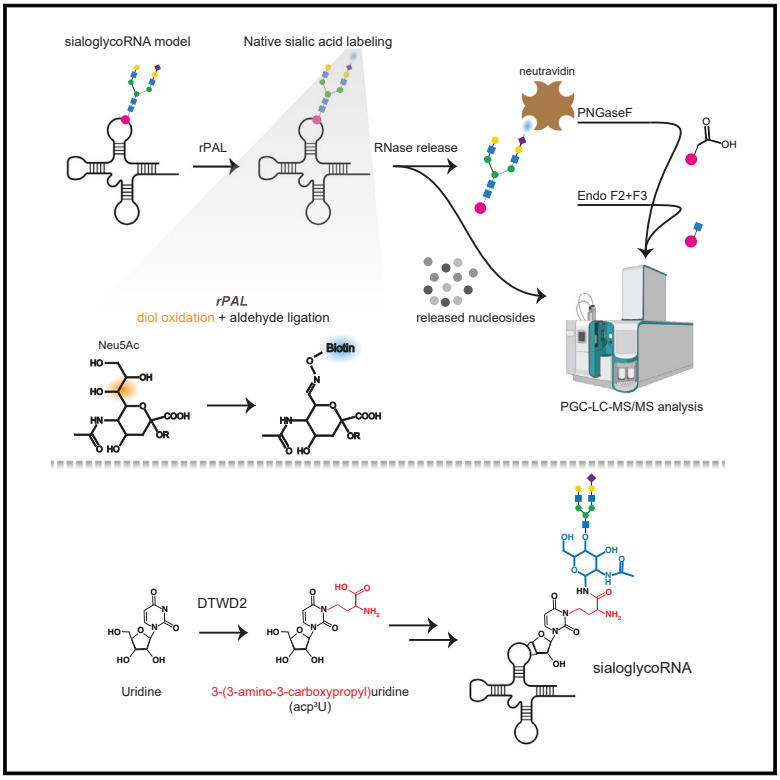
Cell

Short article

The modified RNA base acp^{3}U is an attachment site for N-glycans in glycoRNA

Graphical abstract



Highlights

- rPAL is a chemical tool to label and study native glycoRNAs
- acp^{3}U is an endogenous site of attachment for N-glycans on RNA
- Genetic loss of DTWD enzyme activity impacts glycoRNA biosynthesis

Authors

Yixuan Xie, Peiyuan Chai, Nicholas A. Till, ..., Carolyn R. Bertozzi, Benjamin A. Garcia, Ryan A. Flynn

Correspondence

bagarcia@wustl.edu (B.A.G.), ryan.flynn@childrens.harvard.edu (R.A.F.)

In brief

Characterization of sialic-acid-containing glycoRNAs identifies a modified base in RNA as the N-glycan attachment point.

Xie et al., 2024, Cell *187*, 5228–5237
September 19, 2024
https://doi.org/10.1016/j.cell.2024.07.044

CellPress

CellPress

Cell

Short article

The modified RNA base acp^{3}U is an attachment site for N-glycans in glycoRNA

Yixuan Xie,[1,7] Peiyuan Chai,[2,3,7] Nicholas A. Till,[4] Helena Hemberger,[2,3] Charlotta G. Lebedenko,[2,3] Jennifer Porat,[2,3] Christopher P. Watkins,[2,3] Reese M. Caldwell,[2,3] Benson M. George,[2,3] Jonathan Perr,[2,3] Carolyn R. Bertozzi,[4,5] Benjamin A. Garcia,[1,*] and Ryan A. Flynn[2,3,6,8,*]

[1]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO, USA
[2]Stem Cell Program and Division of Hematology/Oncology, Boston Children's Hospital, Boston, MA, USA
[3]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA
[4]Department of Chemistry and Sarafan ChEM-H, Stanford University, Stanford, CA, USA
[5]Howard Hughes Medical Institute, Stanford, CA, USA
[6]Harvard Stem Cell Institute, Harvard University, Cambridge, MA, USA
[7]These authors contributed equally
[8]Lead contact
*Correspondence: bagarcia@wustl.edu (B.A.G.), ryan.flynn@childrens.harvard.edu (R.A.F.)
https://doi.org/10.1016/j.cell.2024.07.044

SUMMARY

GlycoRNA consists of RNAs modified with secretory N-glycans that are presented on the cell surface. Although previous work supported a covalent linkage between RNA and glycans, the direct chemical nature of the RNA-glycan connection was not described. Here, we develop a sensitive and scalable protocol to detect and characterize native glycoRNAs. Leveraging RNA-optimized periodate oxidation and aldehyde ligation (rPAL) and sequential window acquisition of all theoretical mass spectra (SWATH-MS), we identified the modified RNA base 3-(3-amino-3-carboxypropyl)uridine (acp^{3}U) as a site of attachment of N-glycans in glycoRNA. rPAL offers sensitivity and robustness as an approach for characterizing direct glycan-RNA linkages occurring in cells, and its flexibility will enable further exploration of glycoRNA biology.

INTRODUCTION

Metabolic chemical reporters (MCRs) have been widely implemented to investigate various aspects of all biopolymers, including discovery-based efforts to identify biopolymers in different cellular locations and tissues and provide insight into their biogenesis pathways (reviewed in Pedowitz and Pratt[1]). We recently showed that a sialic-acid-specific MCR, N-azidoacetylmannosamine-tetraacylated (Ac_4ManNAz, Figures 1A and 1B), is able to incorporate into sialic-acid-containing N-glycans that are conjugated to small noncoding RNAs in mammalian cells.[2] These glycoRNAs, or, more precisely, sialoglycoRNAs (those with sialic acid), were produced from a wide variety of cell types, both *in vitro* and *in vivo*, and, importantly, were demonstrated to be presented on the external surface of living cells.[2] Missing from this work, as well as other recent reports supporting the glycoRNA discovery,[3–5] was a detailed characterization of the direct chemical linkage that enables RNA glycosylation.

Ac_4ManNAz enabled the glycoRNA discovery effort, but its limitations, including sub-stoichiometric and biased incorporation,[6] motivate the development of new tools. In addition, MCRs can only be taken up by metabolically active cells before their biosynthetic processing and incorporation into nascent biopolymers, like glycans. This limits MCR use in contexts where access to the organism or cells with a reporter is not possible, such as primary or patient samples. Glycan MCRs are also known to not fully saturate the sugar pool, resulting in the modification of only a portion of the total glycans.[7] Furthermore, glycan MCRs can be variably incorporated across glycoforms and cell types.[7] Additionally, the substitutions on MCRs (e.g., azide-group) can bias against their incorporation by certain glycosyltransferases.[8] Further characterization of glycoRNA species and insights into their biogenesis will necessitate the expansion of specific chemical tools that enable precise quantification and analysis of their dynamic cell display and can be performed on the native glycoRNA substrate.

Although methods exist to derivatize and label native glycoproteins and glycolipids, no such method exists for the modification of native sialoglycoRNAs. Here, we developed an RNA-optimized periodate oxidation and aldehyde labeling (rPAL) technique for the labeling of native sialoglycoRNAs and demonstrate its >25-fold increased sensitivity as compared with Ac_4ManNAz. We leveraged rPAL coupled to a highly sensitive analytical strategy called sequential window acquisition of all theoretical mass spectra (SWATH-MS) to identify natural glyconucleosides. Specifically, we focused on chemical and genetic characterization of the modified RNA base 3-(3-amino-3-carboxypropyl)uridine (acp^{3}U) as a site of attachment of N-glycans in RNA. We anticipate that our approach will

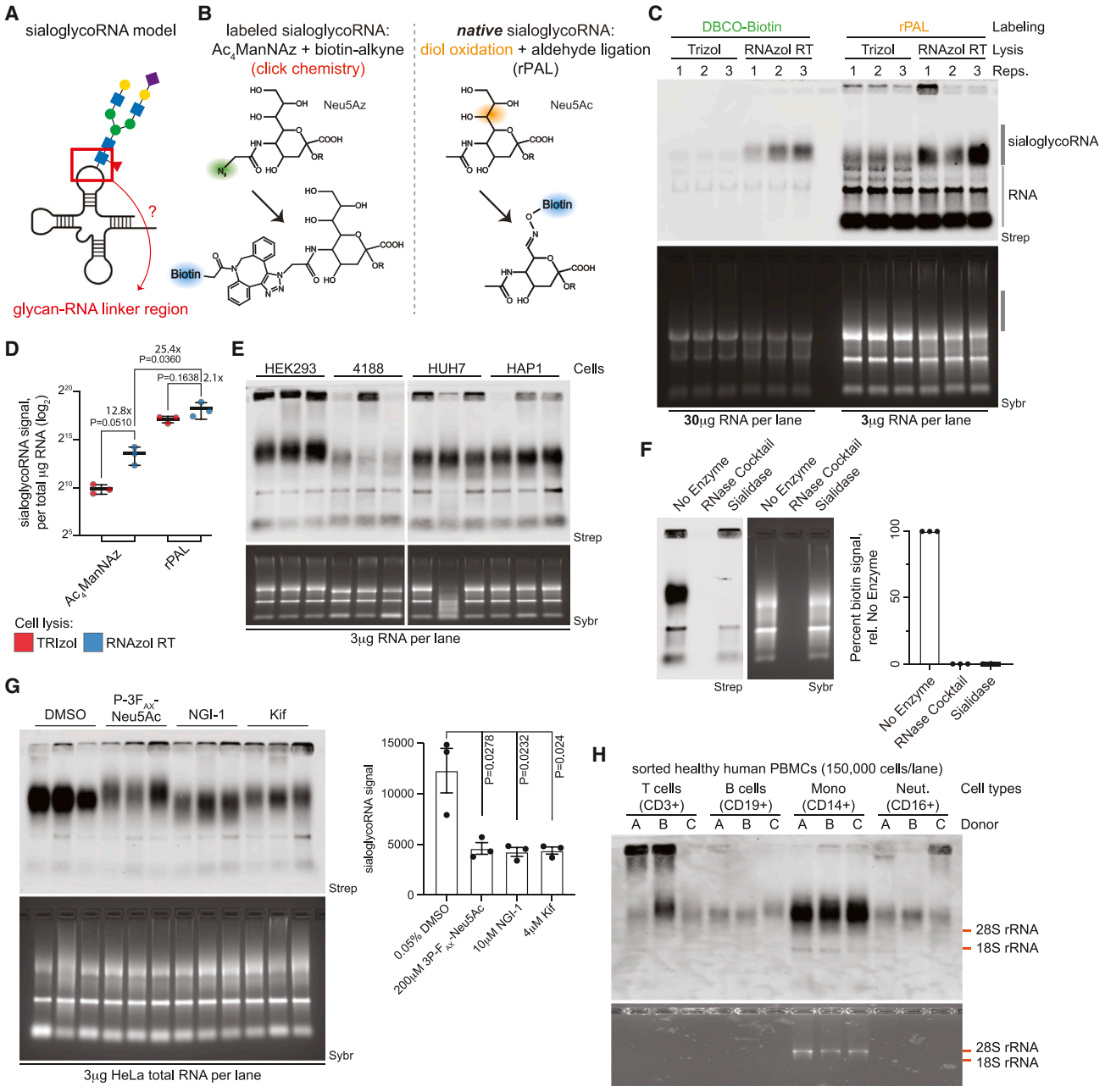


Figure 1. Selective oxidation and aldehyde labeling enables sensitive detection of native sialoglycoRNAs

(A) Cartoon of a sialoglycoRNA highlighting the glycan-RNA linker region, which has not been previously defined.

(B) Schematic of sialic-acid-labeled strategies, including Ac_4ManNAz labeling (left) or rPAL labeling (right).

(C) RNA blotting of HeLa total RNA, labeled and detected with Ac_4ManNAz and dibenzocyclooctyne-PEG4-biotin (DBCO-biotin), or native sialic acids paired with rPAL labeling. In-gel detection of total RNA with SybrGold (Sybr, bottom) and on-membrane detection of biotin (Streptavidin-IR800, top) is shown. The region labeled "sialoglycoRNA" was quantified.

(D) Quantification of data in (C). Each datapoint (biological triplicate) is displayed, with the standard error of the mean (SEM) and *p* values representing unpaired t tests.

(E) RNA blotting of HEK293, 4188, HUH7, and HAP1 total RNA labeled with rPAL as in (C).

(F) RNA blotting of HeLa total RNA that was treated *in vitro* with no enzyme, an RNase A and RNase T1 cocktail, or sialidase labeled with rPAL as in (C). Each datapoint (biological triplicate) is displayed with the SEM.

(G) RNA blotting of total RNA from HeLa cells treated with various chemical inhibitors. Inhibitors were added to cells in complete media for 24 h: 0.05% DMSO, 200 μM P-3_{FAX}-Neu5Ac, 10 μM NGI-1, or 4 μM kifunensine, after which RNA was collected and processed for rPAL labeling. Each datapoint (biological triplicate) is displayed with the SEM, and statistical analysis was performed using an unpaired t test.

(legend continued on next page)

accelerate the characterization of glycoRNAs structure and function across many biological contexts, as the identification of RNA modifications that template glycan addition will reveal insights into their biogenesis and functions.

RESULTS

Native sialoglycoRNA detection

Our previous work relied on metabolic conversion of an azide-modified sialic acid precursor (Ac_4ManNAz) to the corresponding azido sialic acid and its biosynthetic incorporation into the N-glycans of glycoRNA (Figure 1B). Motivated to move beyond the reliance of this MCR, we pursued an alternative approach to label and enrich glycoRNAs that leverages the periodate-mediated oxidation of vicinal diols to aldehydes and their subsequent ligation to amine-containing reagents or solid supports. This method has been used widely to characterize sialoglycolipids and proteins.[9–13] Paired with aminooxy-containing molecules, this reaction forms a stable oxime bond without requiring further derivatization. Although this approach is reasonably selective for sialic-acid-containing glycans, the presence of a 2′,3′ vicinal diol at the 3′-terminal ribose of RNA poses a possible side reaction when applied to glycoRNA. Although the 7′,8′ diol in sialic acid is quite reactive—and thus can be derivatized at physiological pH with short reaction times[12]—complete oxidation of ribose diols in RNA requires more acidic buffers and longer reaction times.[14] With this insight into the differential reactivity of sialic acid diols and ribose diols, we reasoned that, in a mixture of RNA nucleotides and sialic acid sugars, mild oxidation conditions could achieve selective sialic acid diol labeling (Figure 1B). We examined cell lysis methods, RNA extraction reagents (TRIzol vs. RNAzol RT), and specific precipitation parameters for RNA column cleanups (STAR Methods) to develop a robust protocol for high recovery of small RNA (Figures S1A–S1C).

