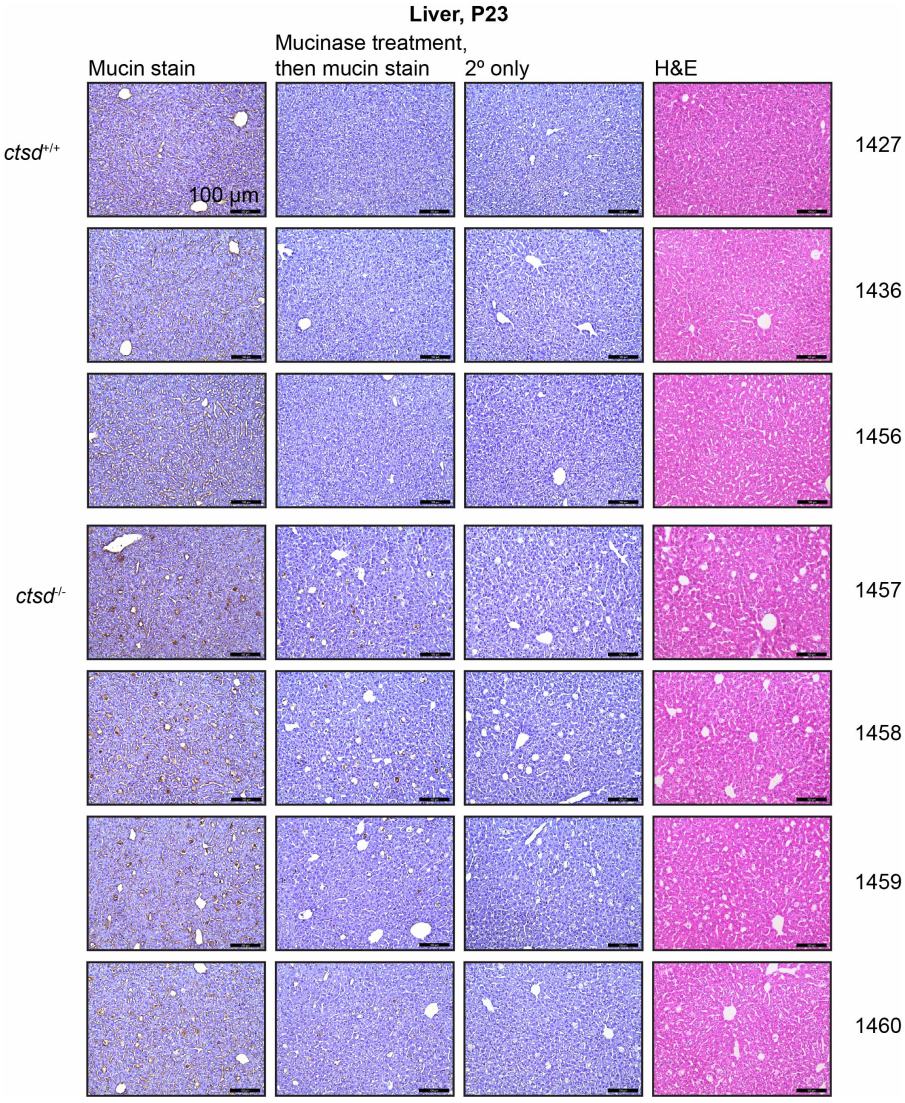


Figure S13. Liver P23 tissues. Fixed and paraffin-embedded liver from P23 wild-type (*ctsd*$^{+/+}$) and cathepsin D knockout (*ctsd*$^{-/-}$) animals were stained for mucins using Biotin-StcE^{E447D}, with hematoxylin counterstaining ("Mucin stain", see *Methods*). As controls, sections from the same tissues were processed with omission of Biotin-StcE^{E447D} ("2° only"). For confirmation of the specificity of the mucin stain, tissues were pre-treated with 50 nM of the mucinase StcE for 20 hours in PBS, then stained for mucins ("Mucinase treatment, then mucin stain"). Numbers at the right correspond to individual mice, as tabulated in Figure S11. Greater mucin staining is observed in the knockout animals relative to the wild-type animals.

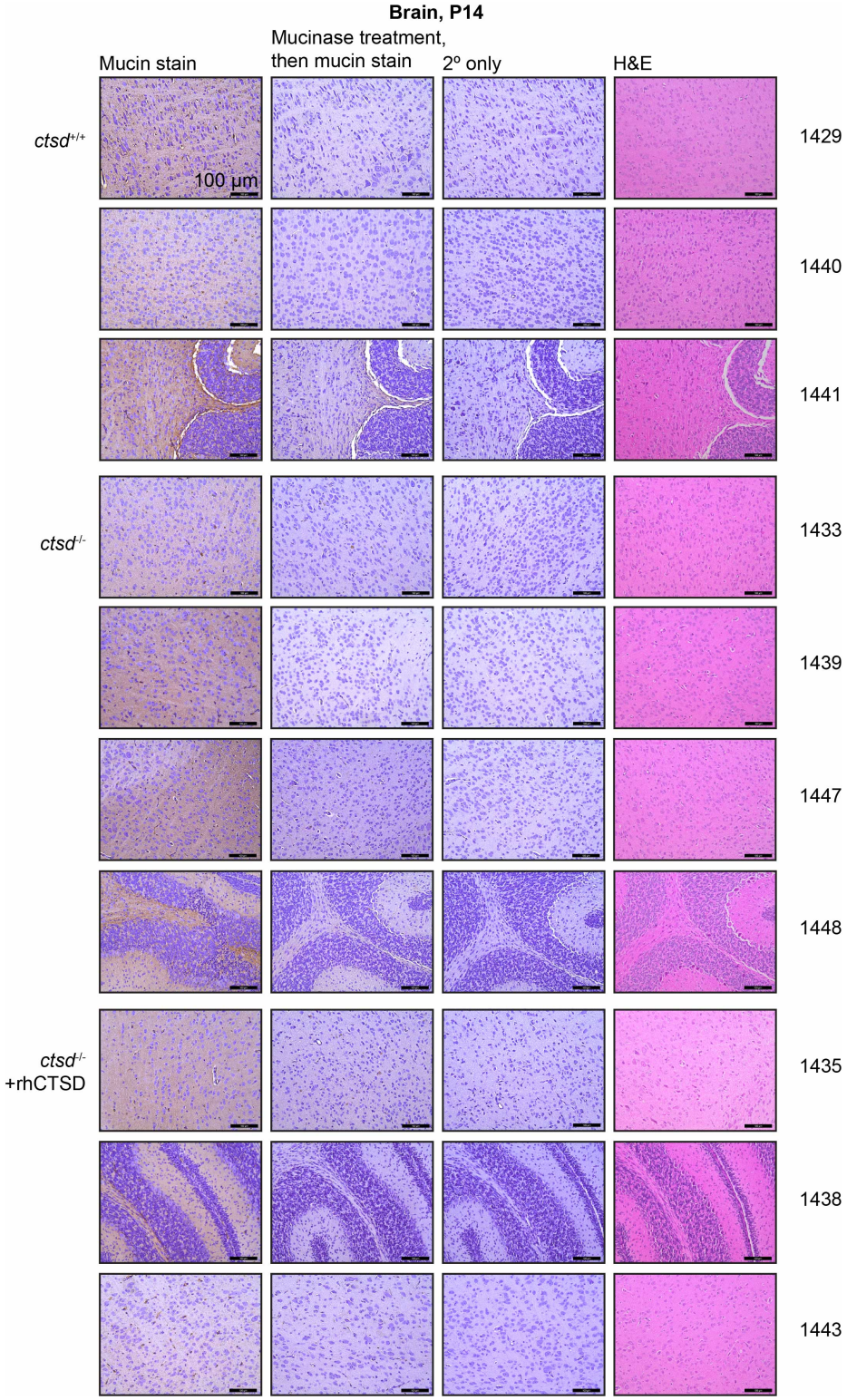

Brain, P14
Mucin stain
Mucinase treatment, then mucin stain
2º only
H&E
100 µm
ctsd+/+
ctsd-/-
ctsd-/- +rhCTSD
1429
1440
1441
1433
1439
1447
1448
1435
1438
1443

Figure S14. Brain P14 tissues, represented as in Figure S12. Mucin staining patterns did not differ across animals, possibly due to the greater resolution needed to see accumulated storage material in the brain[21].

