# Sulfated glycosaminoglycans and low-density lipoprotein receptor contribute to *Clostridium difficile* toxin A entry into cells

Liang Tao[1,2,8,9], Songhai Tian[1,2,9], Jie Zhang[1,2,9], Zhuoming Liu[2], Lindsey Robinson[2], Shin-Ichiro Miyashita[1,2], David T. Breault[3,6,7], Ralf Gerhard[4], Siam Oottamasathien[5], Sean Whelan[2], and Min Dong[1,2]

[1]Department of Urology, Boston Children's Hospital, and Department of Surgery, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, USA
[3] Division of Endocrinology, Boston Children's Hospital, Boston, MA 02115, USA
[4] Institute of Toxicology, Hannover Medical School, 30625 Hannover, Germany
[5] Department of Surgery and Pediatric Urology, University of Utah/Primary Children's Hospital, Salt Lake City, UT 84113, USA
[6] Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA
[7] Harvard Stem Cell Institute, Cambridge, MA 02138, USA
[8] Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Westlake University, Hangzhou, Zhejiang 310024, China

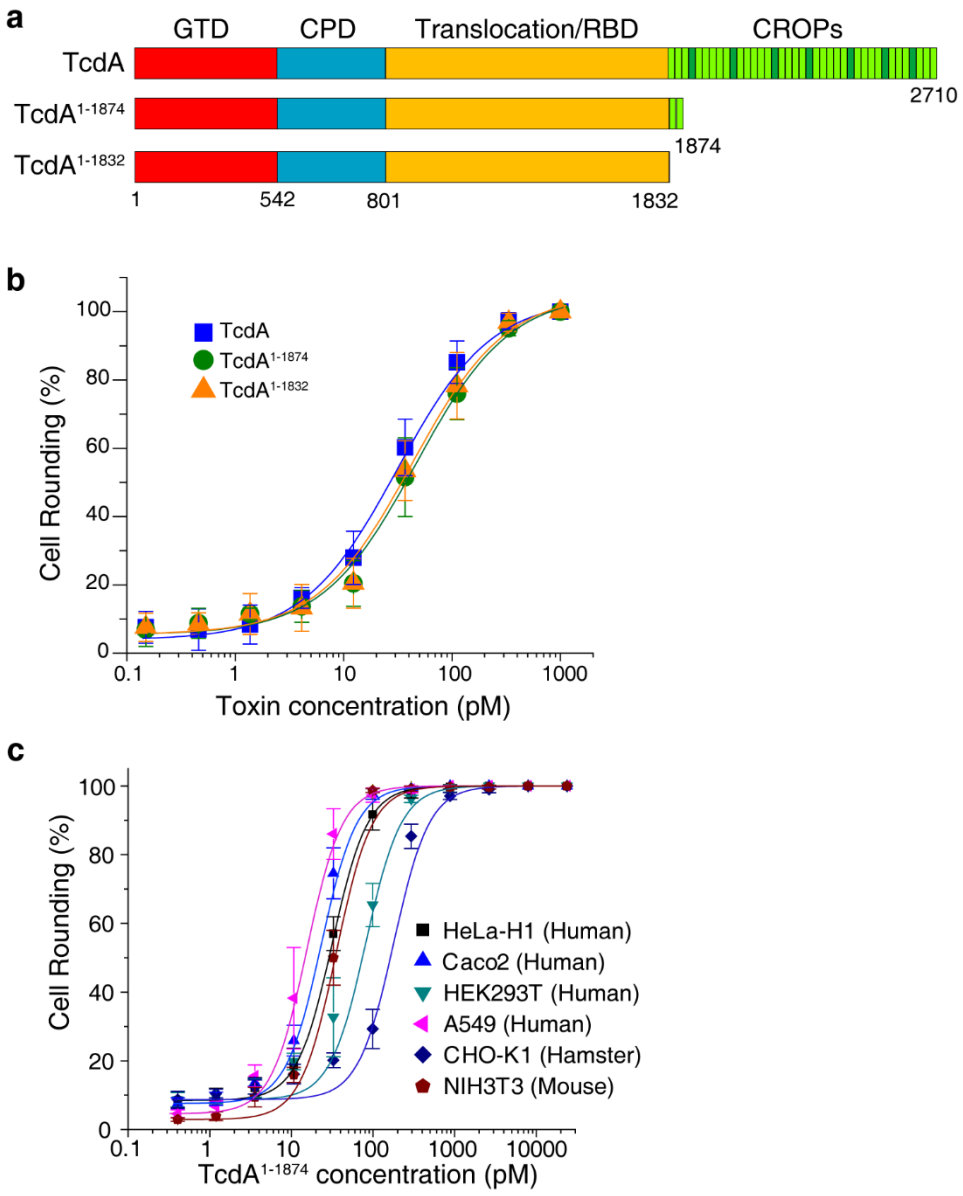
[9]These authors contributed equally to this work.