To maintain a simple procedure, we set out to develop a reaction scheme that would both oxidize diols and ligate newly generated aldehydes (PAL) to a labeling reagent in the same reaction without purification. Conditions including pH, salt concentration, salt type, and temperature were screened using RNAzol RT-extracted HeLa cell total RNA (Figures S1D–S1F). An rPAL protocol (STAR Methods) includes a pre-blocking step with a free aldehyde reagent, which reduces background signal, as well as mucinase digestion (Figures S1G and S1H), over 2.5 h. Further development of the RNA northern blot transfer buffer conditions (pH, salt, and time, Figures S1I–S1L) resulted in enhanced transfer. Extracting RNA from Ac_4ManNAz labeled HeLa cells with TRIzol and either performing copper-free click[2] or rPAL, we found that rPAL generates approximately 150× the amount of signal (Figures 1C and 1D). We repeated this comparison from Ac_4ManNAz-labeled cells and, while there was not a significant difference in total RNA extracted with TRIzol vs. RNAzol RT (Figure S1A), we saw a 12.8× and 2.1× gain in glycoRNA signal with Ac_4ManNAz and rPAL, respectively (Figure 1D). Integrating both the rPAL and RNAzol RT RNA extraction, we can achieve at least 25-fold increased signal recovery per mass of RNA compared with an updated Ac_4ManNAz strategy that uses RNAzol RT extractions (Figure 1D). Although rPAL improves sensitivity of apparent high molecular weight (MW) glycoRNA species, it also induces background labeling, most notably within the 18S ribosomal RNA (rRNA) and the small RNA pool (Figure 1D and elsewhere). This is expected, given that the 3′ ends of RNAs should (mostly) contain a 2′,3′ vicinal diol, which can also undergo periodate-based oxidation.[14]

To demonstrate the flexibility of rPAL, we labeled total RNA stocks that were collected more than 3 years ago from cells treated with Ac_4ManNAz for 24 h prior to collection. rPAL readily detects sialoglycoRNAs from the four archived RNA stocks sourced from HEK293, 4188, HUH7, and HAP1 cells (Figure 1E). We can more directly interpret the levels of rPAL signal across each cell type due to differences in sialylation, while Ac_4ManNAz labeling requires consideration of metabolism and incorporation. Finally, to establish that the high MW signal is indeed sialoglycoRNA, we evaluated its sensitivity to enzymatic digestion. Incubation of purified RNA with RNase or sialidase, and subsequent rPAL labeling, results in total loss of all rPAL signal after RNase treatment but selective and complete loss of the high MW signal after sialidase treatment (Figure 1F), consistent with rPAL-mediated sialic acid labeling of sialoglycoRNAs.

rPAL characterization

To verify that the high MW signal is, indeed, small glycoRNAs, we used column-based separation of large and small RNAs; we found that the high MW rPAL signal, as well as the signal overlapping with the small RNA pool, is enriched with small RNAs but not the long RNAs (Figure S2A). Cell fractionation demonstrated that the high MW signal accumulates in crude cellular membrane extracts while it is depleted in the cytosolic fraction (Figures S2B and S2C). Additionally, and in line with our reported topology of Ac_4ManNAz-labeled sialoglycoRNAs, the rPAL-labeled high MW signal is sensitive to live-cell sialidase treatment. Treatment of live HeLa cells with *Vibrio cholerae* (VC) sialidase resulted in robust loss of sialoglycoRNA signal in as little as 15 min after addition of the VC sialidase (Figure S2D).

Because inhibition of N-glycosyltransferases resulted in loss of signal from Ac_4ManNAz-labeled sialoglycoRNAs,[2] we assessed how the rPAL signal from HeLa cells changes upon treatment with the following inhibitors: (1) P-3F_{AX}-Neu5Ac (sialic acid biosynthesis),[15] (2) NGI-1 (oligosaccharyltransferase [OST]),[16] and (3) kifunensine (α-mannosidase-I).[17] In all three cases, we observed loss of the rPAL signal (Figure 1G), consistent with our previous results.[2] P-3F_{AX}-Neu5Ac and kifunensine treatment, respectively, caused a larger shift in the apparent MW of the sialoglycoRNAs (Figure 1G), which is in line with the effects seen with Ac_4ManNAz. NGI-1 resulted in an apparent lower

(H) RNA blotting of total RNA labeled with rPAL from four sorted populations of human PBMCs, including CD19 (B cells), CD3 (T cells), CD14 (monocytes and macrophages), and CD16 (NK cells, monocytes, and neutrophils)-positive cells; RNA from approximately 150,000 cells was used for each lane.
See also Figure S1.

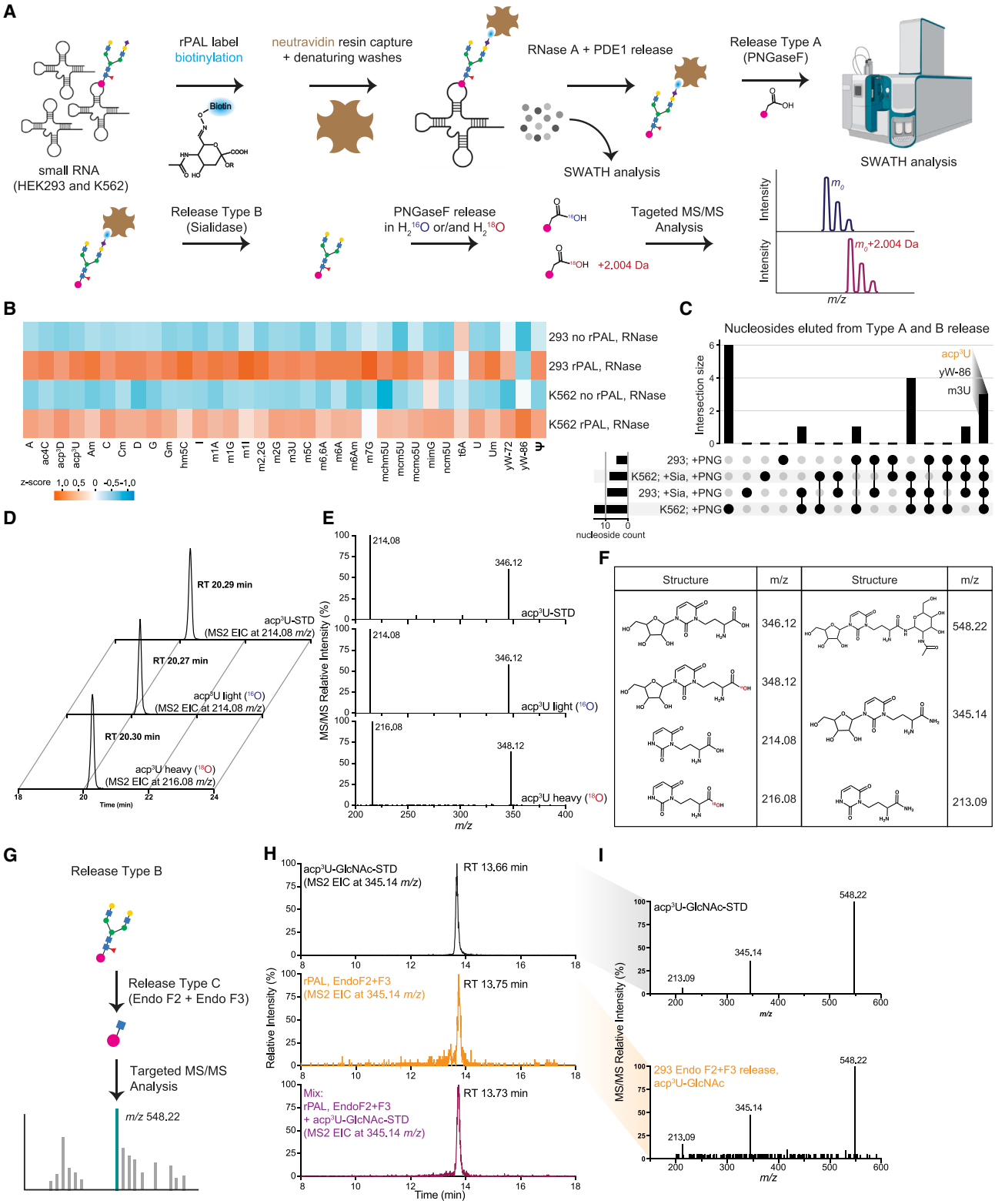


Figure 2. acp^{3}U is an endogenous template for RNA glycosylation

(A) Schematic of rPAL labeling, enrichment, RNase digestion, on-bead PNGaseF release (top, release type A), or on-bead sialidase release, followed by in-solution PNGaseF digestion (bottom, release type B) and SWATH-MS analysis.

(legend continued on next page)

MW smearing of the signal (Figure 1G), which was not previously seen with Ac_4ManNAz.

To demonstrate one utility of the significantly increased sensitivity of rPAL and the ability to access samples where MCR labeling would be challenging, we examined the sialoglycoRNAs from human peripheral blood mononuclear cells (PBMCs). We stained healthy human PBMCs with antibodies targeting CD19 (B cells), CD3 (T cells), CD14 (monocytes and macrophages), and CD16 (natural killer [NK] cells, monocytes, and neutrophils) and performed cell sorting (Figure S2E). We sorted 150,000 cells per tube per donor (three total donors) and labeled extracted RNA with rPAL. Across all four cell types and the three donors, we detected sialoglycoRNAs, with the sorted $CD14^+$ cells yielding the most total RNA as well as the strongest sialoglycoRNA signal (Figure 1H). Some samples from each of the cell types also displayed ultrahigh MW signal (Figure 1H), which is evident but less common in material isolated from cultured lines with higher cell numbers (e.g., Figure S2D; lanes 1, 3, and 4). We note that minor MW change could be seen between donors of T cells (donor 3) and B cells (donor 2) (Figure 1H). Taken together, these data demonstrate that rPAL labeling is highly sensitive and useful in low-input material contexts. Given the robustness of the protocol, we predicted it would be easily scaled up for large enrichments useful for *de novo* discovery of the putative glycan-RNA linkage.

Glyconucleoside identification with rPAL enrichment

In the broader context of RNA modifications, methylation and pseudouridylation are the most abundant RNA modifications; in particular, Watson-Crick face methylations (N-1-methyladenosine [m1A], N-3-methylcytosine [m3C], 5-methylcytosine [m5C]).[18] More chemically complex, but less abundant, modifications are also well represented in small RNAs, especially transfer RNAs (tRNAs).[19] For tRNA and rRNA post-transcriptional modifications (PTMs), we have nucleotide-scale resolution of a majority of RNA modifications.[18,19] Unlike these, N-glycosylation of RNA lacks a direct characterization of the chemical linkage between a particular atom of RNA (base, ribose, or phosphate) and the chitobiose core of an N-glycan.

To address this, we coupled large-scale rPAL labeling, high-efficiency capture, biochemical purification, enzymatic release (type A, B, and C below), and SWATH-MS[20] (Figure 2A). Specifically, we performed the rPAL labeling as described above and captured sialoglycoRNAs on neutravidin resin. After high-temperature (45°C) and stringent (4 M NaCl and 2 M GuHCl) washes, RNA was hydrolyzed with a cocktail of nucleases. Released nucleotides from this step could be putatively from regions of rPAL-labeled RNAs that are surrounding the glyconucleoside. In total, we found 34 unique nucleosides eluted across the three elution types, including unmodified A, C, U, and G, from HEK293 and K562 cells (Figure 2B; Table S1). Subsequently, to release nucleotides that are connected to N-glycan core structures, we treated the resin with PNGaseF (Figure 2A, type A). To release intact glyconucleosides, we instead subjected the final material to a sialidase elution followed by off-bead PNGaseF treatment (Figure 2A, type B). We compared the abundance of each nucleoside between control and rPAL-labeled RNA: we found that rPAL-enriched RNA from HEK293 cells released 4 modified nucleosides from type A releases and 7 modified nucleosides from type B releases, while rPAL-enriched RNA from K562 cells released 15 and 7 modified nucleosides from type A and B releases, respectively (Figure 2C; Table S1). When examining the structures of these released nucleosides, we noticed that a number of them contained carboxylic acid groups (7-aminocarboxypropylwyosine [yW-72]; 7-aminocarboxypropyldemethylwyosine [yW-86]; acp^3U). This was of interest because the products of PNGaseF cleavage of N-glycoproteins are asparagine to aspartic acid diagnostic "scars."[21]

Of acp^3U, yW-72, and yW-86, acp^3U was most enriched across both HEK293 and K562 cells and in both type A and B releases. Further, its synthesis has been described,[22] can be incorporated into oligonucleotide synthesis,[23] is found in the core region of bacterial and eukaryotic tRNAs,[24–26] and has several functional roles.[27,28] To confirm that our identification of the acp^3U nucleoside was a result of PNGaseF cleavage and not contamination of an endogenous modified nucleoside, we repeated the PNGaseF digestion in the presence of heavy ($H_2{}^{18}O$) or light ($H_2{}^{16}O$) water from the sialidase-eluted material, as previously developed for determining the glycosite on glycoproteins.[29] We expected that this would result in PNGaseF-released acp^3U having a mass shift of 2.004 *m/z* pair in MS1 level while displaying another paired pyrimidine ion during the tandem mass spectrometry (MS/MS) (Figure 2A, bottom). Examination of the liquid chromatogram (LC) retention time of the heavy- and light-cleaved material demonstrated overlapping peaks at 20.3 min (Figure 2D). MS/MS of these peaks revealed the expected parent masses and fragmentation patterns of heavy and light acp^3U (Figure 2E). To validate the identity of acp^3U, we used a synthetic standard. Commercially available acp^3U exists as a diastereomeric mixture consisting of the natural amino acid stereochemistry and the unnatural epimer

(B) Heatmap analysis of the nucleosides identified by SWATH-MS from HEK293 and K562 cells after RNase digestion. *Z* scores were calculated for each nucleoside between samples and colored from −1.0 to 1.0.