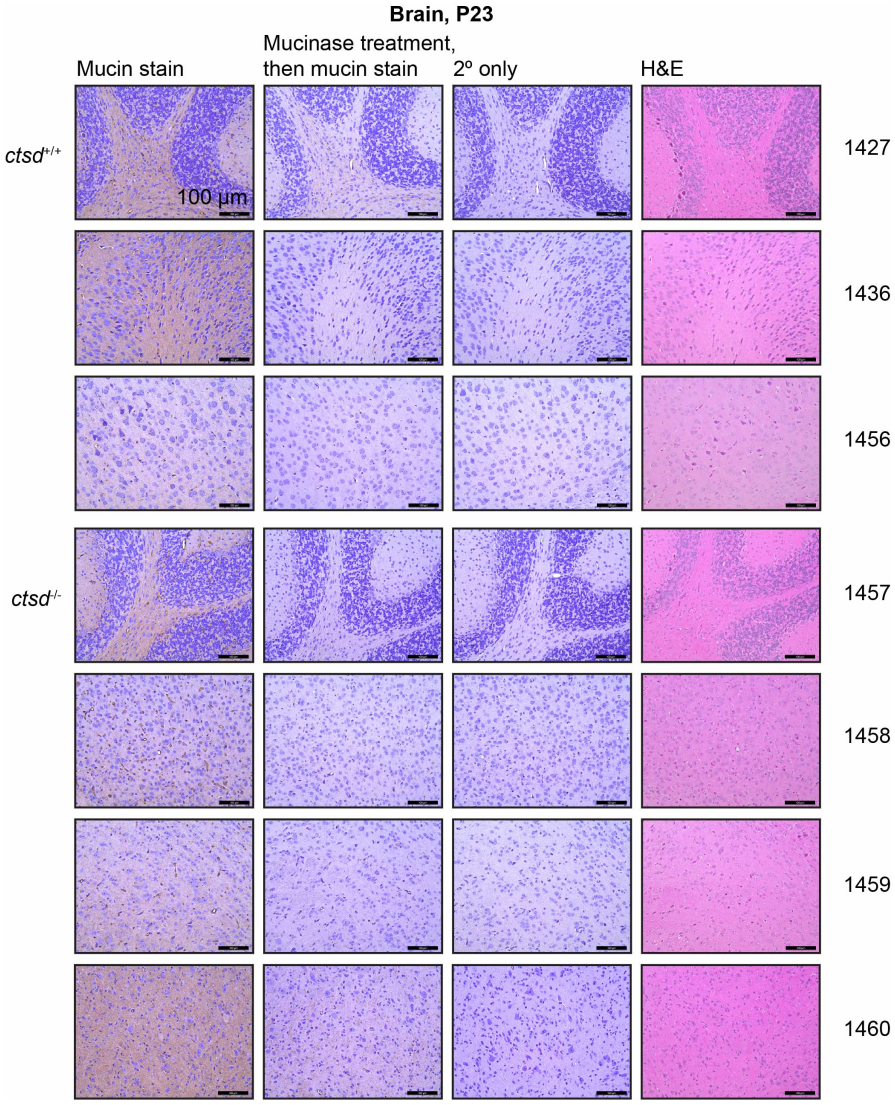


Figure S15. Brain P23 tissues, represented as in Figure S13. Mucin staining patterns did not differ across animals, possibly due to the greater resolution needed to see accumulated storage material in the brain[21].