**This file includes:**

**Supplementary Figures 1-9**

**Raw Image Figure**

**Supplementary Data Legend**

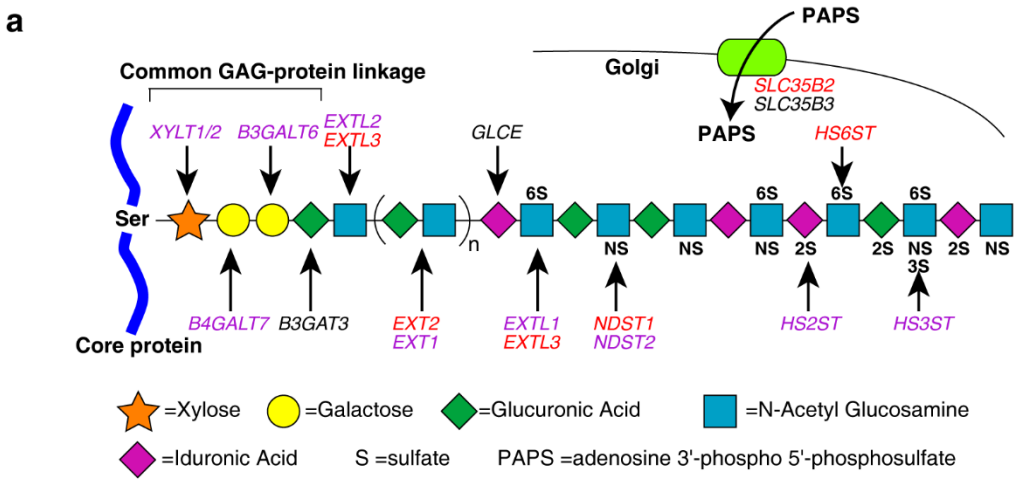


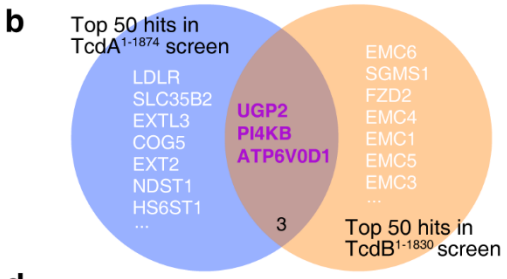

**Supplementary Figure 1. Two truncated TcdA (TcdA$^{1-1832}$ and TcdA$^{1-1874}$) lacking the CROPs retain potent toxicity on human cells.**

**a.** Schematic drawing of TcdA, TcdA$^{1-1874}$, and TcdA$^{1-1832}$. GTD: glucosyltransferase domain; CPD: cysteine protease domain; RBD: receptor binding domain; CROPs: combined repetitive oligopeptides.

**b.** The potency of the three TcdA variants on HeLa cells was determined using the cytopathic cell-rounding assays. The percentage of rounded cells were quantified, plotted, and fitted (data represent as mean ± s.d., experiments were repeated in triplicates, n=4 (biological replicates)).