(C) Upset plot intersection of the nucleosides identified by SWATH-MS from HEK293 and K562 cells from both the RNase digestion and the on-bead PNGaseF release experiments.

(D) Extracted ion chromatogram (EIC, specific m/z shown) from the liquid chromatogram (LC) of acp^3U-STD, type B digestion with light (^{16}O) water, and type B digestion with heavy (^{18}O) water.

(E) Tandem mass spectrometry (MS/MS) analysis of the three peaks shown in (D). m/z values for the two major peaks in each trace are annotated.

(F) Chemical structures of relevant molecules.

(G) Schematic of the release type C.

(H) EIC (specific m/z shown) from the LC of acp^3U-GlcNAc-STD, Endo F2 + F3 released material from rPAL-enriched RNA, and a mixture of these two samples (mix).

(I) MS/MS analysis of two peaks shown in (H). m/z values for the three major peaks in each trace are annotated.

See also Figure S2.

(Methods S1). Stereoisomerically pure acp^{3}U was prepared from the corresponding enantiopure homoserine derivative through a Mitsunobu coupling step (Methods S1). This synthetic acp^{3}U standard (acp^{3}U-STD) exhibited a high-performance liquid chromatography (HPLC) retention time matching that of the cell-derived material (Figure 2D; Methods S1). MS/MS fragmentation of the acp^{3}U-STD, light, and heavy PNGaseF-released material resulted in peaks as expected (Figures 2E and 2F), with the heavy released material shifted by 2 Da. These data suggest that we do indeed observe PNGaseF-released acp^{3}U from rPAL-enriched RNA.

To obtain direct evidence of a glyconucleoside, we repeated the process as outlined in Figure 2A (type B release); however, we performed an *in vitro* digestion with a cocktail of Endo F2 and Endo F3 (type C release, Figure 2G) to release N-glycoconjugates, with the first GlcNAc still connected to the modified polymer. We predicted that a carboxamide version of acp^{3}U could serve as a template for N-glycosylation and, thus, synthesized an acp^{3}U-GlcNAc standard (acp^{3}U-GlcNAc-STD, Methods S1) with a predicted mass of 548.20 *m/z* (Figures 2F and 2G) to examine this hypothesis. We compared the elution profile of the acp^{3}U-GlcNAc-STD to that of released material from rPAL-enriched RNA, eluted with Endo F2 and Endo F3 (Figure 2H), and found co-eluting peaks at 13.66 and 13.75 min, respectively. MS/MS analysis of these peaks demonstrated the expected fragmentation patterns of acp^{3}U-GlcNAc (Figures 2I and 2F). To additionally confirm the elution profile of acp^{3}U-GlcNAc from the RNA sample, we spiked the acp^{3}U-GlcNAc-STD material into the rPAL-enriched Endo F2 + F3 cleaved material. This "mix" sample showed the same elution profile as the rPAL-enriched Endo F2 + F3 cleaved material alone (Figure 2H).

Characterization of DTWD2 enzyme activity on glycoRNAs

The release of the modified nucleoside acp^{3}U after PNGaseF treatment of cellular glycoRNA provided evidence that PNGaseF directly acts on glycoRNAs. Digestion with PNGaseF could lead to less total rPAL signal or MW shifts, given modulation of the glycoforms attached to the small RNAs. We therefore applied PNGaseF to total RNA from U2OS cells, performed rPAL labeling, and evaluated the resulting material by gel analysis. PNGaseF treatment did not statistically significantly impact the levels of rPAL intensity (Figures 3B and 3C) (however, there is a slight decrease in all cells), although there was a highly significant and robust shift down in the MW of the rPAL signal (Figures 3B and 3D). We observed this downshifted rPAL signal from PNGaseF cleavage across RNA sources from three other cell lines (Figures S3A–S3C). The enzymatic digestion of RNA by PNGaseF in these experiments was restricted to 60 min so as not to impact the integrity of the total RNA (to avoid misinterpretation of intensity changes due to simple loss of RNA signal) while the PNGaseF treatment used to liberate material for SWATH-MS analysis was allowed to proceed for 16 h. To confirm the direct action of PNGaseF on proposed N-glycan-RNA linker, we used this acp^{3}U-GlcNAc-STD to evaluate the enzymatic activity of PNGaseF *in vitro*. Incubating acp^{3}U-GlcNAc-STD in buffer with PNGaseF resulted in ~7-fold more release of acp^{3}U as compared with a buffer-only reaction (Figure S3D).

To further investigate aspects of the linkage, with a focus on acp^{3}U biosynthesis as the source for sialoglycoRNA production, we leveraged recent work describing the human DTWD family as responsible for installing acp^{3}U in tRNAs.[28] We generated two knockout (KO) clones in U2OS cells of DTWD2, which has been reported to exist in both the nucleus and cytosol.[28] Analysis of the genomic DNA from the two isogenic clones demonstrated genomic perturbations expected to result in gene disruption (Figure S3E). We isolated small RNA from the wild-type (WT) and DTWD2 KO cells and performed the SWATH-MS analysis of the resulting nucleosides to unbiasedly evaluate how the loss of DTWD2 impacted the nucleoside distribution. We detected 19 nucleosides across all three cell lines (Figure 3E; Table S2) and found that acp^{3}U and acp^{3}D (D = dihydrouridine) were the only modifications to consistently change upon DTWD2 KO; in both cases, we found a loss of the total levels of acp^{3}U and acp^{3}D upon DTWD2 KO (Figures 3F and S3F). We next examined the effect of DTWD2 loss on rPAL signal. Loss of DTWD2 results in a consistent downshifting in the rPAL signal as well as a minor (~10%) loss in total signal intensity (Figures 3G–3I). Thus, the rPAL phenotype of *in vitro* cutting of N-glycans off RNA with PNGaseF mirrors that of the genetic loss of acp^{3}U after DTWD2 deletion.

DISCUSSION

Here, we describe the identification of acp^{3}U as an attachment site for N-glycans on tRNA (Figure 3A), which was facilitated by the development of the rPAL method. rPAL coupled to SWATH-MS enabled the identification of a direct molecular linker between an N-glycan and an RNA, furthering the initial discovery of glycoRNAs[2] and laying a path for examining other chemical mechanisms of RNA glycosylation.

A series of carboxylic-acid-containing nucleosides were eluted from rPAL-enriched RNA with PNGaseF, including acp^{3}U, yW-72, and yW-86 (Figure S3H); however, acp^{3}U was most consistently observed and was quantitatively most enriched in the cell lines examined. Coupled to our genetic KO data of the DTWD2 enzyme and our enzymatic digestion of N-glycans with PNGaseF, our data strongly support a model where acp^{3}U is one of the primary RNA modifications that enables cells to N-glycosylate RNA. We therefore propose a biosynthetic process for generation of cell-surface sialoglyco-tRNAs (highlighting tRNAs, the currently best-characterized small RNAs modified with acp^{3}U) bearing N-glycans on acp^{3}U residues as shown in Figure 4. In this model, tRNAs are transcribed and modified with acp^{3}U in the nucleus and/or cytosol, which is then converted to the carboxamide form by as-yet unknown enzymes. Translocation into the endoplasmic reticulum (ER) lumen would then situate the modified tRNA for modification by OST. Notably, Ren et al. recently identified several small structured RNAs within the ER lumen, including tRNAs and other RNAs that we previously identified within the glycoRNA pool.[30] Trafficking through the secretory pathway accompanied by N-glycan trimming and branch extensions would then produce mature sialoglycoRNAs on the cell surface.

In proteins, N-glycosylation occurs on asparagine side chains at the nitrogen atom of the carboxamide functionality.

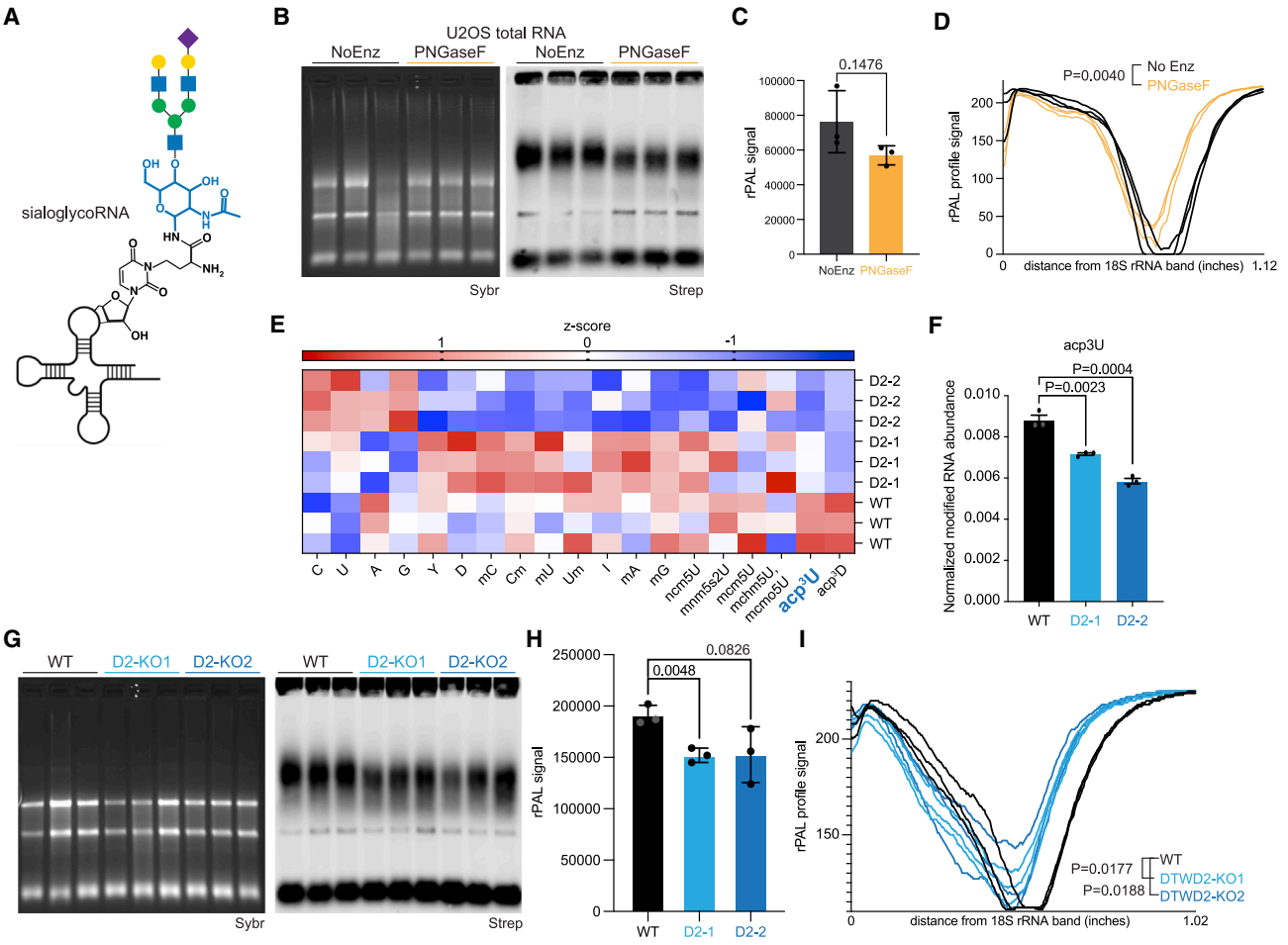


Figure 3. The acp^{3}U biosynthetic pathway contributes to glycoRNA production

(A) Model of a sialoglycoRNA.

(B) RNA blotting of U2OS total RNA that was subjected to PNGaseF digestion after isolation, rPAL labeling, in-gel detection of total RNA (Sybr), and on-membrane detection of biotin (Strep).

(C) Quantification of the sialoglycoRNA signal in (B). Statistical assessment was accomplished by an unpaired t test, *p* value shown.

(D) Plot profile analysis of the six lanes in (B), drawn upward from the 18S rRNA band in the Strep image using ImageJ. Statistical assessment of the differences in the histogram profiles was assessed using a bootstrapping method (STAR Methods), *p* values shown.

(E) Heatmap plot of *Z* scores of the nucleoside abundances (normalized for total nucleoside intensity) from small RNA isolated from the three U2OS cell lines and analyzed by SWATH-MS.

(F) Quantification of the normalized levels of acp^{3}U in (E), t test performed as in (C).

(G) rPAL blotting of total RNA samples extracted from U2OS wild-type (WT) or two individual knockout clones of the DTWD2 gene.

(H) Quantification of the sialoglycoRNA signal from each lane in (G), t test performed as in (C).

(I) Plot profile analysis of the nine lanes in (G), drawn upward from the 18S rRNA band in the Strep image using ImageJ. Bootstrapping test performed as in (D).

See also Figure S3.

