Figure S1: Genomic characterization of Shpk-knockout mice. (A) PCR gel depicting flanking genotyping primers of normal controls (arrow) or 15 potentially edited Shpk<sup>ΔATG</sup> founders. Expected WT band size is 820 bp. Starred lanes had detectable deletions by Sanger Sequencing. Sequenced traces of bracketed band, representing the selected founder, shown to right. (B-C) PCR-based characterization of F1 generation of Shpk<sup>ΔATG</sup> crossed to WT C57BL/6 mice (B) and cross-bred F2 offspring (C). (D) Representative PCR-based genotyping of Shpk<sup>ΔE2</sup> F2 offspring. Expected WT band size is 638 bp. Sanger sequencing of numbered WT (1) or mutant (2) bands displayed to right.
Figure S2: Elimination of Shpk has limited impact on metabolomics outside of the PPP. With the exception of a decrease of hexose sugars (glucose and fructose) and AMP in Shpk<sup>-/-</sup> samples, no significant differences in the liver were observed between the different groups for energy carriers, hexose sugars, or acetyl-coA.
Figure S3: (A) Calibration curves used for quantification of erythritol, sedoheptulose, and erythronic acid. Linear regression equation and correlation coefficient (r) shown above each curve. No weighting was used in the regression analysis. (B) Chromatograms for erythritol (B1), d6-erythritol (B2), erythronic acid (B3), U-13C6-glucose (B4), sedoheptulose (B5) in CN mutant, and sedoheptulose (B6) in WT. Note, there were isobaric peaks slightly overlapping with sedoheptulose, with similar magnitude in all samples. Peaks were integrated to best isolate the area corresponding to sedoheptulose as shown above. (C) Accuracy and precision statistics for the quantification of erythritol, erythronic acid, and sedoheptulose. Four replicate extractions and runs were performed for each QC level. Averages of the accuracy are shown in addition to the %CV. All deviations in accuracy and precision met standard guidelines for bioanalytical assays (ref), ≤20% for lloq, and ≤15% for low, medium, and high QC. There was a single outlier (medium QC for erythritol) that did not meet those criteria and was removed, but the other three replicates, as well as all other QC replicates for all compounds met those criteria.
Table S1: Primer sequences for qPCR and genotyping. All primers reconstituted at 100 μM and used at a working concentration of 10 μM.

| GENE  | FULL NAME                     | PURPOSE           | DIRECTION | PRIMER SEQUENCE                  |
|-------|--------------------------------|-------------------|-----------|-----------------------------------|
| Ctns  | Cystinosin                    | mRNA Expression   | sense     | CAAATACCCCAACGGAGTGAA             |
|       |                                |                   | antisense | GCGTGAGGCTGAAGAAGAC               |
| Shpk  | Sedoheptulokinase             | mRNA Expression   | sense     | CAGGCAAGGCTGTGAAT                 |
|       |                                |                   | antisense | GCCAGCTGCATCATAGGACT              |
| GapDH | Glyceraldehyde 3-phosphate     | qPCR              | sense     | GCACAGTCAAGGCCGAAGAAT             |
|       | dehydrogenase                 | Housekeeping      | antisense | GCCTTCTCCATGTTGGTGAA              |
| Shpk  | Sedoheptulokinase             | Shpk ΔATG         | sense     | ACCAATAGATCCCGCTTCT              |
|       |                                | Genotyping        | antisense | TGATGGGCGGATAGTAGGAGG             |
| Shpk  | Sedoheptulokinase /Carbohydrate-like Kinase | Shpk ΔE2 | sense     | CTCGTCACCTCTGTTCTCTCC            |
|       |                                | Genotyping        | antisense (WT) | CTTGGCTGCTTTCAAAAATAGG          |
|       |                                |                   | antisense (Mutant) | CACCTGCATTCCATCCCTAGACC         |
Table S2: Complete blood count analysis in the different groups of mice.