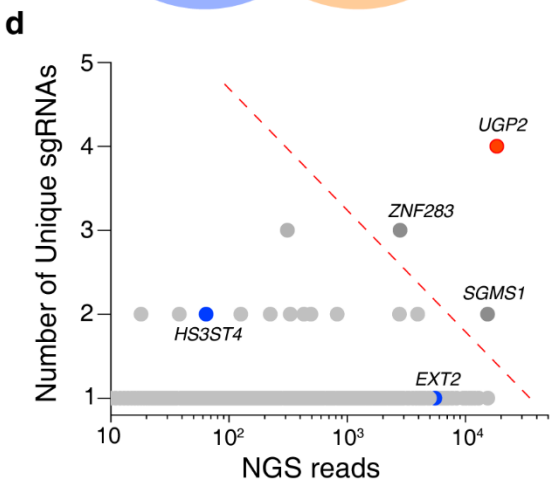
**c.** The sensitivities of HeLa, Caco-2, HEK293T, A549, CHO-K1, and NIH3T3 to TcdA$^{1-1874}$ were estimated using the cytopathic cell-rounding assays. The percentage of rounded cells were quantified, plotted, and fitted (data represent as mean ± s.d., experiments were repeated in duplicates, n=4 (biological replicates)).



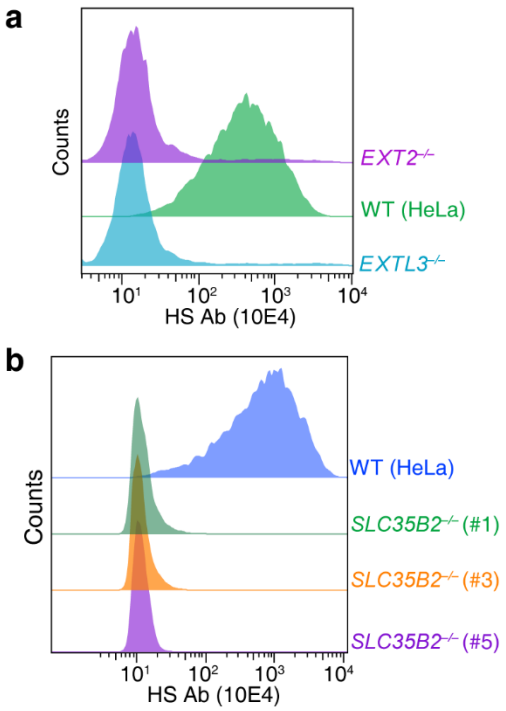

c

| Identified COG genes | Number of Unique sgRNA | NGS reads (Total=1294189) |
|---|---|---|
| *COG5* | 5 | 16182 |
| *COG6* | 4 | 1420 |
| *COG7* | 4 | 1245 |
| *COG8* | 2 | 410 |
| *COG1* | 2 | 225 |
| *COG3* | 1 | 1522 |
| *COG2* | 1 | 91 |
| *COG4* | 1 | 4 |




**Supplementary Figure 2. Schematic drawing of the HS synthesis pathways and analysis of top-ranked genes from CRISPR-Cas9 screens using TcdA$^{1-1874}$ and full-length TcdA.**
**a.** Schematic illustration shows genes involved in HS biosynthesis. Genes identified in R3 with more than 2 unique sgRNAs were marked in red, genes identified with 1 or 2 sgRNAs were marked in purple.
**b.** Comparison of top-50 ranked genes identified in genome-wide CRISPR-Cas9 screens for TcdA$^{1-1874}$ versus in our previous screen with TcdB$^{1-1830}$ (Ref.20). *UGP2, PI4KB,* and *ATP6V0D1* were the only three that overlapped between two screens.
**c.** All eight members of COG complex were identified in the TcdA$^{1-1874}$ screen (R3).
**d.** Genome-wide CRISPR-Cas9-mediated screen was carried out on HeLa cells for full-length TcdA. Genes identified after TcdA screening were ranked and plotted. The *y* axis is the number of

unique sgRNA for each gene. The *x* axis represents the number of sgRNA reads for each gene. The top-ranking genes marked in red (*UGP2*) and dark gray (*ZNF283* and *SGMS1*). Two identified genes involved in HS biosynthesis are marked in blue.