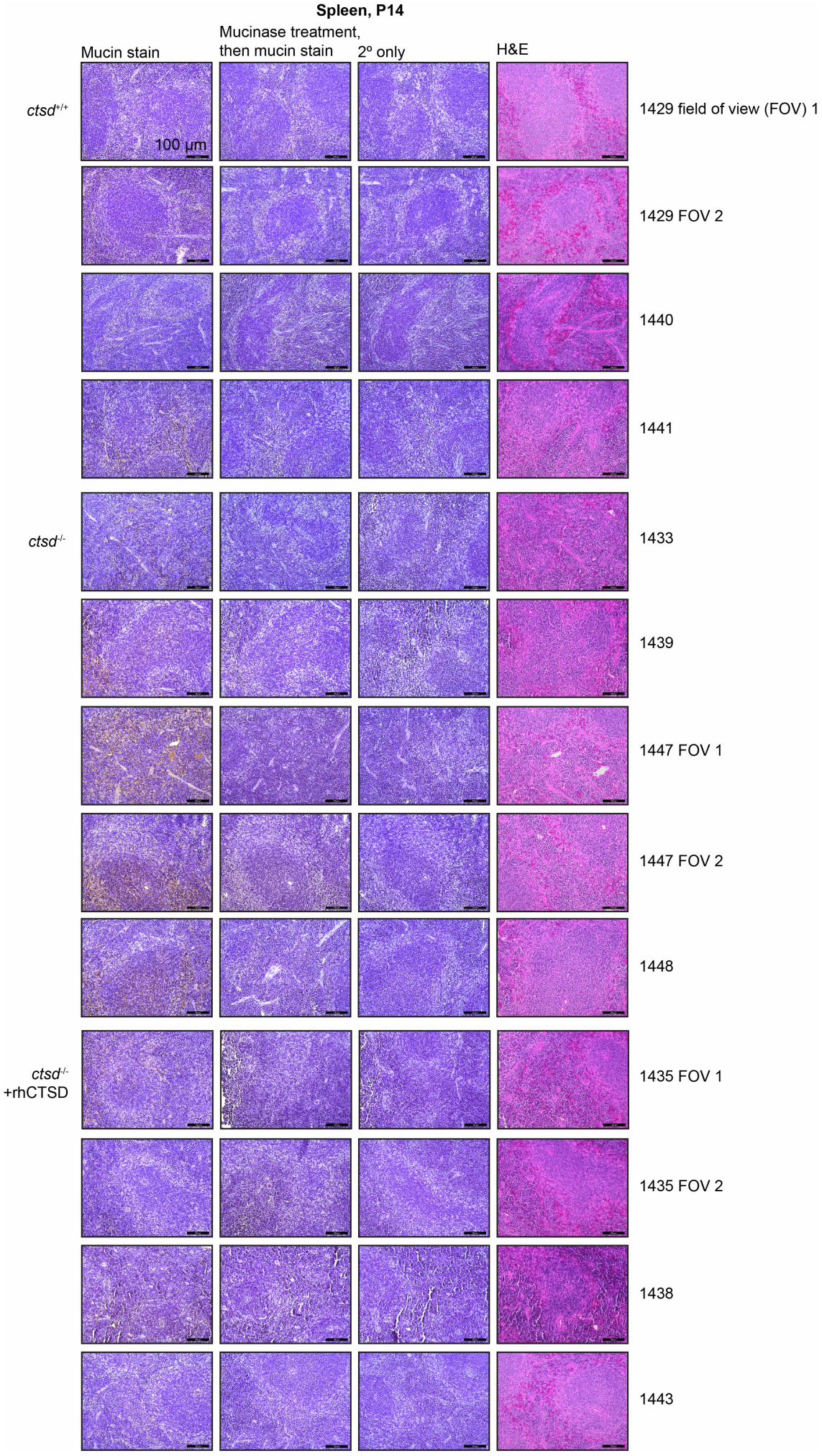

Spleen, P14
Mucin stain
Mucinase treatment, then mucin stain
2º only
H&E
100 µm
ctsd+/+
ctsd-/-
ctsd-/- +rhCTSD
1429 field of view (FOV) 1
1429 FOV 2
1440
1441
1433
1439
1447 FOV 1
1447 FOV 2
1448
1435 FOV 1
1435 FOV 2
1438
1443

Figure S16. Spleen P14 tissues, represented as in Figure S12. Greater mucin staining is observed in the knockout animals relative to the wild-type and rescued animals. Mucin accumulation appears to be distributed across the tissue, surrounding cells. FOV = field of view.

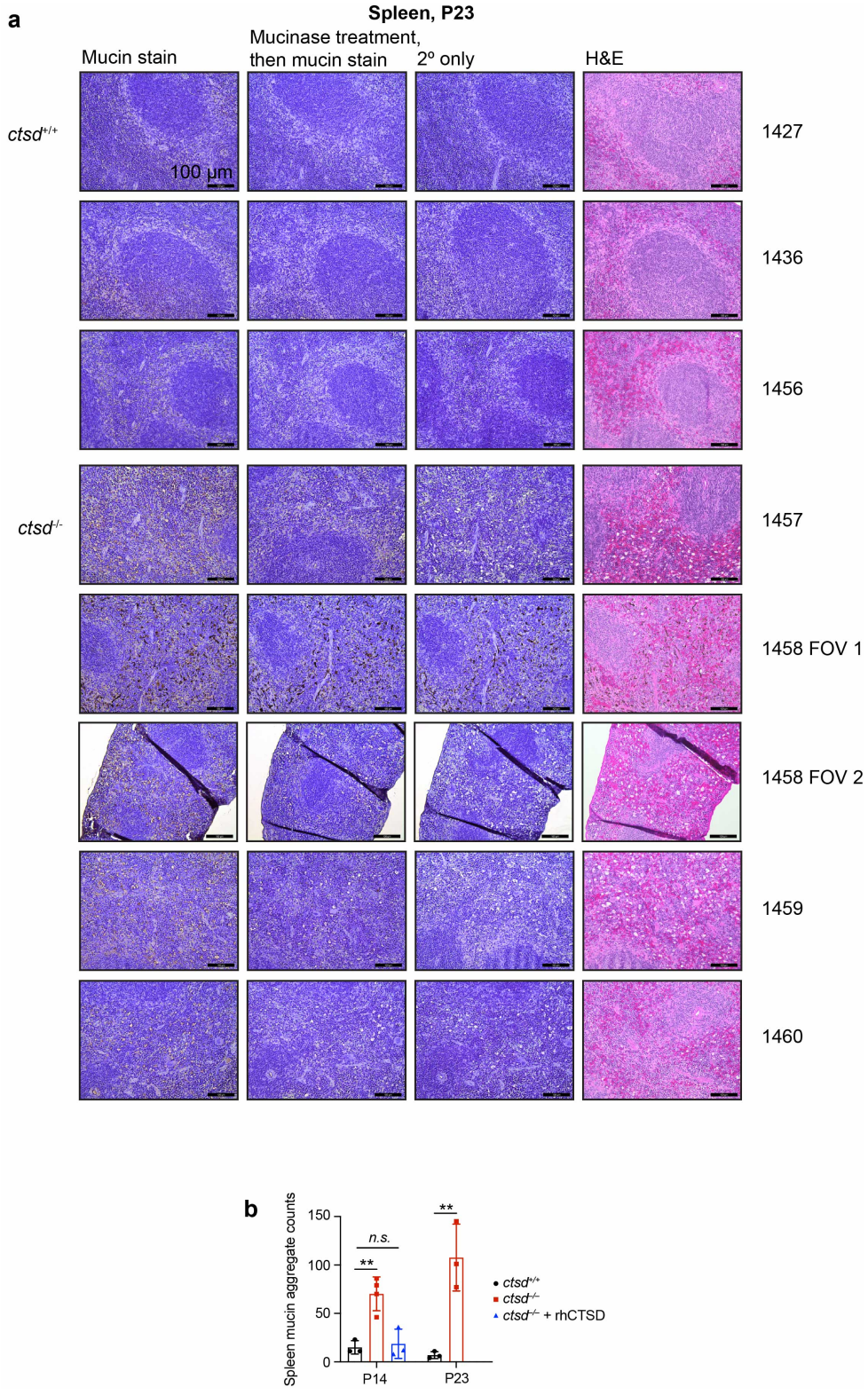

a
Spleen, P23
Mucin stain
Mucinase treatment, then mucin stain
2º only
H&E
ctsd+/+
ctsd-/-
100 µm
1427
1436
1456
1457
1458 FOV 1
1458 FOV 2
1459
1460
b
Spleen mucin aggregate counts
150
100
50
0
n.s.
**
**
P14
P23
ctsd+/+
ctsd-/-
ctsd-/- + rhCTSD

Figure S17. Spleen P23 tissues, represented as in Figure S13. (a) Greater mucin staining is observed in the knockout animals relative to the wild-type and rescued animals. Mucin accumulation appears to be distributed in vacuoles, which may represent tissue damage, as previously reported[21]. FOV = field of view. (b) Quantification of mucin aggregates for P14 and P23 spleen as in Figure 4c and described in Methods. Error bars are standard deviation. ** = $p < 0.01$, n.s. = $p > 0.05$.