Analogously, we speculate that glycosylation of acp^{3}U would involve prior formation of a carboxamide functionality from the acp^{3}U side chain carboxylic acid. Currently, only one RNA modification, 5-carbamoylmethyluridine (ncm5U), is known to possess a carboxamide group.[19] However, ncm5U is generated from 5-carboxymethyluridine (cm5U),[31] indicating that an amidating enzyme must exist and that other RNA modifications with carboxylic acids may be subject to amidation, making them compatible with N-linked glycosylation. Identification of a putative acp^{3}U amidation pathway is an immediate future goal. Beyond uracil, we found yW-86, a hyper-modified G, as also eluted with PNGaseF (Figure 2C; Table S1), which provides an additional putative site for N-glycans in RNA. yW-86 could offer a linkage mechanism for the putative sites of glycosylation we initially identified via sequencing.[32]

More broadly, the biosynthesis of sialoglycoRNAs observed in many cell types[2–5] suggests that the biochemical activity should be well conserved. In addition, acp^{3}U is found across bacterial and eukaryotic RNAs,[24–26] which opens the exciting possibility that glycosylated RNAs may exist in non-mammalian species. Despite its conservation, little is known about the cellular role of acp^{3}U: structural analysis suggests it could bind Mg^{2+} [27] but

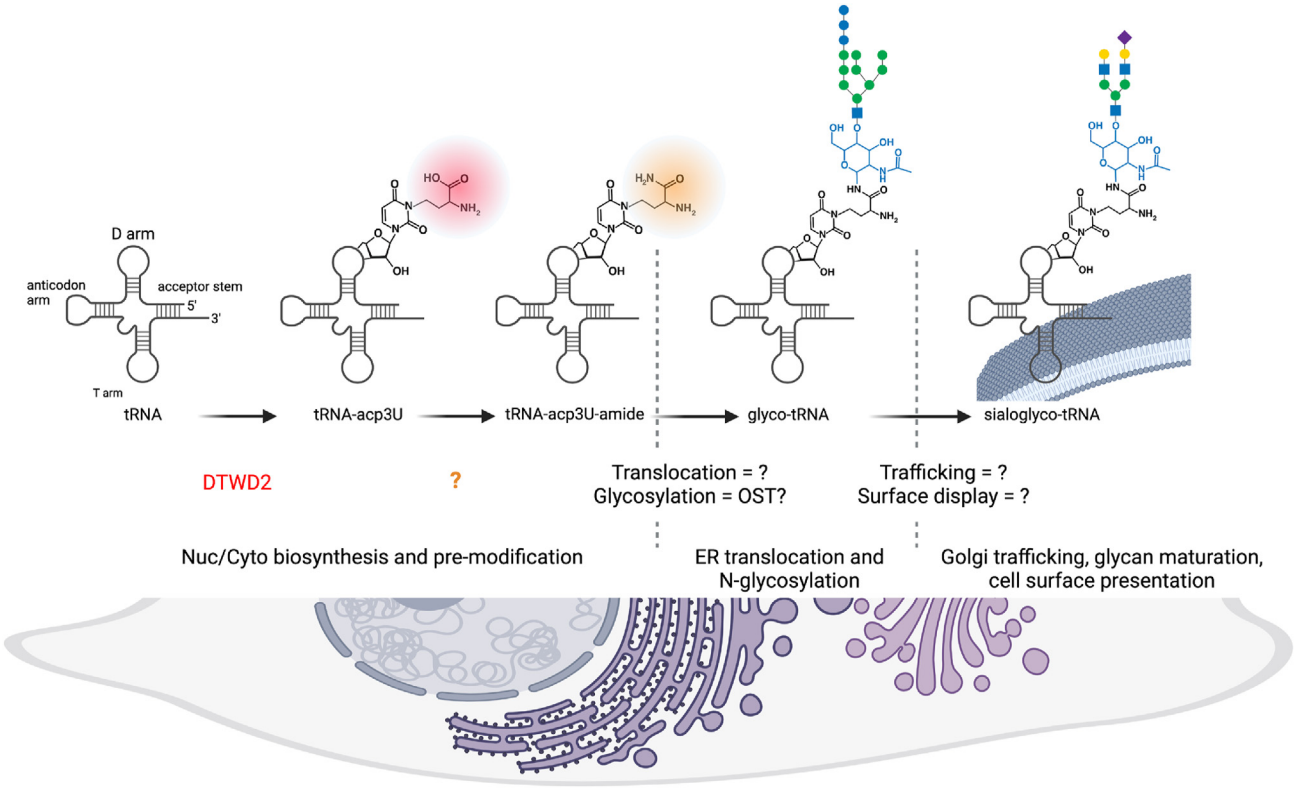


Figure 4. Proposed biosynthetic pathway for generation of cell-surface sialoglycoRNAs

(A) Cellular tRNAs are synthesized in the nucleus and can be modified there or in the cytosol to create acp^{3}U residue. A subsequent conversion to the carboxamide functionality, by an as-yet unknown enzyme, then allows translocation into the ER lumen. Once in the ER lumen, the carboxamide form of acp^{3}U would then enable modification by OST for N-glycosylation. Further trafficking through the secretory pathway, accompanied by N-glycan trimming and branch extensions, would then produce mature sialoglycoRNAs on the cell surface. Created with BioRender.com.

does not impact the conformation of the ribose[22]; however, mutations in the bacterial acp^{3}U biogenesis (TapT enzyme) cause growth defects in continuous heat stress, and double KOs of DTWD1 and DTWD2 cause growth defects in HEK293T cells.[28] In this context, we speculate that one of the important cellular roles for acp^{3}U is to provide a site for glycosylation by OST. Consistent with this, we and others have found oligo-mannose and hybrid N-glycans attached to RNA,[33] suggesting that immature N-glycans are initially added and then processed on the RNA template. Together, identifying a specific chemical linkage between N-glycans and tRNA brings us one step closer to understanding the biosynthesis and biological functions of glycoRNA.

Limitations of the study

The development of rPAL affords greater sensitivity and flexibility than most glycan MCRs; however, rPAL remains focused on sialic-acid-containing glycoRNAs, which may be an incomplete survey of all cellular glycoRNAs. Using a newly developed MS-based approach called glycanDIA, we have found asialoglycans on RNA,[33] which highlights a need for additional biochemical strategies to characterize glycoRNAs not dependent on sialic acid. Considering more broadly RNA PTMs and the reported possible synergy between DTWD1 and DTWD2 in depositing acp^{3}U on tRNA, additional investigation may be needed to establish the relative contribution of each enzyme to the steady-state production of sialoglycoRNAs. Separately, here, we proposed that the acp^{3}U linkage is likely on tRNAs, which we previously reported as among the class of RNAs that are modified with N-glycans.[2] The focus on tRNA here is due to the established knowledge of acp^{3}U modification in tRNA, although it is possible that acp^{3}U may be incorporated in other (specific evidence in rRNA exists[34,35]) RNAs and could therefore also be a template for glycosylation of other RNA classes. Further work identifying acp^{3}U in other RNA classes would be important to better contextualize the scope of this modification and its possible templating of glycans in those RNAs.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Mammalian cell culture
- METHOD DETAILS

- Chemical inhibitors and metabolic chemical reporters
- RNA extraction and enzymatic cleanups
- In vitro and On Cell enzymatic digestions
- Isolation of crude cellular membranes
- Periodate oxidation and aldehyde ligation for glycoRNA labeling
- glycoRNA blotting
- Peripheral blood mononuclear cell staining and flow cytometry sorting
- rPAL labeling, glycoRNA capture, and enzymatic release
- Porous graphitic carbon liquid chromatography mass spectrometry analysis
- LC-MS/MS analysis
- In vitro PNGaseF treatment of acp^{3}U_GlcNAc
- Data analysis
- Generation of knockout U2OS cell lines
- rPAL profile analysis
- Synthetic procedures

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.07.044.

ACKNOWLEDGMENTS

We thank Kayvon Pedram, Stacy Malaker, Maurice Wong, Kevin Janssen, Tiandi Yang, Janelle Sauvageau, and members of the Flynn lab for helpful comments and discussions. We also thank Melissa Gray for the expression and purification of VC sialidase. This work was supported by grants from Burroughs Wellcome Fund, Career Award for Medical Scientists (R.A.F.); The Rita Allen Foundation (R.A.F.); National Institutes of Health under award numbers GM151157 (R.A.F.), GM058867 (C.R.B.), AI118891 (B.A.G.), and HD106051 (B.A.G.); the Herchel Smith-Harvard Undergraduate Science Research Program (R.M.C.); and an NIH F32 Postdoctoral Fellowship (N.A.T.).

AUTHOR CONTRIBUTIONS

R.A.F. conceived the project. R.A.F., B.A.G., and C.R.B. supervised the project and obtained funding. H.H., P.C., C.G.L., R.M.C., J. Porat, and R.A.F. performed RNA labeling optimization experiments and cell-based assays. B.M.G. performed PBMC staining and fluorescence-activated cell sorting (FACS). C.G.L. performed the large-scale RNA isolation and labeling. C.P.W. proposed the DTWD pathway, and P.C. and C.P.W. developed the KO cells. J. Perr developed the statistical testing of rPAL profiles. C.R.B. and N.A.T. designed the chemical synthesis routes. N.A.T. performed the chemical synthesis and purification and interpreted NMR and MS data. Y.X. prepared samples for SWATH-MS data. Y.X., B.A.G., R.A.F., N.A.T., and C.R.B. analyzed SWATH-MS data. R.A.F. and C.R.B. wrote the manuscript. All authors discussed the results and revised the manuscript.

DECLARATION OF INTERESTS

R.A.F. is a stockholder of ORNA Therapeutics and is a board of directors' member and stockholder of Chronus Health and Blue Planet Systems. C.R.B. is a cofounder of Firefly Biologics and a cofounder and scientific advisory board member of Lycia Therapeutics, Palleon Pharmaceuticals, Enable Bioscience, Redwood Biosciences (a subsidiary of Catalent), ReNAgade Therapeutics, and InterVenn Biosciences.

Received: February 25, 2024
Revised: June 17, 2024
Accepted: July 24, 2024
Published: August 21, 2024

REFERENCES

1. Pedowitz, N.J., and Pratt, M.R. (2021). Design and synthesis of metabolic chemical reporters for the visualization and identification of glycoproteins. RSC Chem. Biol. *2*, 306–321. https://doi.org/10.1039/D1CB00010A.
2. Flynn, R.A., Pedram, K., Malaker, S.A., Batista, P.J., Smith, B.A.H., Johnson, A.G., George, B.M., Majzoub, K., Villalta, P.W., Carette, J.E., et al. (2021). Small RNAs are modified with N-glycans and displayed on the surface of living cells. Cell *184*, 3109–3124.e22. https://doi.org/10.1016/j.cell.2021.04.023.
3. Zhang, N., Tang, W., Torres, L., Wang, X., Ajaj, Y., Zhu, L., Luan, Y., Zhou, H., Wang, Y., Zhang, D., et al. (2024). Cell surface RNAs control neutrophil recruitment. Cell *187*, 846–860.e17. https://doi.org/10.1016/j.cell.2023.12.033.
4. Ma, Y., Guo, W., Mou, Q., Shao, X., Lyu, M., Garcia, V., Kong, L., Lewis, W., Ward, C., Yang, Z., et al. (2024). Spatial imaging of glycoRNA in single cells with ARPLA. Nat. Biotechnol. *42*, 608–616. https://doi.org/10.1038/s41587-023-01801-z.
5. Li, J., Yue, S., Gao, Z., Hu, W., Liu, Z., Xu, G., Wu, Z., Zhang, X., Zhang, G., Qian, F., et al. (2023). Novel approach to enriching glycosylated RNAs: specific capture of GlycoRNAs via solid-phase chemistry. Anal. Chem. *95*, 11969–11977. https://doi.org/10.1021/acs.analchem.3c01630.
6. Batt, A.R., Zaro, B.W., Navarro, M.X., and Pratt, M.R. (2017). Metabolic chemical reporters of glycans exhibit cell-type-selective metabolism and glycoprotein labeling. Chembiochem *18*, 1177–1182. https://doi.org/10.1002/cbic.201700020.
7. Park, D.D., Xu, G., Wong, M., Phoomak, C., Liu, M., Haigh, N.E., Wongkham, S., Yang, P., Maverakis, E., and Lebrilla, C.B. (2018). Membrane glycomics reveal heterogeneity and quantitative distribution of cell surface sialylation. Chem. Sci. *9*, 6271–6285. https://doi.org/10.1039/C8SC01875H.
8. Liu, F., Chen, H.-M., Armstrong, Z., and Withers, S.G. (2022). Azido groups hamper glycan acceptance by carbohydrate processing enzymes. ACS Cent. Sci. *8*, 656–662. https://doi.org/10.1021/acscentsci.1c01172.
9. Gahmberg, C.G., and Andersson, L.C. (1977). Selective radioactive labeling of cell surface sialoglycoproteins by periodate-tritiated borohydride. J. Biol. Chem. *252*, 5888–5894. https://doi.org/10.1016/S0021-9258(17)40107-4.
10. Liu, T., Qian, W.-J., Gritsenko, M.A., Camp, D.G., Monroe, M.E., Moore, R.J., and Smith, R.D. (2005). Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. J. Proteome Res. *4*, 2070–2080. https://doi.org/10.1021/pr0502065.
11. Reuter, G., Schauer, R., Szeiki, C., Kamerling, J.P., and Vliegenthart, J.F. (1989). A detailed study of the periodate oxidation of sialic acids in glycoproteins. Glycoconj. J. *6*, 35–44. https://doi.org/10.1007/BF01047888.
12. Zeng, Y., Ramya, T.N.C., Dirksen, A., Dawson, P.E., and Paulson, J.C. (2009). High-efficiency labeling of sialylated glycoproteins on living cells. Nat. Methods *6*, 207–209. https://doi.org/10.1038/nmeth.1305.
13. Zhang, H., Li, X.-J., Martin, D.B., and Aebersold, R. (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat. Biotechnol. *21*, 660–666. https://doi.org/10.1038/nbt827.
14. Qiu, C., Liu, W.-Y., and Xu, Y.-Z. (2015). Fluorescence labeling of short RNA by oxidation at the 3′-end. Methods Mol. Biol. *1297*, 113–120. https://doi.org/10.1007/978-1-4939-2562-9_8.
15. Rillahan, C.D., Antonopoulos, A., Lefort, C.T., Sonon, R., Azadi, P., Ley, K., Dell, A., Haslam, S.M., and Paulson, J.C. (2012). Global metabolic inhibitors of sialyl- and fucosyltransferases remodel the glycome. Nat. Chem. Biol. *8*, 661–668. https://doi.org/10.1038/nchembio.999.
16. Lopez-Sambrooks, C., Shrimal, S., Khodier, C., Flaherty, D.P., Rinis, N., Charest, J.C., Gao, N., Zhao, P., Wells, L., Lewis, T.A., et al. (2016). Oligosaccharyltransferase inhibition induces senescence in RTK-driven tumor

cells. Nat. Chem. Biol. *12*, 1023–1030. https://doi.org/10.1038/nchembio.2194.