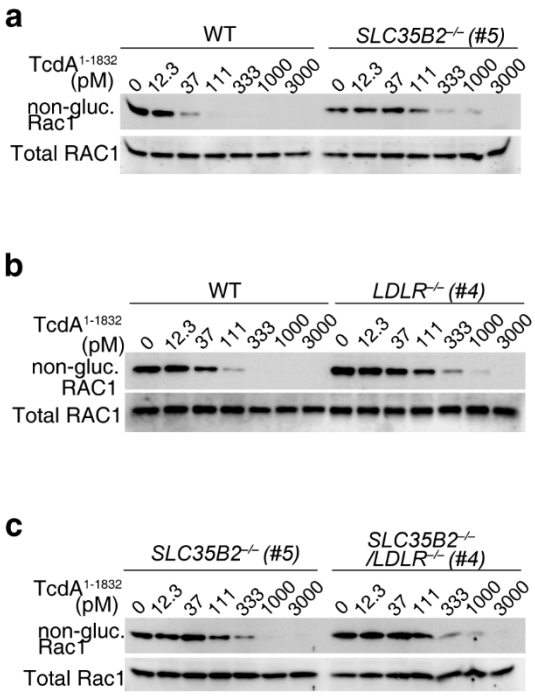



**Supplementary Figure 3. Cells lacking EXT2, EXTL3, or SLC35B2 lose surface HS expression.**

**a.** The levels of cell surface HS in WT, *EXT2*$^{-/-}$, and *EXTL3*$^{-/-}$ HeLa cells were assessed by flow cytometry using a HS-specific antibody (10E4). *EXT2*$^{-/-}$ and *EXTL3*$^{-/-}$ cells showed greatly reduced levels of HS.

**b.** The cell surface HS in three *SLC35B2*$^{-/-}$ cell lines was assessed by flow cytometry using 10E4 antibody. All of the three cell lines showed greatly reduced levels of HS.

For panels **a** and **b**, experiments were repeated at least twice independently with similar results.



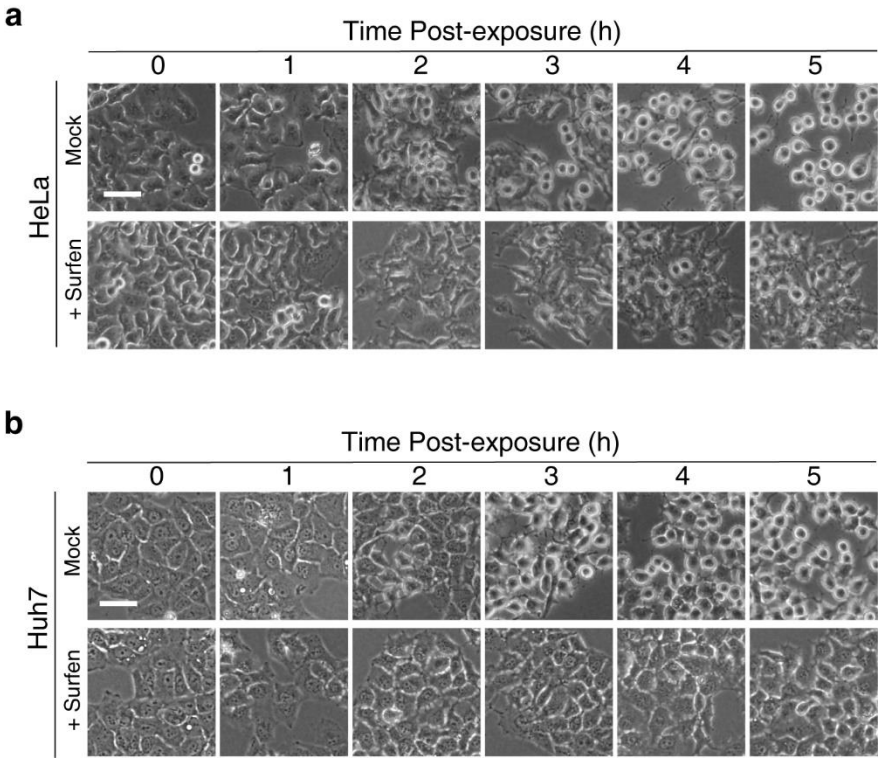

**Supplementary Figure 4. RAC1 glucosylation in HeLa WT cells and the indicated KO cells after exposure to TcdA$^{1\text{-}1832}$.**

**a.** The sensitivities of HeLa *SLC35B2*$^{-/-}$ (#5) cells versus the WT cells to TcdA$^{1\text{-}1832}$ were compared by analyzing the glucosylation level of RAC1 via immunoblot assays.