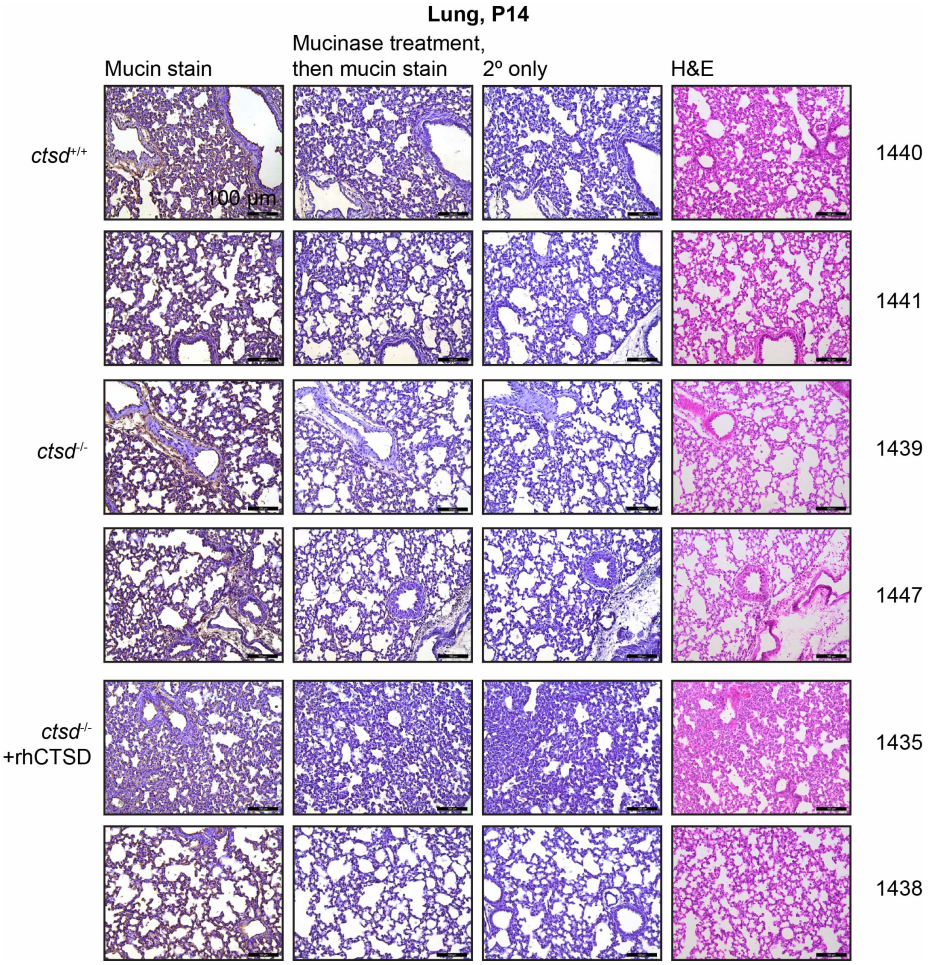


Figure S18. Lung P14 tissues, represented as in Figure S12. Mucin staining patterns did not differ across animals.

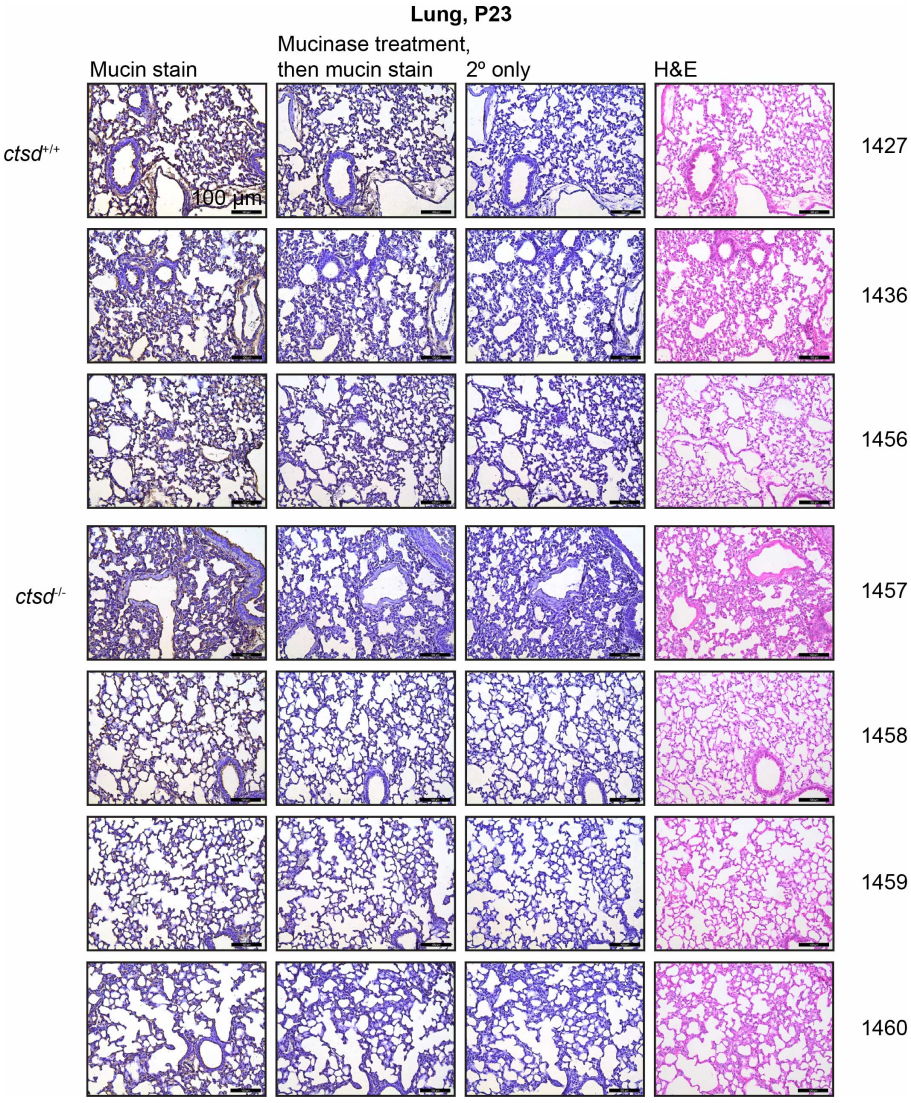


Figure S19. Lung P23 tissues, represented as in Figure S13. Mucin staining patterns did not differ across animals.

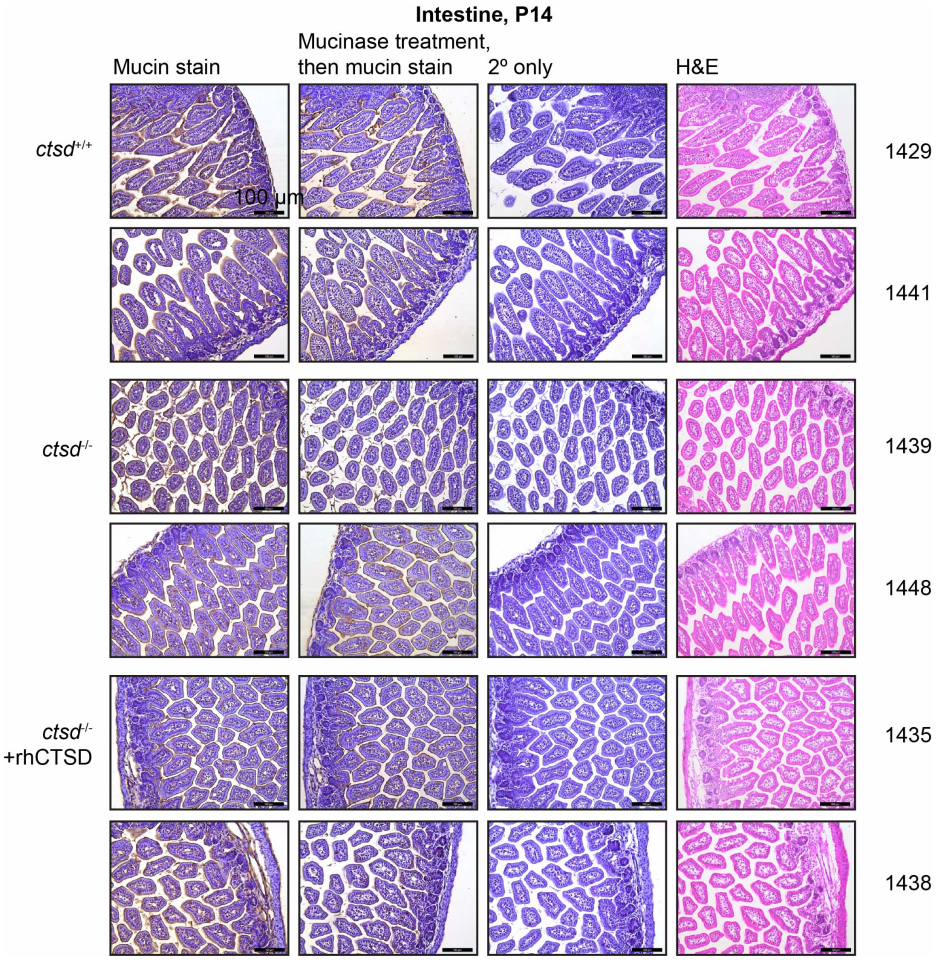


Figure S20. Intestine P14 tissues, represented as in Figure S12. Mucin staining patterns did not differ across animals.