17. Elbein, A.D., Tropea, J.E., Mitchell, M., and Kaushal, G.P. (1990). Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. J. Biol. Chem. *265*, 15599–15605. https://doi.org/10.1016/S0021-9258(18)55439-9.
18. Zhang, W., Foo, M., Eren, A.M., and Pan, T. (2022). tRNA modification dynamics from individual organisms to metaepitranscriptomics of microbiomes. Mol. Cell *82*, 891–906. https://doi.org/10.1016/j.molcel.2021.12.007.
19. Boccaletto, P., Stefaniak, F., Ray, A., Cappannini, A., Mukherjee, S., Purta, E., Kurkowska, M., Shirvanizadeh, N., Destefanis, E., Groza, P., et al. (2022). MODOMICS: a database of RNA modification pathways. 2021 update. Nucleic Acids Res. *50*, D231–D235. https://doi.org/10.1093/nar/gkab1083.
20. Xie, Y., De Luna Vitorino, F.N., Chen, Y., Lempiäinen, J.K., Zhao, C., Steinbock, R.T., Lin, Z., Liu, X., Zahn, E., Garcia, A.L., et al. (2023). SWAMNA: a comprehensive platform for analysis of nucleic acid modifications. Chem. Commun. (Camb) *59*, 12499–12502. https://doi.org/10.1039/D3CC04402E.
21. Maley, F., Trimble, R.B., Tarentino, A.L., and Plummer, T.H. (1989). Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. Anal. Biochem. *180*, 195–204. https://doi.org/10.1016/0003-2697(89)90115-2.
22. Chang, Y.-C., Herath, J., Wang, T.H.-H., and Chow, C.S. (2008). Synthesis and solution conformation studies of 3-substituted uridine and pseudouridine derivatives. Bioorg. Med. Chem. *16*, 2676–2686. https://doi.org/10.1016/j.bmc.2007.11.039.
23. Nainytė, M., Amatov, T., and Carell, T. (2019). Synthesis of an acp3U phosphoramidite and incorporation of the hypermodified base into RNA. Chem. Commun. (Camb) *55*, 12216–12218. https://doi.org/10.1039/c9cc06314e.
24. Friedman, S., Li, H.J., Nakanishi, K., and Van Lear, G. (1974). 3-(3-amino-3-carboxy-n-propyl)uridine. The structure of the nucleoside in Escherichia coli transfer ribonucleic acid that reacts with phenoxyacetoxysuccinimide. Biochemistry *13*, 2932–2937. https://doi.org/10.1021/bi00711a024.
25. Jühling, F., Mörl, M., Hartmann, R.K., Sprinzl, M., Stadler, P.F., and Pütz, J. (2009). tRNAdb 2009: compilation of tRNA sequences and tRNA genes. Nucleic Acids Res. *37*, D159–D162. https://doi.org/10.1093/nar/gkn772.
26. Ohashi, Z., Maeda, M., McCloskey, J.A., and Nishimura, S. (1974). 3-(3-amino-3-carboxypropyl)uridine: a novel modified nucleoside isolated from Escherichia coli phenylalanine transfer ribonucleic acid. Biochemistry *13*, 2620–2625. https://doi.org/10.1021/bi00709a023.
27. Stuart, J.W., Basti, M.M., Smith, W.S., Forrest, B., Guenther, R., Sierzputowska-Gracz, H., Nawrot, B., Malkiewicz, A., and Agris, P.F. (1996). Structure of the trinucleotide D-acp3U-A with coordinated Mg^{2+} demonstrates that modified nucleosides contribute to regional conformations of RNA. Nucleos. Nucleot. *15*, 1009–1028. https://doi.org/10.1080/07328319608002031.
28. Takakura, M., Ishiguro, K., Akichika, S., Miyauchi, K., and Suzuki, T. (2019). Biogenesis and functions of aminocarboxypropyluridine in tRNA. Nat. Commun. *10*, 5542. https://doi.org/10.1038/s41467-019-13525-3.
29. Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N., and Isobe, T. (2003). Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. Nat. Biotechnol. *21*, 667–672. https://doi.org/10.1038/nbt829.
30. Ren, Z., Li, R., Zhou, X., Chen, Y., Fang, Y., and Zou, P. (2023). Enzyme-mediated proximity labeling identifies small RNAs in the endoplasmic reticulum lumen. Biochemistry *62*, 1844–1848. https://doi.org/10.1021/acs.biochem.3c00142.
31. Chen, C., Huang, B., Anderson, J.T., and Byström, A.S. (2011). Unexpected accumulation of ncm(5)U and ncm(5)S(2) (U) in a trm9 mutant suggests an additional step in the synthesis of mcm(5)U and mcm(5)S(2)U. PLoS One *6*, e20783. https://doi.org/10.1371/journal.pone.0020783.
32. Flynn, R.A., Smith, B.A.H., Johnson, A.G., Pedram, K., George, B.M., Malaker, S.A., Majzoub, K., Carette, J.E., and Bertozzi, C.R. (2019). Mammalian Y RNAs are modified at discrete guanosine residues with N-glycans. Preprint at bioRxiv. https://doi.org/10.1101/787614.
33. Xie, Y., Liu, X., Zhao, C., Chen, S., Wang, S., Lin, Z., Robison, F.M., George, B.M., Flynn, R.A., Lebrilla, C.B., et al. (2024). Development and application of GlycanDIA workflow for glycomic analysis. Preprint at bioRxiv. https://doi.org/10.1101/2024.03.12.584702.
34. Meyer, B., Wurm, J.P., Sharma, S., Immer, C., Pogoryelov, D., Kötter, P., Lafontaine, D.L.J., Wöhnert, J., and Entian, K.-D. (2016). Ribosome biogenesis factor Tsr3 is the aminocarboxypropyl transferase responsible for 18S rRNA hypermodification in yeast and humans. Nucleic Acids Res. *44*, 4304–4316. https://doi.org/10.1093/nar/gkw244.
35. Taoka, M., Nobe, Y., Yamaki, Y., Yamauchi, Y., Ishikawa, H., Takahashi, N., Nakayama, H., and Isobe, T. (2016). The complete chemical structure of Saccharomyces cerevisiae rRNA: partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. Nucleic Acids Res. *44*, 8951–8961. https://doi.org/10.1093/nar/gkw564.
36. Gray, M., Stanczak, M.A., Xiao, H., Pijnenborg, J.F.A., Malaker, S.A., Weidenbacher, P.A., Tanzo, J.T., Ahn, G., Woods, E.C., Läubli, H., et al. (2019). Targeted desialylation overcomes glyco-immune checkpoints and potentiates the anticancer immune response in vivo. Preprint at chemRxiv. https://doi.org/10.26434/chemrxiv.8187146.v1.
37. Adams, K.J., Pratt, B., Bose, N., Dubois, L.G., St. John-Williams, L., Perrott, K.M., Ky, K., Kapahi, P., Sharma, V., MacCoss, M.J., et al. (2020). Skyline for small molecules: A unifying software package for quantitative metabolomics. J. Proteome Res. *19*, 1447–1458. https://doi.org/10.1021/acs.jproteome.9b00640.
38. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308. https://doi.org/10.1038/nprot.2013.143.
39. Dong, L., Shen, S., Chen, W., Lu, H., Xu, D., Jin, S., Yang, Q., and Zhang, J. (2019). Glycosyl triazoles as novel insect β-N-acetylhexosaminidase OfHex1 inhibitors: design, synthesis, molecular docking and MD simulations. Bioorg. Med. Chem. *27*, 2315–2322. https://doi.org/10.1016/j.bmc.2018.11.032.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
APC/Cyanine7 anti-human CD14 Antibody	BioLegend	Cat# 325619; RRID:AB_830692
FITC anti-human CD3 Antibody	BioLegend	Cat# 317305; RRID:AB_571906
IRDye 800CW Goat anti-Mouse IgG	LI-COR Biosciences	Cat# 926-32210; RRID:AB_621842
IRDye 800CW Goat anti-Rabbit IgG	LI-COR Biosciences	Cat# 926-32211; RRID:AB_621843
Mouse GAPDH	Santa Cruz Biotechnology	Cat# sc-47724; RRID: AB_627678
Mouse IgG1,κ Isotype	BioLegend	Cat# 400130; RRID: AB_2800436
PE/Cyanine7 anti-human CD19 Antibody	BioLegend	Cat# 302215; RRID:AB_314245
PerCP/Cyanine5.5 anti-human CD16 Antibody	BioLegend	Cat# 302027; RRID:AB_893263
Rabbit DTWD2	Proteintech	Cat# 24873-1-AP; RRID: AB_2879770
Chemicals, peptides, and recombinant proteins		
0.5M EDTA pH 8.0	Invitrogen	Cat# AM9260G
10% SDS	Sigma-Aldrich	Cat# 15553027
4′,6-diamidino-2-phenylindole (DAPI)	Sigma-Aldrich	Cat# D1306
AccuGENE 10X MOPS Buffer	Lonza	Cat# 50876
Acetic Anhydride	Sigma Aldrich	Cat# 242845
Acetone	Fisher Scientific	Cat# A18-20
Acetonitrile, Optima LC/MS Grade	Fisher Scientific	Cat# A955-4
Agarose Molecular Biology Grade	IBI Scientific	Cat# IB70042
Agilent prep column (C18, 5 μm)	Agilent	Cat# 446905-702
Aldehyde Reactive Probe (trifluoroacetate salt)	Cayman Chemical	Cat# 10009350
Ammonium Acetate (5 M), RNase-free	Invitrogen	Cat# AM9070G
Anhydrous Tetrahydrofuran (THF)	Sigma Aldrich	Cat# 401757
Biotage® Sfär pre-packed silica gel columns (60 μm particle size)	Biotage	Cat# FSRS-0445-0025
Boc-homoserine-CO2Me ((S)-Methyl 2-((tert-butoxycarbonyl)amino)-4-hydroxybutanoate)	Chem Scene	Cat# CS-0088997
Bovine Serum Albumin	Sigma-Aldrich	Cat# A9418
CD_3Cl	Cambridge Isotope Laboratories	Cat# DLM-7-PK
CD_3OD	Cambridge Isotope Laboratories	Cat# DLM-15-PK
D_2O	Cambridge Isotope Laboratories	Cat# DLM-4DR-PK
DCM (dichloromethane)	Sigma Aldrich	Cat# D65100
diisopropyl azidodicarboxylate (DIAD)	Sigma Aldrich	Cat# 225541
Dimethyl Sulfoxide for molecular biology	Sigma-Aldrich	Cat# D8418-50ML
DMAP (N,N-dimethylaminopyridine)	Sigma Aldrich	Cat# 107700
DMSO-d6	Cambridge Isotope Laboratories	Cat# DLM-10-PK
Ethanol	Decon Labs	Cat# V1001
Formaldehyde Solution	Sigma-Aldrich	Cat# F8775-500ML
Formamide	Thermo Fischer Scientific	Cat# 15515026
Formic Acid, Optima LC/MS Grade	Fisher Scientific	Cat# A117-50
Glyco Buffer #1 (10x)	New England Biolabs	Cat# P0722L
Guanidium Chloride	Thermo Fischer Scientific	Cat# G3272-1KG

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HATU	Sigma Aldrich	Cat# 445460
HEPES Buffer	Sigma-Aldrich	Cat# 83264-100ML-F
Hexane (HPLC grade)	Sigma Aldrich	Cat# 293253
Human TruStain FcX	BioLegend	Cat# 422302
Hydrochloric Acid	Sigma-Aldrich	Cat# 320331-500ML
Hydrogen Chloride, 4M in 1,4-dioxane	Fisher Scientific	Cat #AAA16935JT
Hypercarb Porous Graphitic Carbon HPLC Column (1 mm x 100 mm, 3 μm)	Thermo Scientific	35003-101030
Hypercarb SPE 96-well plate	Thermo Scientific	60302-609
Intercept Protein-Free Blocking Buffer	Li-Cor Biosciences	Cat# 927-80001
Isopropanol	Thermo Fischer Scientific	Cat# BP26181
KCl (2 M), RNase-free	Invitrogen	Cat# AM9640G
Kifunensine	Sigma-Aldrich	Cat# K1140
Lithium Hydroxide	Fisher Scientific	Cat# 11386728
m-PEG3-aldehyde	BroadPharm	Cat# BP-23750
Magnesium Chloride Hexahydrate	Sigma-Aldrich	Cat# M2670-100G
Magnesium Sulfate	Sigma-Aldrich	Cat# M7506-500G
Methanol	Fisher Scientific	Cat# A452-4
$MgCl_2$ (1 M)	Invitrogen	Cat# AM9530G
N-azidoacetylmannosamine-tetraacylated (Ac4ManNAz)	Click Chemistry Tools	Cat# CCT-1084
N,N-diisopropylethylamine	Sigma Aldrich	Cat# D125806
NGI-1	Sigma-Aldrich	Cat# SML1620
Nitrocellulose Membranes, 0.45 uM	Bio-Rad Laboratories	Cat# 1620115
Nuclease P1	Sigma-Aldrich	Cat# N8630-1VL
P-3FAX-Neu5Ac	Tocris	Cat# 5760
PBS (1X), pH 7.4, without Ca & Mg	Quality Biological	Cat# 114-058-101
Peracetylated azido-GlcNAc (2-Acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-glucopyranosyl azide)	Synthose	Cat# AG931
Phosphate Buffered Saline (10X)	Genesee Scientific	Cat# 25-507X
Phosphodiesterase I from Crotalus adamanteus venom	Sigma-Aldrich	Cat# P3243
Phosphodiesterase II	Worthington Biochemical	Cat# LS003603
Pierce NeutrAvidin Agarose	Thermo Fischer Scientific	Cat# 29201
PNGase F PRIME-LY	Bulldog-Bio	Cat# NZPP050LY
PNGase F, Glycerol-free	New England Biolabs	Cat# P0705L
Proteinase K, Molecular Biology Grade	New England Biolabs	Cat# P8107S
Puromycin	Invivogen	Cat# ant-pr-1
RNA Binding Buffer	Zymo Research	Cat# R1013-2-100
RNase A	Sigma-Aldrich	Cat# 10109169001
RNase Cocktail	Thermo Fischer Scientific	Cat# AM2288
RNAzol RT	Molecular Research Center, Inc.	Cat# RN 190-500ML
Shrimp Alkaline Phosphatase	New England Biolabs	Cat# M0371L
Silicycle 0.25 mm silica gel F-254 plates	Silicycle	Cat# TLG-R10014B-323
Sodium (meta)periodate	Sigma-Aldrich	Cat# S1878-25G
Sodium Acetate	Sigma-Aldrich	Cat# 71183
Sodium Chloride	Sigma-Aldrich	Cat# S9888
Sodium Sulfite	Sigma-Aldrich	Cat# S0505-250G