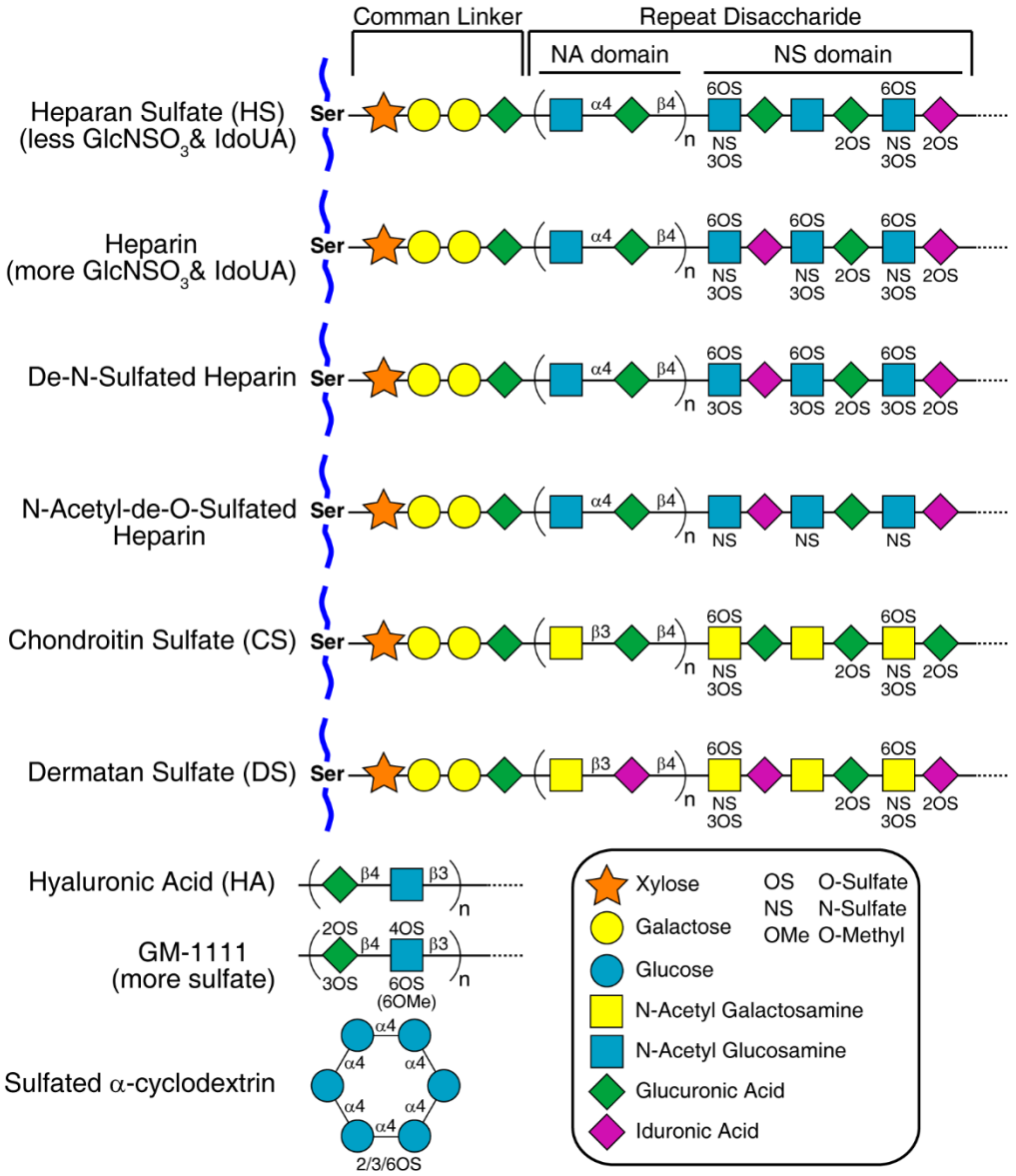
**b.** The sensitivities of HeLa *LDLR*$^{-/-}$ (#4) cells versus the WT cells to TcdA$^{1\text{-}1832}$ were compared by analyzing the glucosylation level of RAC1 via immunoblot assays.

**c.** The sensitivities of HeLa *SLC35B2*$^{-/-}$ (#5) cells versus *LDLR*$^{-/-}$/*SLC35B2*$^{-/-}$ (#4) cells to TcdA$^{1\text{-}1832}$ were evaluated based on the level of glucosylation of RAC1 protein.