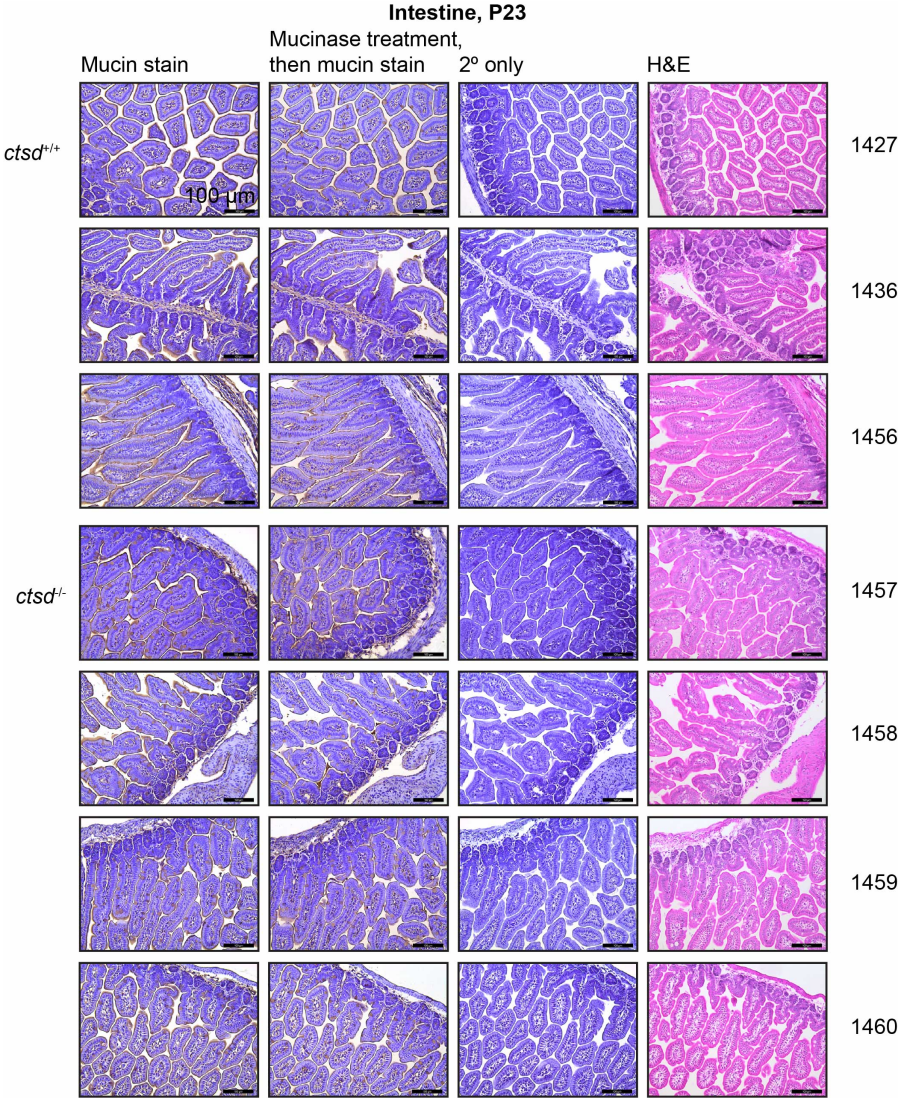


Figure S21. Intestine P23 tissues, represented as in Figure S13. Mucin staining patterns did not differ across animals.

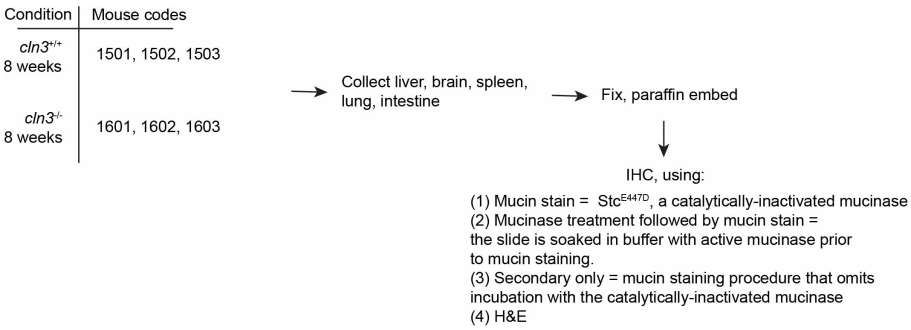


Figure S22. Table of identifying codes for wild-type (*cln3*$^{+/+}$) and CLN3 knockout (*cln3*$^{-/-}$), along with workflow for tissue processing. Weights were consistent across animals.