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
StcE Enzyme	Sigma-Aldrich	Cat# SAE0202
Streptavidin-IR800	Li-Cor Biosciences	Cat# 926-32230
SybrGold (10,000x)	Thermo Fischer Scientific	Cat# S11494
TRI Reagent	Molecular Research Center	Cat # TR118
Triethylamine	Sigma Aldrich	Cat# 471283
Trifluoroacetic Acid, Optima LC/MS Grade	Fisher Scientific	Cat# A116-10X1AMP
Triphenylphosphine	Sigma Aldrich	Cat# T84409
Tween-20	Sigma-Aldrich	Cat# P1379
UltraPure DNase/RNase-Free Distilled Water	Invitrogen	Cat# 10977015
UltraPure™ 1M Tris-HCl, pH 8.0	Invitrogen	Cat# 15568025
Uridine	Sigma Aldrich	Cat# U3750
Water, Optima LC/MS Grade	Fisher Chemical	Cat# W64
Whatman Paper	GE Healthcare	Cat# 10426890
Zinc Chloride	TCI Chemicals	Cat# Z0014
α2-3,6,8 Neuraminidase	New England Biolabs	Cat# P0720L
α2-3,6,8,9 Neuraminidase A	New England Biolabs	Cat# P0722L
Critical commercial assays		
Plasma Membrane Protein Extraction Kit	Abcam	Cat# ab65400
Zymo-Spin I Column	Zymo Research	Cat# C1003
Zymo-Spin II Column	Zymo Research	Cat# C1008
Zymo-Spin III Column	Zymo Research	Cat# C1005
Deposited data		
Mass Spectrometry Data	This Study	MassIVE ID: MSV000093213
Mass Spectrometry Data	This Study	MassIVE ID: MSV000094102
Experimental models: Cell lines		
HEK293	ATCC	Cat# ATCC-CRL-3216
HeLa	ATCC	Cat# ATCC-CCL-2
Human Peripheral Blood Mononuclear Cells (PBMCs)	StemCell Technologies	Cat# 70025.1
K562	ATCC	Cat# ATCC-CCL-243
U2OS	ATCC	Cat# HTB96
U2OS, DTWD2 KO (clone 1, clone 2)	This paper	This paper
Software and algorithms		
Adobe Illustrator CC	Adobe	https://www.adobe.com
Fiji	ImageJ	https://imagej.net/software/fiji/downloads
FlowJo	BD Biosciences	https://www.flowjo.com/
Geneious Prime	Geneious	https://www.geneious.com/
NuMoFinder	Dr. Benjamin Garcia	https://github.com/ChenfengZhao/NuMoFinder
Python	Python	https://www.python.org/
SCIEX OS Explorer	SCIEX	https://sciex.com/products/software/sciex-os-software
Skyline	Dr. Mike MacCoss	https://skyline.ms/project/home/software/Skyline/begin.view

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ryan Flynn (ryan.flynn@childrens.harvard.edu).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

- Raw Mass Spectrometry data have been deposited in the MassIVE database and are publicly available as of the date of publication. The accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Mammalian cell culture

All cells were grown at 37°C and 5% CO_2. HeLa (ATCC), HEK293 (ATCC), K562 (ATCC), and U2OS (ATCC) cells were cultured in DMEM media supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (P/S).

METHOD DETAILS

Chemical inhibitors and metabolic chemical reporters

Stocks of N-azidoacetylmannosamine-tetraacylated (Ac_4ManNAz, Click Chemistry Tools) were made to 500 mM in sterile dimethyl sulfoxide (DMSO). For cell treatments, Ac_4ManNAz was used at a final concentration of 100 μM. Working stocks of glycan-biosynthesis inhibitors were all made in DMSO at the following concentrations and stored at -80°C: 5 mM NGI-1 (Sigma), 10 mM Kifunensine (Kif, Sigma), 50 mM P-3F_{AX}-Neu5Ac (Tocris). All compounds were used on cells for 24 hours.

RNA extraction and enzymatic cleanups

TRIzol extractions were performed as previously described in detail.[2] For RNAzol RT (Molecular Research Center, Inc.) extractions, the manufacturer's protocol was followed with the following details. First, RNAzol RT was added to lyse and denature cells or tissues, and denaturing was further encouraged by placing the samples at 50°C and shaking for 5 min. To phase separate the RNA, 0.4X volumes of water was added, vortexed, let to stand for 5 minutes at 25°C and lastly spun at 12,000x g at 4°C for 15 min. The aqueous phase was transferred to clean tubes and 1.1X volumes of isopropanol was added. The RNA is then purified over a Zymo column (Zymo Research). We found that preconditioning Zymo columns with water before binding nucleic acids produces more consistent results. For all column cleanups, we followed the following protocol. First, 350 μL of pure water was added to each column and spun at 10,000x g for 30 seconds, and the flowthrough was discarded. Next, precipitated RNA from the RNAzol RT extraction (or binding buffer precipitated RNA, below) is added to the columns, spun at 10,000x g for 10-20 seconds, and the flowthrough is discarded. This step is repeated until all the precipitated RNA is passed over the column once. Next, the column is washed three times total: once using 400 μL RNA Prep Buffer (3M GuHCl in 80% EtOH), twice with 400 μL 80% ethanol. The first two spins are at 10,000x g for 20 seconds, the last for 30 sec. The RNA is then treated with Proteinase K (Ambion) on the column. Proteinase K is diluted 1:19 in water and added directly to the column matrix (Zymo-I = 20 μL, Zymo-II = 50 μL, Zymo-IIICG = 60 μL, all from Zymo Research), and then allowed to incubate on the column at 37°C for 45 min. The column top is sealed with either a cap or parafilm to avoid evaporation. After the digestion, the columns are brought to room temperature for 5 min; lowering the temperature is important before proceeding. Next, eluted RNA is spun out into fresh tubes and a second elution with water is performed (Zymo-I = 30 μL, Zymo-II = 50 μL, Zymo-IIICG = 60 μL). To the eluate, 1.5 μg of the mucinase StcE (Sigma-Aldrich) is added for every 50 μL of RNA, and placed at 37°C for 30 minutes to digest. The RNA is then cleaned up again using a Zymo column. Here, 2X RNA Binding Buffer (Zymo Research) was added and vortexed for 10 seconds, and then 2X (samples + buffer) of 100% ethanol was added and vortexed for 10 sec. An example would be 50 μL of RNA, 100 μL of RNA Binding Buffer, and 300 μL of 100% ethanol. This is then bound to the column, cleaned up as described above, and eluted twice with water (Zymo-I = 25 μL, Zymo-II = 50 μL, Zymo-IIICG = 60 μL). The final enzymatically digested RNA is quantified using a Nanodrop under the manufacturer's RNA settings.

Binding conditions to efficiently precipitate small RNAs as highlighted in Figure S1C were optimized by varying the amount of ethanol added post RNA Binding Buffer mixing with RNA. Isopropanol was also exchanged for the ethanol at this step to assess its ability to facilitate small RNA capture on the Zymo columns.

After total RNA extraction, the RNA can be further processed in order to fractionate small (17–200 nts) and large RNA (>200 nts) using Zymo columns, largely as recommended by the manufacturer. First, an adjusted RNA binding buffer is made by mixing equal

volumes of RNA binding buffer and 100% ethanol. Two volumes of the adjusted buffer are added to the total RNA and are vortexed thoroughly to mix. The sample is then bound to the column as described above, but the flow through - which contains the small RNA - is saved. One volume of 100% ethanol is added to the small RNA and vortexed to mix. The small RNA is then bound to new columns. Both the large and small RNA (bound to their respective columns) are then cleaned up using 2X 400 μL 80% ethanol, the second clean being centrifuged for 30 seconds. The RNA is then eluted using 2X 50 μL of water.

In vitro and On Cell enzymatic digestions

To digest RNA, the following was used: 2 μL of RNase cocktail (0.5U/mL RNaseA and 20U/mL RNase T1, Thermo Fisher Scientific) with 20 mM Tris-HCl (pH 8.0), 100 mM KCl and 0.1 mM $MgCl_2$. To digest sialic acid: 1.5 μL of a2-3,6,8 Neuraminidase (50U/mL, New England Biolabs, NEB) with 1x GlycoBuffer 1 (NEB). Reactions were performed on 3 μg total RNA from indicated cell sources for 60 minutes at 37°C. For live cell treatments, VC-Sia was expressed and purified as previously described[36] and added to cells at 150 nM final concentration in complete growth media for between 15 and 60 minutes at 37°C.

Isolation of crude cellular membranes

Crude membranes were isolated using the Plasma Membrane Protein Extraction Kit (ab65400, Abcam): cultured cells first had growth media removed and cells were then washed twice with ice-cold 1x Phosphate Buffered Saline (PBS). In the second PBS wash, cells were scraped off the plate and spun down at 400x g for 4 minutes at 4°C or suspension cells were directly pelleted. Cell pellets were resuspended in 1mL of Homogenization Buffer Mix per 10,000,000 cells. Cell suspension was Dounce Homogenized on ice for 40-70 strokes, care was taken to stop douncing when the processing resulted in approximately 60% free nuclei so as to not generate excess lysis of nuclei. Homogenate was then spun at 700x g for 10 minutes at 4°C. This pellet contained the nuclear fraction and supernatants were transferred to new tubes and spun again at 10,000x g for 30 minutes at 4°C. The pellets generated from this spin were crude membranes and the supernatant was soluble cytosol. RNA extraction was performed as above and labeling as described below.

Periodate oxidation and aldehyde ligation for glycoRNA labeling

Starting with a maximum of 3 ug of lyophilized, enzymatically treated RNA, the first step of the rPAL labeling protocol is to block any aldehyde reactive species. To make the blocking buffer, 1μL 16 mM mPEG3-Ald (BP-23750, BroadPharm), 15 μL 1 M $MgSO_4$ and 12 μL 1 M NH_4OAc pH5 (with HCl) are mixed together (final buffer composition: 570 μM mPEG3-Ald + 500 mM $MgSO_4$ + 450 mM NH_4OAc pH5). 28μL of the blocking buffer is added to the lyophilized RNA, mixed completely by vortexing, and then incubated for 45 minutes at 35°C to block. The samples are briefly allowed to cool to room temperature (2-3 min), then working quickly, 1 μL 30 mM aldehyde reactive probe (ARP/aminooxy biotin, 10009350, Cayman Chemicals, stock made in water) is added first, then 2μL of 7.5 mM $NaIO_4$ (periodate, stock made in water) is added. The periodate is allowed to perform oxidation for exactly 10 minutes at room temperature in the dark. The periodate is then quenched by adding 3 μL of 22 mM sodium sulfite (stock made in water). The quenching reaction is allowed to proceed for 5 minutes at 25°C. Both the sodium periodate and sodium sulfite stocks were made fresh within 20 minutes of use. Next, the reactions are moved back to the 35°C heat block, and the ligation reaction is allowed to occur for 90 min. The reaction is then cleaned up using a Zymo-I column. 19 μL of water is added in order to bring the reaction volume to 50 μL, and the Zymo protocol is followed as per the above details. If samples will be analyzed on an agarose gel for glycoRNA visualization, the RNA is then eluted from the column using 2X 6.2μL water (final volume approximately 12 μL).

glycoRNA blotting

In order to visualize the periodate labeled RNA, it is run on a denaturing agarose gel, transferred to a nitrocellulose (NC) membrane, and stained with streptavidin in a manner similar to[2] with some modifications. After elution from the column as described above, the RNA is combined with 12 μL of Gel Loading Buffer II (GLBII, 95% formamide, 18 mM EDTA, 0.025% SDS) with a final concentration of 2X SybrGold (ThermoFisher Scientific) and denatured at 55°C for 10 minutes. It is important to not use GLBII with dyes. Immediately after this incubation, the RNA is placed on ice for at least 2 minutes. The samples are then loaded into a 1% agarose, 0.75% formaldehyde, 1.5x MOPS Buffer (Lonza) denaturing gel. Precise and consistent pouring of these gels is critical to ensure a similar thickness of the gel for accurate transfer conditions; we aim for approximately 1 cm thick of solidified gel. RNA is electrophoresed in 1x MOPS at 115V for between 34 or 45 min, depending on the length of the gel. Subsequently, the RNA is visualized on a UV gel imager, and excess gel is cut away; leaving ~0.75 cm of gel around the outer edges of sample lanes will improve transfer accuracy. The RNA is transferred as previously described,[2] however various buffer conditions and times were screened to determine the optimal method (Figure S1). Finally, we determined that a 3M NaCl solution at pH 1, achieved with HCl, yields the most consistent and efficient transfer of material to the NC membrane. Transfer occurs for 90 minutes at 25°C. Post transfer, the membrane is rinsed in 1x PBS and dried on Whatman Paper (GE Healthcare). Dried membranes are rehydrated in Intercept Protein-Free Blocking Buffer, TBS (Li-Cor Biosciences), for 30 minutes at 25°C. After the blocking, the membranes are stained using Streptavidin-IR800 (Li-Cor Biosciences) diluted 1:5,000 in Intercept blocking buffer for 30 minutes at 25°C. Excess Streptavidin-IR800 was washed from the membranes using three washes with 0.1% Tween-20 (Sigma) in 1x PBS for 3 minutes each at 25°C. The membranes were then briefly rinsed with PBS to remove the Tween-20 before scanning. Membranes were scanned on a Li-Cor Odyssey CLx scanner (Li-Cor Biosciences).