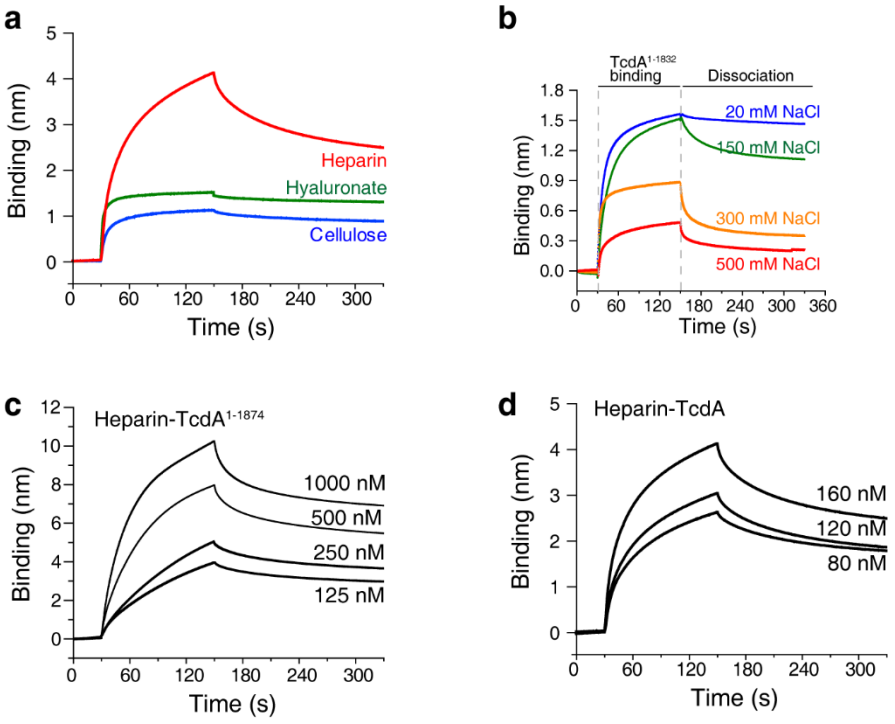
For panels **a**, **b** and **c**, experiments were repeated at least twice independently with similar results.




**Supplementary Figure 5. Surfen protected cells from TcdA$^{1\text{-}1832}$.**
**a.** Representative images showing that pre-incubation of surfen (5 nM) reduced the cytopathic effect cause by TcdA$^{1\text{-}1832}$ (2 nM) in HeLa cells. Scale bar, 50 µm.
**b.** Representative images showing that pre-incubation of surfen (5 nM) reduced the cytopathic effect cause by TcdA$^{1\text{-}1832}$ (2 nM) in Huh7 cells. Scale bar, 50 µm.
For panels **a** and **b**, experiments were repeated at least three times independently with similar results.