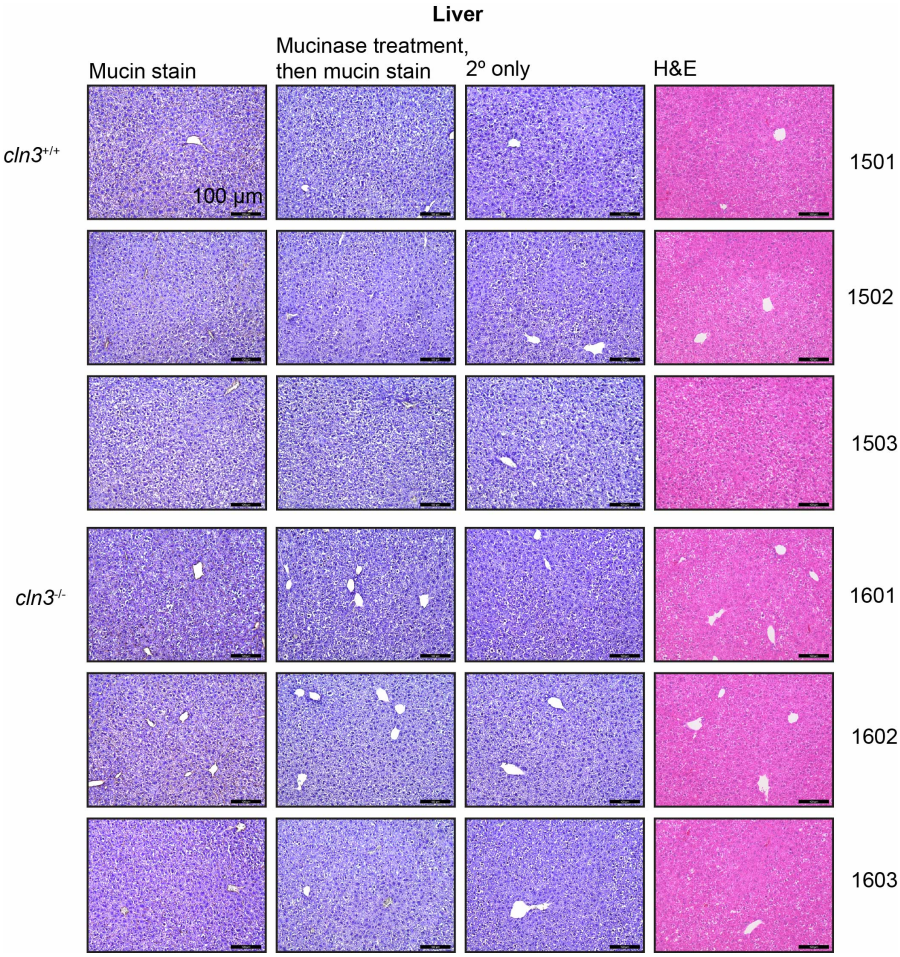


Figure S23. Fixed and paraffin-embedded liver from ~8-week old wild-type (*cln3*$^{+/+}$) and CLN3 knockout (*cln3*$^{-/-}$) animals were stained for mucins using Biotin-StcE^{E447D}, with hematoxylin counterstaining ("Mucin stain"). As controls, sections from the same tissues were processed with omission of Biotin-StcE^{E447D} ("2° only"). For confirmation of the specificity of the mucin stain, tissues were pre-treated with 50 nM of the mucinase StcE for 20 hours in PBS, then stained for mucins ("Mucinase treatment, then mucin stain"). Numbers at the right correspond to individual mice, as tabulated in Figure S22. Mucin staining patterns did not differ across animals.

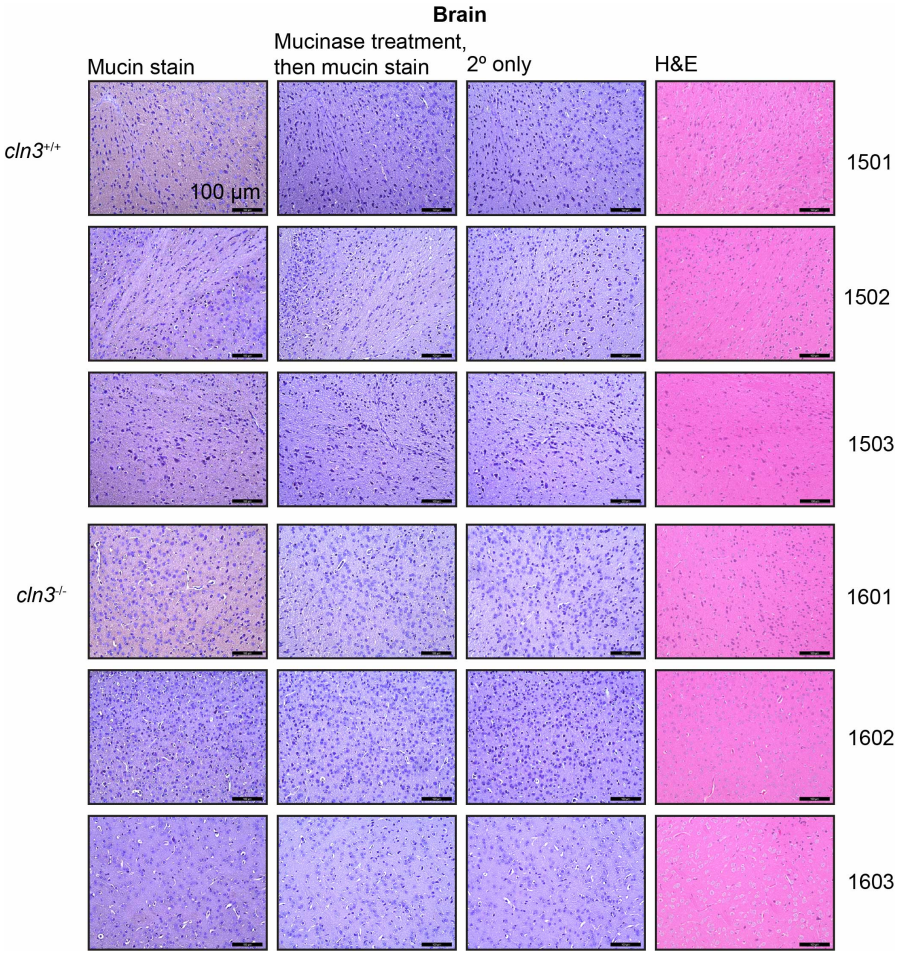


Figure S24. Brain tissues prepared as in Figure S23. Mucin staining patterns did not differ across animals.

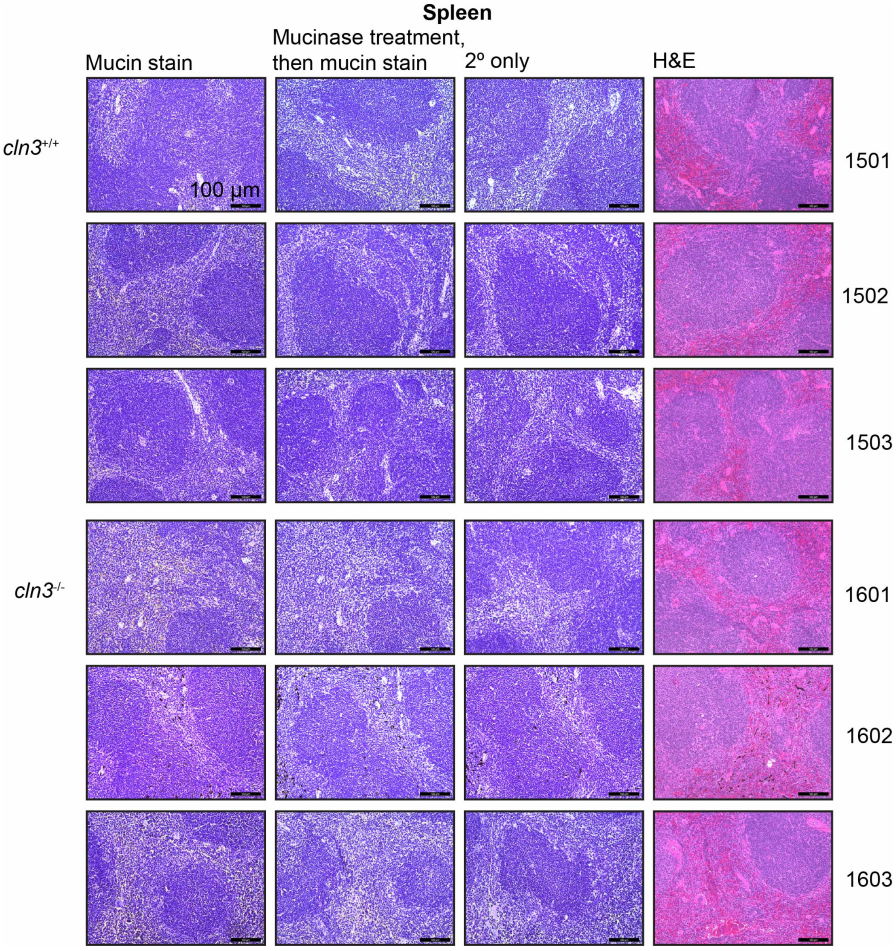


Figure S25. Spleen tissues prepared as in Figure S23. Mucin staining patterns did not differ across animals.