| Parameter (units) | WT (n=5) | Ctns\(^{+/-}\) (n=9) | Ctns\(^{+/-}\) HSPC (n=11) | WT HSPC (n=13) | Shpk\(^{+/-}\) HSPC (n=22) | Charles River Laboratories Normal Range for C57BL/6 mice |
|------------------|----------|-----------------------|-----------------------------|----------------|-----------------------------|----------------------------------------------------|
| White Blood Cells (K/µL) | 3.5 ± 1.0 | 6.4 ± 0.8 | 5.2 ± 0.6 | 7.0 ± 1.3 | 7.8 ± 0.8 | 3.9 - 13.9 |
| Neutrophils (K/µL) | 0.9 ± 0.4 | 1.8 ± 0.2 | 1.5 ± 0.2 | 2.1 ± 0.4 | 2.1 ± 0.2 | 0.4 - 2.6 |
| Lymphocytes (K/µL) | 2.4 ± 0.7 | 4.0 ± 0.6 | 3.3 ± 0.5 | 4.3 ± 0.9 | 4.8 ± 0.6 | 2.9 - 10.9 |
| Monocytes (K/µL) | 0.2 ± 0.1 | 0.4 ± 0.1 | 0.4 ± 0.0 | 0.4 ± 0.1 | 0.4 ± 0.1 | 0.2 - 0.7 |
| Eosinophils (K/µL) | 0.0 ± 0.0 | 0.2 ± 0.1 | 0.1 ± 0.0 | 0.2 ± 0.0 | 0.2 ± 0.1 | 0.0 - 0.5 |
| Basophils (K/µL) | 0.0 ± 0.0 | 0.1 ± 0.0 | 0.0 ± 0.0 | 0.1 ± 0.0 | 0.1 ± 0.0 | 0.0 - 0.1 |
| Neutrophils (%) | 21.2 ± 4.8 | 28.9 ± 1.6 | 30.3 ± 1.6 | 30.9 ± 3.3 | 28.7 ± 1.7 | 7.4 - 22.7 |
| Lymphocytes (%) | 71.9 ± 6.3 | 59.6 ± 2.7 | 58.6 ± 2.2 | 60.0 ± 3.3 | 61.5 ± 2.9 | 70.2 - 87.8 |
| Monocytes (%) | 5.9 ± 2.0 | 7.6 ± 1.2 | 7.8 ± 0.9 | 6.0 ± 0.8 | 5.7 ± 0.8 | 2.2 - 7.1 |
| Eosinophils (%) | 0.8 ± 0.6 | 2.9 ± 0.8 | 2.5 ± 0.5 | 2.3 ± 0.3 | 3 ± 0.7 | 0.2 - 4.5 |
| Basophils (%) | 0.7 ± 0.1 | 1.0 ± 0.4 | 0.9 ± 0.3 | 0.9 ± 0.1 | 1.2 ± 0.3 | 0.0 - 1.3 |
| Red Blood Cells (M/µL) | 9.8 ± 0.5 | 8.7 ± 0.3 | 8.3 ± 0.6 | 7.4 ± 0.4 | 8.5 ± 0.3 | 7.4 - 11.5 |
| Hemoglobin (g/dL) | 14.0 ± 0.7 | 12.6 ± 0.4 | 11.8 ± 0.7 | 10.7 ± 0.4 | 12.1 ± 0.4 | 10.9 - 18.1 |
| Hematocrit (%) | 44.4 ± 2.2 | 37.9 ± 1.4 | 36.0 ± 2.2 | 33.5 ± 1.6 | 37.3 ± 1.5 | 37.2 - 58.0 |
| Mean corpuscular volume (fL) | 45.4 ± 0.6 | 43.6 ± 0.6 | 43.9 ± 0.9 | 45.3 ± 1 | 43.8 ± 0.6 | 42.6 - 55.6 |
| Mean corpuscular hemoglobin (Pg) | 14.3 ± 0.2 | 14.5 ± 0.2 | 14.4 ± 0.3 | 14.6 ± 0.3 | 14.2 ± 0.2 | 13.0 - 16.8 |
| Mean corpuscular hemoglobin concentration (g/dL) | 31.5 ± 0.2 | 33.5 ± 0.7 | 32.6 ± 0.7 | 32.2 ± 0.8 | 32.5 ± 0.4 | 26.0 - 35.9 |
| Red cell distribution width (%) | 18.4 ± 0.9 | 19.7 ± 0.3 | 19.7 ± 0.3 | 18.7 ± 0.3 | 19 ± 0.1 | 16.1 - 21.1 |
| Platelets (K/µL) | 648.6 ± 93.1 | 725.0 ± 70.2 | 713.9 ± 90.1 | 704.5 ± 85.7 | 653.2 ± 42.9 | 565.0 - 1849.0 |
| Mean platelet volume (fL) | 5.8 ± 0.0 | 5.6 ± 0.1 | 5.9 ± 0.2 | 5.7 ± 0.1 | 5.7 ± 0.1 | 4.3 - 5.6 |

Data are presented as mean ± sem
Supplementary Methods for Metabolomics

**Metabolite analysis (erythritol, sedoheptulose, and erythronic acid) by LC-MS/MS in urine.**

A novel method was created for the measurement of erythritol, sedoheptulose, and erythronic acid in urine. The details of quantitation by LC-MS/MS are described below, while validation statistics are described in the supplementary data.