Peripheral blood mononuclear cell staining and flow cytometry sorting

Human Peripheral Blood Mononuclear Cells (PBMCs) were purchased from StemCell Technology. Samples from three individual donors were thawed at 37°C for 3 minutes and then diluted into 5 mL of FACS buffer (0.5% BSA in 1x PBS). Freezing media was removed by spinning the cells for 4 minutes at 400g at 4°C. 10 million cells were resuspended in 1 mL of FACS buffer and 50 μL of Human TruStain FcX Fc Receptor Blocking Solution (BioLegend) was added and left to bind the cells on ice for 15 minutes. The blocked PMBCs were stained with anti-CD19 clone HIB19 conjugated to PE-Cy7 (BioLegend), anti-CD3 clone OKT3 conjugated to FITC (BioLegend), anti-CD14 clone HCD14 conjugated to APC-Cy7 (BioLegend), and anti-CD16 clone 3G8 conjugated to PerCP-Cy5.5 (BioLegend) for 30 minutes on ice. Cells were pelleted by spinning for 4 minutes at 400g at 4°C. The supernatant was discarded and the cells were resuspended in 1 mL of FACS buffer supplemented with 4',6-Diamidino-2-phenylindole (DAPI) at 1 μg/mL final concentration as a live/dead stain. Stained cells were sorted into four major subpopulations using a Sony MA900 sorter with a 100 μM sorting chip (Sony). Cells were collected into FACS buffer and spun for 4 minutes at 400g at 4°C. Pellets were lysed in RNAzol RT and rPAL labeling was performed as described above.

rPAL labeling, glycoRNA capture, and enzymatic release

RNA from HEK293 and K562 cells was isolated, enzymatically digested, and removed of large RNA as described above. For each replicate 250 μg of small RNA was used. The rPAL procedure was performed as described above, scaled up linearly, with a 1x reaction assuming 3 μg input RNA. We performed the rPAL with or without the aminooxy-biotin reagent, omitting this reagent served as a control sample. After the rPAL labeling and Zymo cleanups were accomplished we captured biotinylated RNAs using 250 μL Neutravidin bead slurry (ThermoFisher Scientific) per replicate. The capture took place at 4°C, over 2 hours on rotation in a final volume of 1000 μL in 5x PBS. Beads were then washed sequentially with 1000 μL of the following buffers for 30 seconds each, all pre-warmed to 45°C before use: twice with 4 M NaCl, 100 mM HEPES, then twice 2 M GuHCl, 100 mM HEPES, then twice with room temperature 1x PBS. The washed beads were then subjected to nuclease treatment by incubating in 300 μL of 250 μg RNase A (Sigma, 10109169001), 5 units of Phosphodiesterase (PDE1, Sigma, P3243-1VL), and 0.5 mM $MgCl_2$ in 0.5x PBS. Nuclease digestion occurred at 37°C for 4 hours with shaking to 1100 rpm for 10 seconds, every 60 seconds. The supernatant of this reaction was collected in a new tube and the beads washed with 1000 μL of LC-MS grade water (ThermoFisher Scientific) and the water supernatant of this wash was pooled with the nuclease released supernatant. The pooled supernatants were frozen and lyophilized for LC-MS analysis as described below. The nuclease digested beads were next subjected to one of two types of glycosidase treatments, either PNGaseF or Sialidase, both of which occurred at 37°C for 16 hours with shaking to 1100 rpm for 10 seconds, every 60 seconds. Sialidase digestion mix was a final volume of 250 μL in 1x NEB Glyco Buffer #1 with 5 μL of α2-3,6,8 Neuraminidase (NEB, P0720) and 5 μL of α2-3,6,8,9 Neuraminidase A (NEB, P0722). PNGaseF digestion mix was a final volume of 250 μL in 1x PBS with 7.5 μL PNGaseF prime (Bulldog Bio, NZPP050LY). After the overnight reaction, the supernatants were collected again in a fresh tube, beads washed with LC-MS water as above, supernatants pooled, and samples lyophilized for LC-MS analysis.

Porous graphitic carbon liquid chromatography mass spectrometry analysis

Nucleosides sample preparation for rPAL-based experiments

10 mU of shrimp alkaline phosphatase (SAP, New England Biolabs) was added to the samples generated above to remove the phosphates. The reaction was employed in 50 mM sodium acetate (pH 7.2) for 2 hours at 37 °C. The nucleoside samples were further purified using the Hypercarb SPE 96-well plate (Thermo Fisher Scientific). The plate was first activated and equilibrated with a solution of 80% (v/v) Acetonitrile (ACN, LC-MS Grade, Thermo Fisher Scientific) and 0.1% (v/v) TFA and 0.1% (v/v) Trifluoroacetic Acid (TFA, LC-MS Grade, Thermo Fisher Scientific) in LC-MS water, respectively. The dried samples were solubilized, loaded onto the cartridge, and washed with 0.1% (v/v) TFA in water. Nucleosides were eluted with a solution of 60% (v/v) ACN and 0.1% (v/v) TFA and dried using the TurboVap LV Evaporator (Biotage, Sweden).

Nucleosides sample preparation for DTWD2 knock out experiments

Small RNAs (1 μg) were digested into ribonucleosides with 5 mU/μL of nuclease P1, 5 mU/μL of recombinant shrimp alkaline phosphatase, 500 μU/μL of phosphodiesterase I, and 6.25 μU/μL of phosphodiesterase II in 20 μL of digestion buffer (1 mM $ZnCl_2$, 1mM $MgCl_2$, 30 mM NaOAc, pH 7.2) overnight at room temperature. The digested ribonucleosides were purified using the Hypercarb SPE 96-well plate (Thermo Fisher Scientific) and dried using the TurboVap LV Evaporator (Biotage).

LC-MS/MS analysis

The digested nucleosides were reconstituted in MS buffer A (LC-MS water with 0.1% Formic Acid (FA, LC-MS Grade, Fisher Scientific) (v/v)), and 2 μL was injected for the analysis using the SWAMNA platform.[20] Briefly, the samples were introduced using nanoAcquity UPLC System (Waters) equipped with Hypercarb Porous Graphitic Carbon HPLC Column (1 mm x 100 mm, 3 μm). Separation was carried out at a constant flow rate of 40 μL/min using MS buffer A and buffer B (ACN with 0.1% FA (v/v). The gradient was referred from the previous protocol.[21] Specifically, 0–2 min, 0% B; 2–20 min, 0-16% B; 20–40 min, 16%-72% B; 40–42 min, 72%-100% B; 42-52 min, 100% B; 52-54 min 100%-0% B; 54-65 min 0% B. The analytes were ionized using OptiFlow Turbo V ion source (SCIEX) at 5000 V and analyzed using ZenoTOF 7600 (SCIEX). Ion source gas1, gas2, and curtain gas were at 40, 60, and 35 psi, respectively. The ion transfer tube temperature was set at 200 °C. For zenoSWATH analysis, MS spectra were acquired over a mass range of m/z 200–1250 in positive ionization mode with 0.25 s accumulation time. Collision-induced fragmentation was

performed with nitrogen gas using dynamic collision energy. The product ions were monitored from the range m/z 100-1200 with 15 ms accumulation time. For the targeted MRMHR analysis, the monitored masses and charges were listed in Table S1.

In vitro PNGaseF treatment of acp^{3}U_GlcNAc

The acp^{3}U_GlcNAc standard was treated with 2 μL of PNGase F (500 units/μL, NEB) in 50 μL of 50mM HEPES buffer (pH 7.5, Sigma) at 37°C overnight. The product was desalted using the Hypercarb SPE 96-well plate (Thermo Fisher Scientific) and dried using the TurboVap LV Evaporator (Biotage). The acp^{3}U and acp^{3}U_GlcNAc were monitored using targeted MRMHR analysis.

Data analysis

The data was manually inspected using SCIEX OS Explorer (v3.0), NuMo Finder, and Skyline (v22.2).[37] All raw data has been deposited in the MassIVE database with accession number: MSV000093213 and MSV000094102.

Generation of knockout U2OS cell lines

DTWD2 KO cell lines were generated using a CRISPR/Cas9 approach.[38] The target oligonucleotide (5'-AGCGCACCCTCTGCATATCT-3') was synthesized and ligated into PX459 vectors. U2OS cells were then transfected with gRNA vectors. Two days later, puromycin (2 μg/mL) was added into the cell culture and the live cells were selected by flow cytometry (FACS Calibur 2, BD science) for isolation of single clones. The expanded individual clones were screened by genomic DNA sequencing.

rPAL profile analysis

To compare the differences in rPAL signal distribution either with PNGaseF or upon gene knockout, ImageJ software was used to draw lines vertically through signal in each lane to measure signal intensity at different migration distances. Then, this list of intensities was used to generate histograms of signal in which each unit of signal intensity is represented by one unit of RNA signal. Total quantities of RNA from these histograms was then summed across triplicates and normalized to each other to prevent differences in loading from affecting the significance of differences in the migration of the RNA on the blot. Then distributions from triplicates were concatenated. A nonparametric permutation test by bootstrapping was then employed to determine the statistical significance of differences in the distribution of RNA migration distances for each treatment condition or genetic background. In this analysis, the null hypothesis stated that there was a greater than five percent chance that differences in medians between the union of two distributions would be greater than the observed difference. The test was performed with subsamples of size 100 over 10,000 iterations.

Synthetic procedures

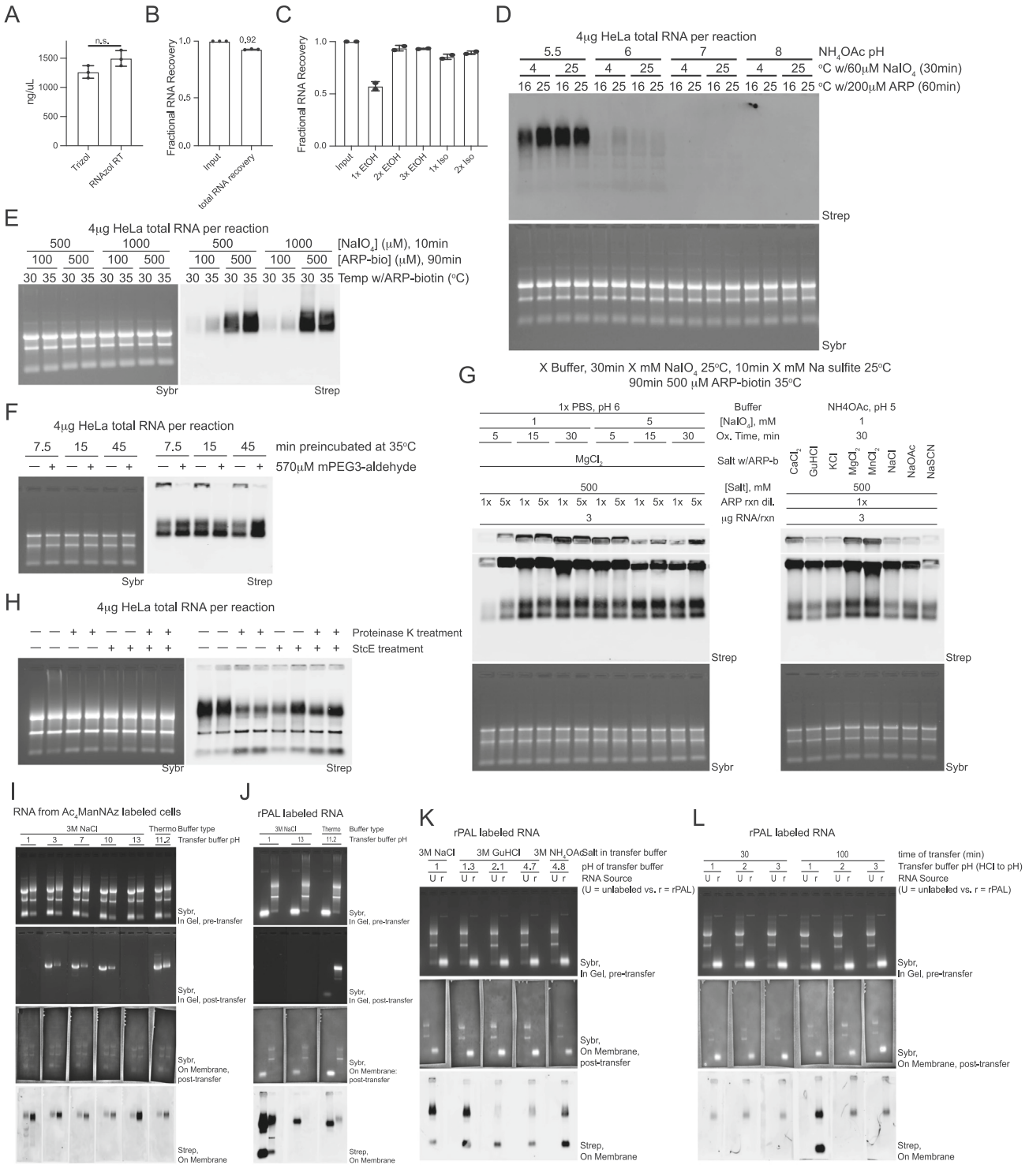
An overview of the synthetic approach to making acp^{3}U and acp^{3}U-GlcNAc and NMR Spectra of the intermediate and final compounds can be found in Methods S1. To begin the synthesis, peracetylated azido-GlcNAc (2-Acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-glucopyranosyl azide) and Boc-homoserine-CO_2Me ((*S*)-Methyl 2-((tert-butoxycarbonyl)amino)-4-hydroxybutanoate) were purchased from Synthose (catalog # AG931) and Chem Scene (catalog # CS-0088997) respectively and used without further purification. Acetylated amino-GlcNAc ((2*R*,3*S*,5*R*,6*R*)-5-acetamido-2-(acetoxymethyl)-6-aminotetrahydro-2H-pyran-3,4-diyl diacetate, S6) was prepared by palladium-catalyzed hydrogenation of the corresponding commercially available azide starting material according to a literature procedure.[39] Solvents were purified with a PureSolv Solvent Purification System or used as commercially available anhydrous sure-seal bottles. Non-aqueous reagents were transferred under nitrogen via syringe. Organic solutions were concentrated under reduced pressure on a Büchi rotary evaporator using a water bath. Normal phase chromatographic purification of products was accomplished using forced-flow chromatography on a Biotage® automated flash column chromatography system equipped with Biotage® Sfär pre-packed silica gel columns (60 μm particle size). Thin-layer chromatography (TLC) was performed on Silicycle 0.25 mm silica gel F-254 plates. Visualization of the developed chromatogram was performed by UV lamp exposure and $KMnO_4$ stain. ^{1}H NMR spectra were recorded on a Bruker NEO-500 MHz and are internally referenced to residual protio D_2O (4.79 ppm), CD_3OD (3.31 ppm), or $CDCl_3$ (7.26 ppm) signals. $CDCl_3$ was stored over K_2CO_3. Data for ^{1}H NMR are reported as follows: chemical shift (δ ppm), integration, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, dd = doublet of doublets, dt = doublet of triplets, br = broad), and coupling constant (Hz). ^{13}C NMR spectra were recorded on a Bruker UltraShield Plus 500 MHz and data are reported in terms of chemical shift relative to $CDCl_3$ (77.00 ppm) or CD_3OD (49.00 ppm). HRMS (High Resolution Mass Spectrometry) data were obtained on an Agilent 6230B Time of Flight (TOF) LC/MS by direct infusion. Reversed-phase preparatory HPLC purification was carried out with an Agilent 1260 Infinity II system equipped with an Agilent preparatory HPLC column (5 μm, C18, 50 x 21.2 mm L x ID).