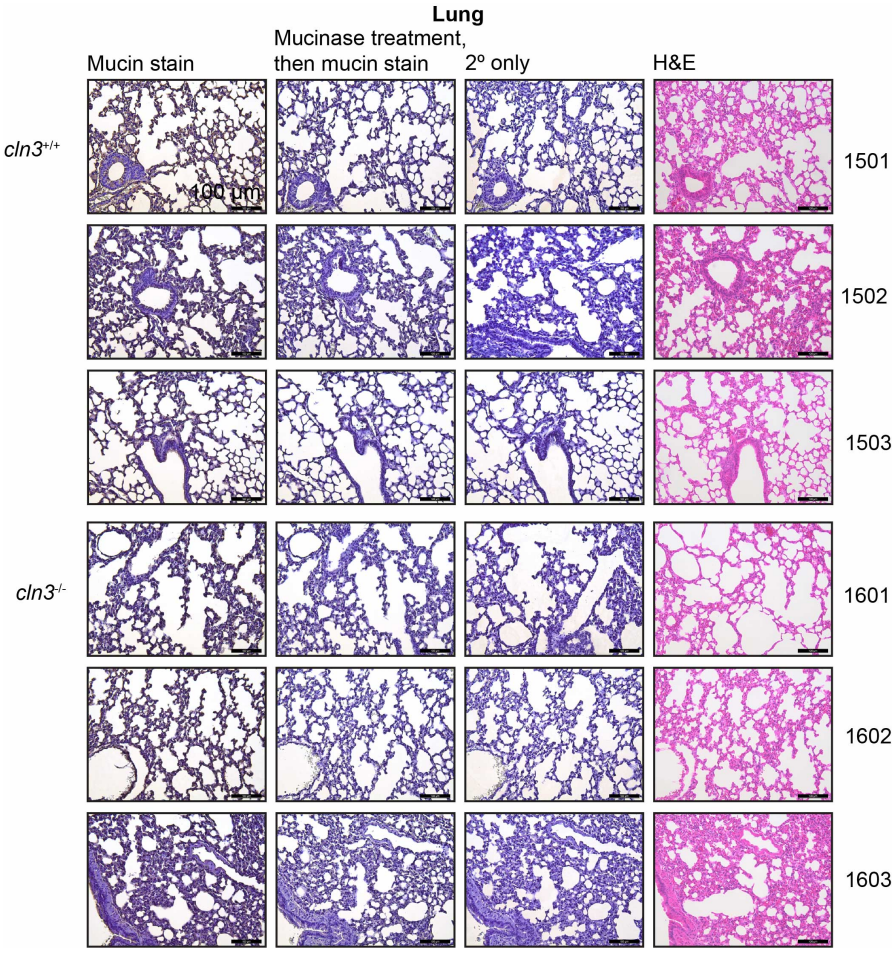


Figure S26. Lung tissues prepared as in Figure S23. Mucin staining patterns did not differ across animals.

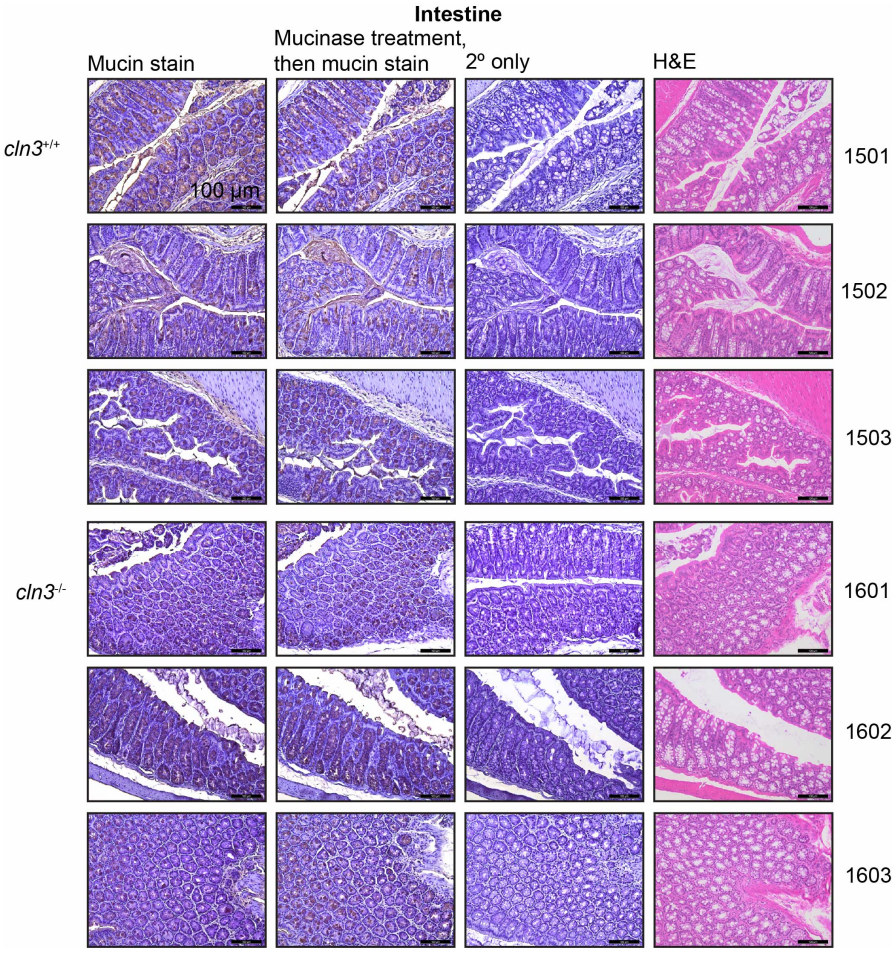


Figure S27. Intestine tissues prepared as in Figure S23. Mucin staining patterns did not differ across animals.