**Preparation of standards, calibration curve, and QC’s**

Erythritol (Sigma cat# E7500), sedoheptulose (sigma cat# 07532), and erythronic acid (sigma: cat# 75025) standards were freshly prepared as 10mg/ml stocks using HPLC grade water. Meso-d6-erythritol (CDN Isotopes, cat# D-7720) and U-13C6-glucose (Cambridge Isotopes, cat# CLM-1396) were purchased for use as internal standards. 10 mg/ml stock solutions of U-13C6-glucose and d6-erythritol were prepared in HPLC grade water. For sedoheptulose, the 10mg/ml stock solution was diluted to a 10 mM calibrator stock solution, which was then serially diluted down to the lowest calibrator stock, which was 120 µM. For erythritol and erythronic acid, 6 mM calibrator stocks were serially diluted down to 120 µM. 10 µL of each calibrator stock was added to 30 µL of analyte stripped serum (Sigma, cat# SAE0012), which served as the matrix for the calibration curve and QC samples. We deemed this to be the most appropriate blank biological matrix available, with negligent levels of the metabolites of interest or other co-eluting interferences. The final concentrations of the 3 metabolites in the 40 µL (total volume) of spiked serum was the following (in µM): 30, 50, 75, 100, 150, 200, 400, 600, 800, 1000, 1250, 1500). For sedoheptulose, which had very high concentrations in the mutants, an additional 2 calibration points were added, with final concentrations of 2000 µM and 2500 µM. Final QC concentrations were as follows for all compounds: (lloq: 40 µM, low: 88 µM, medium: 500 µM, and high: 1375 µM). Four replicates of each QC level were prepared and extracted along with the calibration curve and the study samples in the same batch. The extraction solution contained: 150 µM of d6-erythritol and 20 µM 13C6-glucose dissolved in 95/5 ratio of methanol/water. Blank and double blank samples were also prepared within the same batch.

**Urine sample preparation and LC-MS based quantification**

To 40 µL of urine, added 200 µL of above extraction solution (ice cold). Similar amount was added to all calibrators. QC’s, and blank. 200 µL of 95% methanol was added to the double blank. Vortex for 30 seconds and place in -20°C for deproteinization. Spun at 13,000 x g for 15min at 4°C. Remove supernatant and place in fresh microcentrifuge tube. Dilute supernatant 1/10 in 75% methanol and inject 5 µL per run. Samples were injected on a Unison UK-Amino MF column (Imtakt USA), with dimensions of 150mm x 2mm (product # UKA25F). A column temperature of 60°C and flow rate of 0.4ml.min was used for all runs. A Shimadzu Prominence UHPLC (Shimadzu Corp.) was used coupled to an API 4500 (AB Sciex) mass spectrometer. Runs were collected in negative ionization mode (-4500V), with a CAD set to 9. Curtain gas of 40, GS1 and GS2 set to 40 and 70 respectively. All other parameters were optimized for each individual compound and are shown in the table below: The bolded entries were the mrm transitions used for quantification, while all others were used for confirmation.
### MS parameter table

| Q1   | Q3   | dwell time | name                | DP | EP | CE | CXP |
|------|------|------------|---------------------|----|----|----|-----|
| 121.0| 71.0 | 50         | erythritol_1        | -60| -4 | -15| -6  |
| 121.0| 59.0 | 50         | erythritol_2        | -60| -4 | -15| -5  |
| 121.0| 89.0 | 50         | erythritol_3        | -60| -4 | -12| -6  |
| 135.0| 75.0 | 50         | erythronic_acid_1   | -60| -4 | -17| -5  |
| 135.0| 117.0| 50         | erythronic_acid_2   | -60| -4 | -14| -5  |
| 209.1| 101.0| 50         | sedoheptulose_2     | -60| -4 | -15| -5  |
| 209.1| 119.0| 50         | sedoheptulose_1     | -60| -4 | -10| -5  |
| 209.1| 59.0 | 50         | sedoheptulose_3     | -60| -4 | -27| -5  |
| 127.1| 92.0 | 50         | Iso-erythritol      | -60| -4 | -12| -6  |
| 185.1| 92.0 | 50         | Iso-glucose         | -55| -10| -13| -7  |