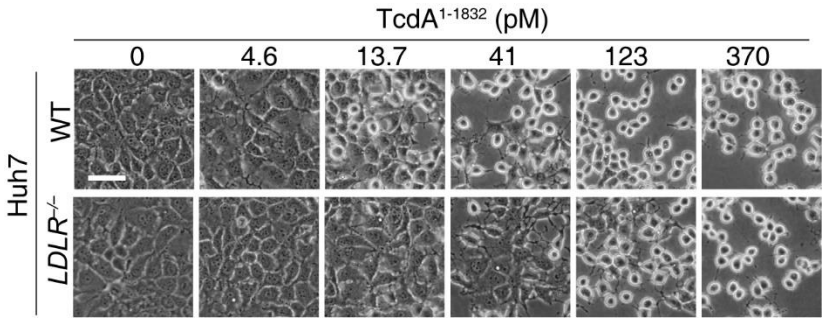

**Supplementary Figure 6. Schematic drawing of sGAGs and other sulfated molecules.**




**e**

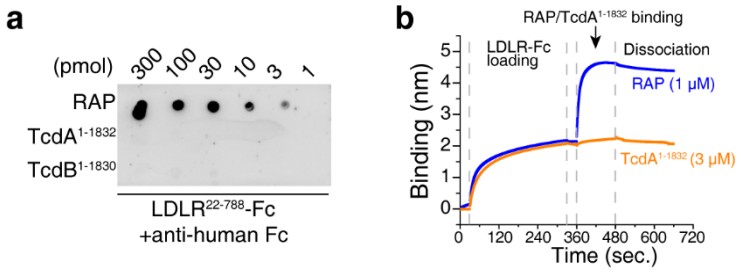
| | Heparin-TcdA$^{1-1874}$ | Heparin-TcdA |
|---|---|---|
| $K_{on}$ ($M^{-1}s^{-1}$) | 3.49E4 ± 7.97E2 | 2.41E5 ± 4.54E3 |
| $K_{off}$ ($s^{-1}$) | 2.98E-3 ± 6.68E-5 | 5.60E-3 ± 9.18E-5 |
| $K_D$ (M) | 8.55E-8 ± 3.87E-9 | 2.32E-8 ± 8.16E-10 |

**Supplementary Figure 7. Characterizing binding of TcdA$^{1-1874}$ to heparin using BLI assay.**
**a.** Direct binding of full-length TcdA to immobilized biotin-heparin was analyzed using BLI assays. Biotin-hyaluronate and biotin-cellulose were analyzed in parallel as controls.
**b.** Direct binding of TcdA$^{1-1874}$ to immobilized heparin was sensitive to salt concentrations.
**c-e.** The binding affinity between heparin and TcdA$^{1-1874}$ (panel **c**) or between heparin and full-length TcdA (panel **d**) in TBS buffer (20 mM Tris-Cl, 150 mM NaCl, pH 7.5) was measured by BLI assay. The binding kinetic data were listed in panel **e** (n=3, Centre values represent mean. Error bars represent ± s.d.).
Representative data was shown from two independent experiments.




**Supplementary Figure 8. Representative images showing that Huh7 *LDLR*$^{-/-}$ cells are less sensitivity to TcdA$^{1\text{-}1832}$ than WT cells.**

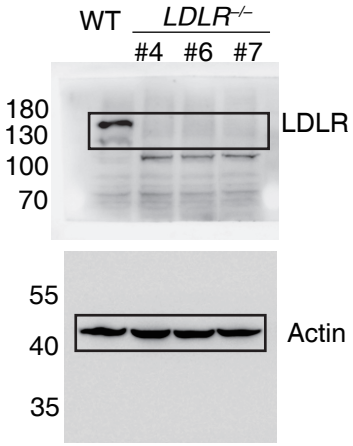
Huh7 *LDLR*$^{-/-}$ cells were less sensitive to TcdA$^{1\text{-}1832}$ compared with the WT cells using the cytopathic cell-rounding assays (12 h incubation). Scale bar, 50 μm. The experiments were repeated at least three times independently with similar results.



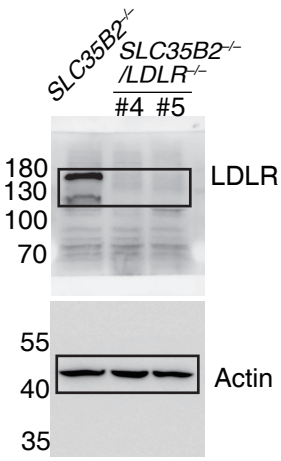

**Supplementary Figure 9. Binding of LDLR-Fc to TcdA$^{1-1874}$ was not detected.**
**a.** *In vitro* dot blot assays did not detect binding of LDLR-Fc to immobilized TcdA$^{1-1874}$. RAP and TcdB$^{1-1830}$ served as positive and negative control respectively.
**b.** BLI analysis showed that 1 μM RAP robustly bound to immobilized LDLR-Fc, while TcdA$^{1-1874}$ at up to 3 μM had no detectable binding.
The experiments were repeated at least three times independently with similar results.