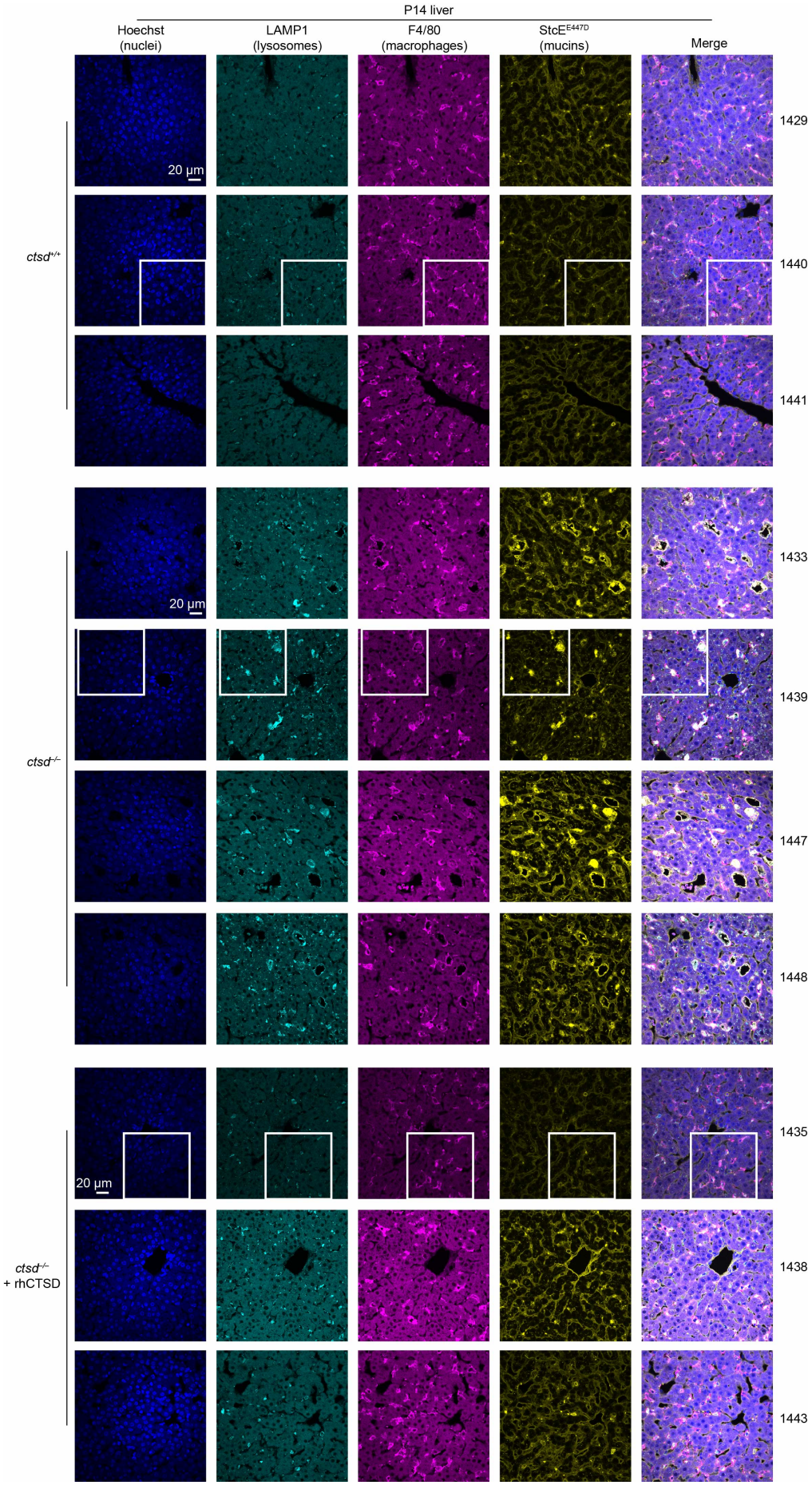

P14 liver
Hoechst (nuclei)
LAMP1 (lysosomes)
F4/80 (macrophages)
StcE^{E447D} (mucins)
Merge
20 µm
ctsd$^{+/+}$
ctsd$^{-/-}$
ctsd$^{-/-}$ + rhCTSD
1429
1440
1441
1433
1439
1447
1448
1435
1438
1443

Figure S28. Immunofluorescence images of liver from P14 *ctsd*$^{+/+}$, *ctsd*$^{-/-}$, and *ctsd*$^{-/-}$ + rhCTSD mice. Fixed and paraffin-embedded slices were de-waxed, hydrated, and subjected to antigen retrieval and permeabilization (Methods). Tissues were stained with the markers shown and imaged by confocal microscopy, taking 15-18 Z-stacks with 1 μm step size. Maximum intensity projections are shown with brightness and contrast normalized across all images. Numbers at the right correspond to individual mice, as tabulated in Figure S11. White squares denote cropped regions shown in Figure 4e.

SI References

1. Wong, N. K. *et al.* Characterization of the Oligosaccharides Associated with the Human Ovarian Tumor Marker CA125. *J. Biol. Chem.* **278**, 28619–28634 (2003).

2. Recktenwald, C. V. & Hansson, G. C. The Reduction-insensitive Bonds of the MUC2 Mucin Are Isopeptide Bonds*. *J. Biol. Chem.* **291**, 13580–13590 (2016).

3. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* (2016) doi:10.1038/nprot.2016.136.

4. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–805 (2011).

5. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics MCP* **13**, 2513–26 (2014).

6. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).

7. Malaker, S. A. *et al.* The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7278–7287 (2019).

8. Shon, D. J. *et al.* An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 21299–21307 (2020).

9. Riley, N. M., Malaker, S. A., Driessen, M. D. & Bertozzi, C. R. Optimal Dissociation Methods Differ for N- and O-Glycopeptides. *J. Proteome Res.* **19**, 3286–3301 (2020).

10. Riley, N. M., Malaker, S. A. & Bertozzi, C. R. Electron-Based Dissociation Is Needed for O-Glycopeptides Derived from OpeRATOR Proteolysis. *Anal. Chem.* **92**, 14878–14884 (2020).

11. Saba, J., Dutta, S., Hemenway, E. & Viner, R. Increasing the Productivity of Glycopeptides Analysis by Using Higher-Energy Collision Dissociation-Accurate Mass-Product-Dependent Electron Transfer Dissociation. *Int. J. Proteomics* **Vol. 2012**, Article ID 560391 (2012).

12. Wu, S. W., Pu, T. H., Viner, R. & Khoo, K. H. Novel LC-MS2 product dependent parallel data acquisition function and data analysis workflow for sequencing and identification of intact glycopeptides. *Anal. Chem.* **86**, 5478–5486 (2014).

13. Singh, C., Zampronio, C. G., Creese, A. J. & Cooper, H. J. Higher energy collision dissociation (HCD) product ion-triggered electron transfer dissociation (ETD) mass spectrometry for the analysis of N-linked glycoproteins. *J. Proteome Res.* **11**, 4517–25 (2012).

14. Čaval, T., Zhu, J. & Heck, A. J. R. Simply Extending the Mass Range in Electron Transfer Higher Energy Collisional Dissociation Increases Confidence in N-Glycopeptide Identification. *Anal. Chem.* **91**, 10401–10406 (2019).

15. Rose, C. M. *et al.* A Calibration Routine for Efficient ETD in Large-Scale Proteomics. *J. Am. Soc. Mass Spectrom.* **26**, 1848–1857 (2015).

16. Lu, L., Riley, N. M., Shortreed, M. R., Bertozzi, C. R. & Smith, L. M. O-Pair Search with MetaMorpheus for O-glycopeptide characterization. *Nat. Methods* **17**, 1133–1138 (2020).

17. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

18. Brademan, D. R., Riley, N. M., Kwiecien, N. W. & Coon, J. J. Interactive peptide spectral annotator: A versatile web-based tool for proteomic applications. *Mol. Cell. Proteomics* **18**, S193–S201 (2019).

19. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

20. Gahmberg, C. G. & Andersson, L. C. Role of Sialic Acid in the Mobility of Membrane Proteins Containing O-Linked Oligosaccharides on Polyacrylamide Gel Electrophoresis in Sodium Dodecyl Sulfate. *Eur. J. Biochem.* **122**, 581–586 (1982).

21. Marques, A. R. A. *et al.* Enzyme replacement therapy with recombinant pro-CTSD (cathepsin D) corrects defective proteolysis and autophagy in neuronal ceroid lipofuscinosis. *Autophagy* **16**, 811–825 (2020).