QUANTIFICATION AND STATISTICAL ANALYSIS

For each experiment and data type, appropriate statistical tests were used and described in the figure legends as well as in the accompanying methods section above.

CellPress

Supplemental figures



(legend on next page)

Figure S1. Development of rPAL labeling strategy, related to Figure 1

(A) NanoDrop analysis of total RNA extracted from HeLa cells. Cells were lysed with either TRIzol or RNAzol RT and processed as per the manufacturer's recommendations with RNA, cleanups performed using Zymo columns (STAR Methods). Each datapoint (biological triplicate) is displayed with the SEM.

(B) Nanodrop analysis of total RNA extracted from HeLa cells isolated with RNAzol RT. After isolation, samples were quantified and then subjected to a Zymo column cleanup, after which each sample was quantified again. Each datapoint (biological triplicate) is displayed with the SEM.

(C) Nanodrop analysis of small RNA extracted from HeLa cells isolated with RNAzol RT. After isolation, samples were quantified and then subjected to a Zymo column cleanup. During the precipitation step, before binding to the Zymo columns, indicated ratios of ethanol or isopropanol were added. After Zymo column cleanup, each sample was quantified again. Each datapoint (biological duplicate) is displayed with the SEM.

(D) RNA blotting of total RNA from HeLa cells. Total RNA was subjected to various reaction conditions, including NH_4OAc buffer pH changes, the temperature at which diol oxidation occurred, and the temperature at which the aldehyde ligation occurred. In-gel detection of total RNA with SybrGold (Sybr, bottom) and on-membrane detection of biotin (Streptavidin-IR800, top) is shown.

(E) Blotting as in (D), with variations in the concentration of $NaIO_4$ used for oxidation and aldehyde reactive probe (ARP) biotin used for aldehyde ligation, as well as the temperature at which the aldehyde ligation occurred. Sybr and Strep detection is as in (D).

(F) Blotting as in (D), with evaluation of performing a blocking step before the $NaIO_4$ oxidation with mPEG3-aldehyde for the indicated times at 35°C. Sybr and Strep detection is as in (D).

(G) Blotting as in (D), with evaluation of buffer conditions, $NaIO_4$ concentration, ARP-biotin concentrations, and oxidation times for generating specific rPAL signal. Sybr and Strep detection is as in (D).

(H) Blotting as in (D), with evaluation of enzymatic digestions of the RNA samples after extraction using RNAzol RT. RNA was either not digested or subjected to digestion with Proteinase K (45 min at 37°C), StcE (35 min at 37°C), or both enzymes (sequentially). Sybr and Strep detection is as in (D).

(I) RNA northern blot transfer optimization. Small and total RNA samples from HeLa cells labeled with Ac_4ManNAz and visualized with copper-free click of DBCO-biotin were electrophoresed in a denaturing agarose gel and imaged with SybrGold (top row), after transferring with buffers of various pHs the gel was again imaged (second row), as well as the membrane to which the RNA was transferred (third row). Finally, sialoglycoRNAs were visualized with Streptavidin-IR800 (Strep, bottom row).

(J) RNA northern blot transfer optimization as in (I), assessing high and low pH conditions for sialoglycoRNAs transfer detecting with rPAL labeling.

(K) RNA northern blot transfer optimization as in (J), screening buffer salt composition as well as pH.

(L) RNA northern blot transfer optimization as in (J), assessing the time and pH dependency of transfer efficiency.

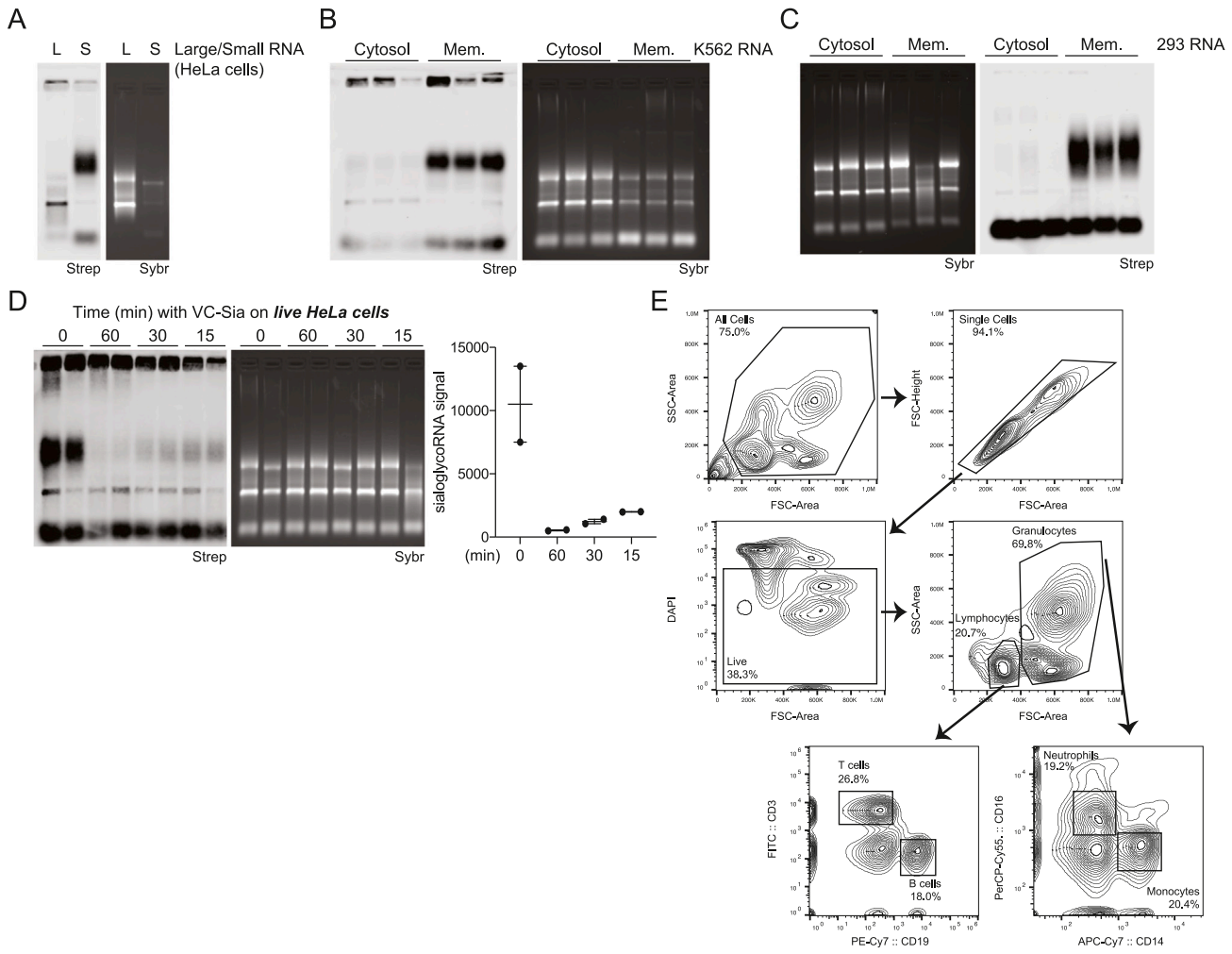


Figure S2. Subcellular localization of rPAL signal and flow cytometry analysis for sorting of human PBMCs, related to Figure 2

(A) RNA blotting of HeLa RNA that, after isolation, was subjected to a size fractionation (STAR Methods), separating out large (L, >~200 nts) from small (S, <~200 nts) RNAs. Detection of the resulting sialoglycoRNA signal was accomplished by rPAL labeling, in-gel detection of total RNA with SybrGold (Sybr, bottom), and on-membrane detection of biotin (Streptavidin-IR800, top) is shown.

(B) RNA blotting of RNA that was isolated from HEK293 cells fractionated for their crude membranes. The soluble cytosol is a fraction from this procedure that serves as a non-membrane-enriched control.

(C) RNA blotting of RNA that was isolated from K562 cells fractionated for their crude membranes. Sybr and Strep detection is as in (B).

(D) RNA blotting of total RNA from HeLa cells that, prior to RNAzol RT lysis, were exposed to VC sialidase treatment in the culture media for the indicated times. After treatment, cells were washed twice with 1× PBS, and then RNA was collected and labeled with rPAL. Sybr and Strep detection is as in (A). SialoglycoRNA signal was quantified and plotted (right). Each datapoint (biological duplicate) is displayed with the SEM.

(E) Scatterplots of one of three human peripheral blood mononuclear cell (PBMC)-sorting experiments. Each plot represents serial gates drawing to isolate live single cells and further fractionation of CD19 (B cells), CD3 (T cells), CD14 (monocytes and macrophages), and CD16 (NK cells, monocytes, and neutrophils). These four populations were then sorted into a FACS buffer for later RNA extraction and sialoglycoRNA analysis. Arrows denote the gating scheme used to isolate the four final populations of cells.

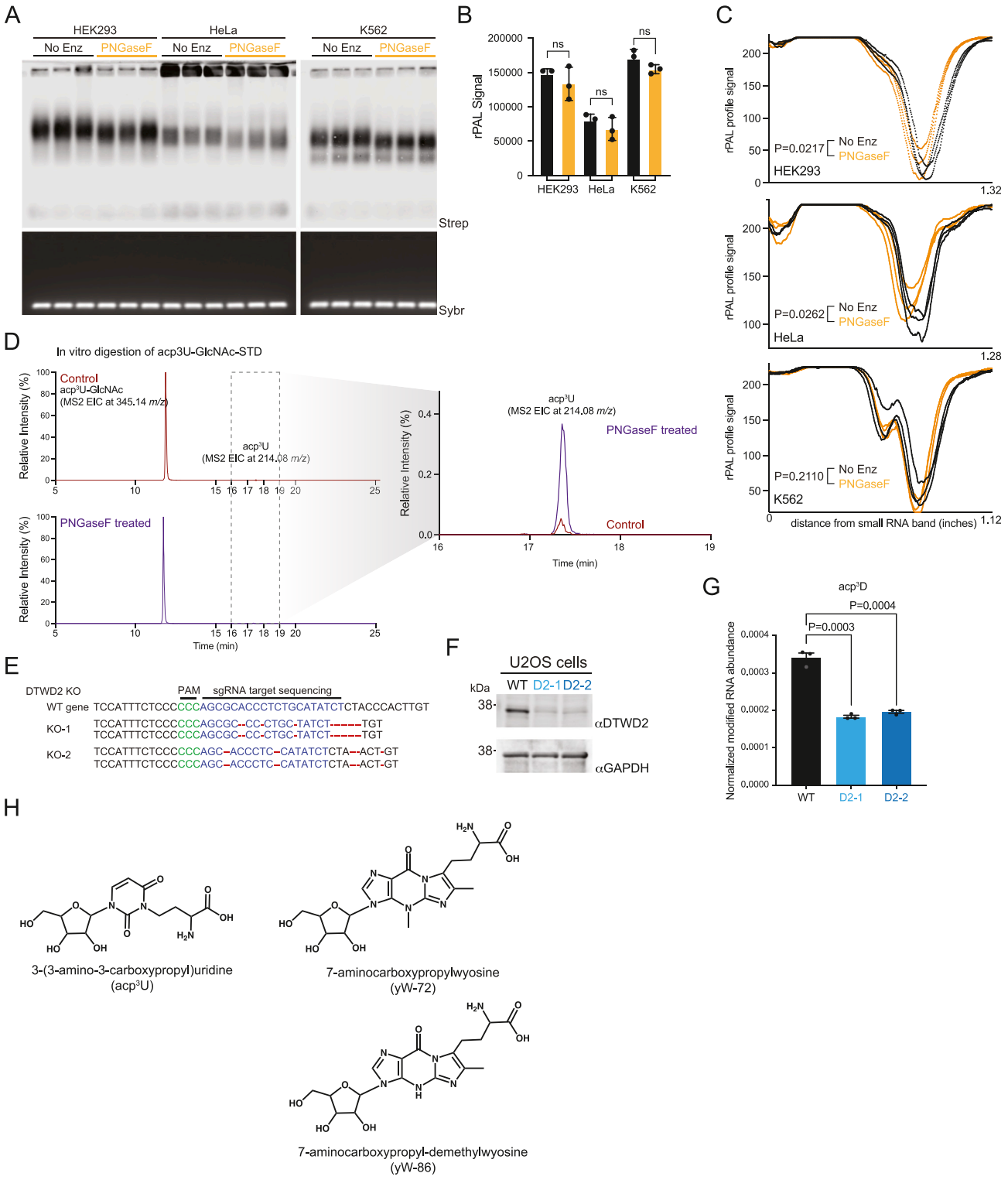


Figure S3. DTWD2 knockout and PNGaseF characterization, related to Figure 3

(A) Small RNA blotting of HEK293, HeLa, and K562 cells that, after isolation, were subjected PNGaseF digestion. Detection of the resulting sialoglycoRNA signal was accomplished by rPAL labeling, in-gel detection of total RNA (Sybr), and on-membrane detection of biotin (Strep).

(B) Quantification of the sialoglycoRNA signal in (A). Statistical assessment was accomplished by an unpaired t test, *p* value shown.

(legend continued on next page)

(C) Plot profile analysis of the lanes in (A), drawn upward from the small RNA band in the Strep image using ImageJ. Statistical assessment of the differences in the histogram profiles was assessed using a bootstrapping method (STAR Methods), *p* values shown.
(D) SWATH-MS analysis of *in vitro* PNGaseF digestion of acp^{3}U-GlcNAc-STD. Samples with (purple) and without (red) PNGaseF were reacted overnight and then analyzed for the levels of acp^{3}U-GlcNAc (left) and acp^{3}U (right breakout).
(E) Sequence alignment of DTWD2 coding sequences targeted for Cas9-mediated cleavage from WT and DTWD2 KO U2OS cell lines. Protospacer-adjacent motif (PAM), green; single guide RNA (sgRNA) target, blue; mutated sequence, red.
(F) Western blot analysis of WT, DTWD2-KO1, and DTWD2-KO2 total protein lysates. Membranes were stained with anti-DTWD2 or anti-GAPDH antibodies.
(G) Quantification of the normalized levels of acp^{3}D from small RNA isolated from WT, DTWD2-KO1, and DTWD2-KO2 U2OS cells. Statistical assessment was accomplished by an unpaired t test, *p* value shown.