# Raw Image Figures

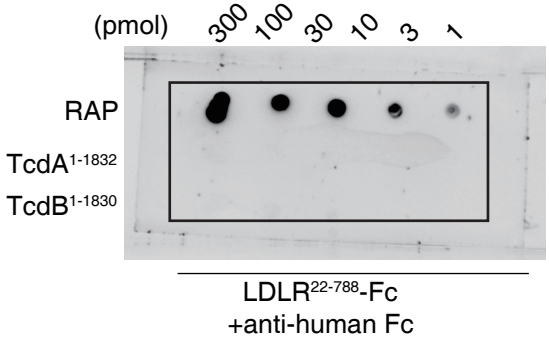
## Figure 3a



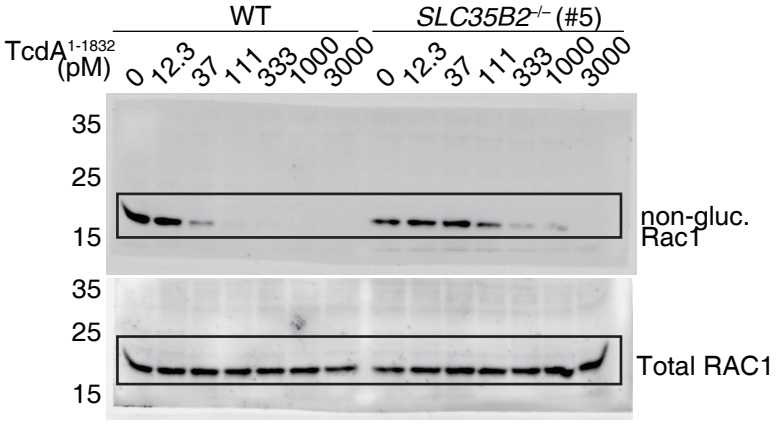

## Figure 4a



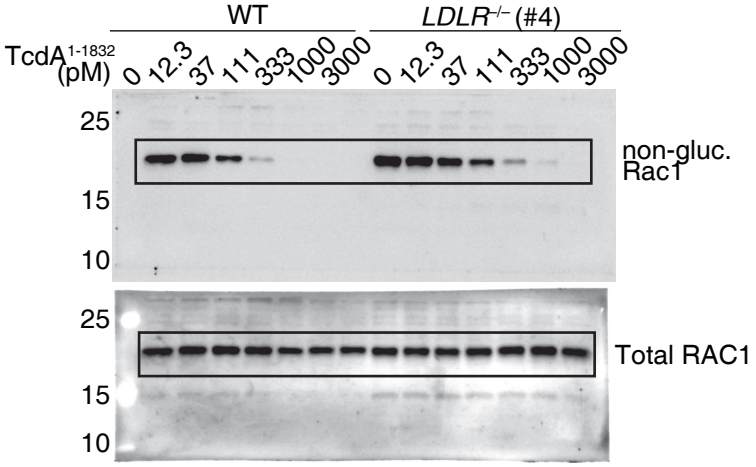

## Supplementary Figure 9a



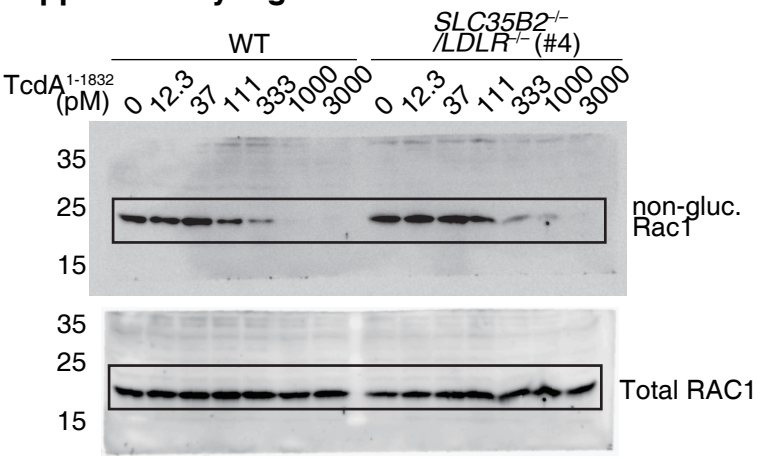

## Supplementary Figure 4a




## Supplementary Figure 4b




## Supplementary Figure 4c



**Supplementary Data (for Figure 1) | Lists of all sgRNA sequences and target genes identified from CRISPR-Cas9 screen.**

Column A: Gene name; Column B: sgRNA code name; Column C: sgRNA sequence; Column D: NGS reads.
Sheet 1: Library A screened with $TcdA^{1-1874}$.
Sheet 2: Library B screened with $TcdA^{1-1874}$.