The following mobile phases were used: Solvent A-0.7% formic acid, Solvent B- Acetonitrile. The following was the LC run schedule:

1. 0-1.0 min at 90% B
2. 1.0-2.0 min: 90%>80% B
3. 2.0-4.0 min: 80% B
4. 4.0-4.1 min: 80%>22%B
5. 4.1-7 min: 22%B
6. 7.0-7.1 min: 22%>2%B
7. 7.1-11.0 min: 2%B
8. 11.0-11.1 min: 2%>90% B
9. 11.1-20.0 min: 90% B

Peak integration, calibration curve and QC analysis, in addition to final determination of concentrations was performed using MuliQuant software (v3.02, AB Sciex). We used d6-erythritol internal standard for normalization of erythritol, while $^{13}$C$_6$-glucose showed good linearity for use in normalization of sedoheptulose and erythronic acid. Calibration curves and QC analysis results can be seen in supplementary figure 1 and supplementary table 1. Intra-day accuracy and precision data met standard guidelines for bioanalytical assays [32], which includes ≤ 20% deviation (accuracy and CV) for lloq and ≤15% for low, medium, and high QC levels.

**PPP, Glycolysis, and nucleotide analysis in liver by LC-MS/MS**

Sample preparation involved harvesting livers, followed by freeze-clamping under liquid nitrogen to preserve biological metabolite levels. Samples were then powderized under liquid nitrogen and 20 mg of each powder was weighed and was re-suspended in 1mL of extraction solution (40% methanol/40% acetonitrile/20% H$_2$O) that was supplemented with 40 μL 13C-labeled internal standards (Cambridge Isotype Laboratories) per mL solution. The isotope mixture has a 13-C labeled version of each metabolite quantified in the method, which are used for normalization.
Samples were homogenized using a Precilys homogenizer, place in a -20°C freezer for 1 hr for deproteinization, then spun at 13000 x g for 15 mins. The supernatant was then dried in a speed vac coupled to a cold trap, resuspended in 300ul of 60% methanol, and 10uL was injected onto each LC-MS run.

The LC-MS method used for PPP intermediates was adapted from Bajad et al. with several modifications. Primarily, instead of the use of an amino-propyl column, we used a polymeric amino column (Supelco aPHera NH2 polymeric column (15cm x 2mm, Cat#56400AST)), at a similar pH (~9.2 vs 9.4 in Bajad et al.). The polymeric amino column MS contamination from unstable silica at high alkaline conditions. Mobile phase A consisted of 2% Acetonitrile in water with 3mM ammonium bicarbonate and 30mM ammonium hydroxide. A Waters-Acquity liquid chromatography system (Waters Corporation, Milford, MA) was used for all runs, with a constant flow rate of 0.3 mL/min. The autosampler was cooled to 10°C. Mobile phase B consisted of acetonitrile. The LC-MS/MS run (42 minutes in length) consisted of the following gradient conditions: 1) t= 0-1.5min, 60% B; 2) t= 1.5-12.0 min, 60% > 0% B gradient; 3) t= 12.0-27.0 min, 0% B isocratic; 4) t= 27.0-42.0min, 60% B isocratic. An API 4500 triple quadrupole instrument (AB Sciex LLC. Foster City, CA) was used for data collection, in negative ionization mode, with a -4500 ionization spray voltage. Other MS parameters included: CUR: 40, GS1: 70 PSI, GS2: 40, and CAD: 9. Source heating gas temperature was set to 450°C. Similar MRM transitions were used for MS collection as previously published in Bajad et al. [33], but were further optimized for the API 4500 using purified standards (parameters optimized for individual standards include collision energy (CE), declustering potential (DP), entrance and exit potentials (EP/XP)). A scheduled MRM method was used for data acquisition to maximize the dwell times for each compound. Each compound had a minimum 20 ms dwell time. All data was integrated and quantified using MultiQuant software (AB Sciex, LLC). Calibration curves used for quantitation are included in the supplementary data (Supplementary Fig. 3). Accuracy and precision data for this method met guidelines described above (≤ 20% deviation [accuracy and CV] for lloq and ≤15% for low, medium, and high QC levels).