



- (51) **International Patent Classification:**
C07K 17/00 (2006.01) *A61K 47/68* (2017.01)
- (21) **International Application Number:**
PCT/US2024/036428
- (22) **International Filing Date:**
01 July 2024 (01.07.2024)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
63/524,354 30 June 2023 (30.06.2023) US
- (71) **Applicant:** WAYNE STATE UNIVERSITY [US/US];
5057 Woodward Ave., Suite 6400, Detroit, Michigan 48202 (US).
- (72) **Inventors:** FEHL, Charlie; 5057 Woodward Avenue,
Suite 6400, Detroit, Michigan 48202 (US). LIU, Yimin;
5057 Woodward Avenue, Suite 6400, Detroit, Michigan
- 48202 (US). NELSON, Zachary M.; 5057 Woodward Avenue,
Suite 6400, Detroit, Michigan 48202 (US).
- (74) **Agent:** STAPLE, Julie K.; 900 Wilshire Drive, Suite 300,
Troy, Michigan 48084 (US).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.
- (84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST,

(54) **Title:** GLYCAN PROXIMITY LABELING METHODS AND APPLICATIONS THEREOF

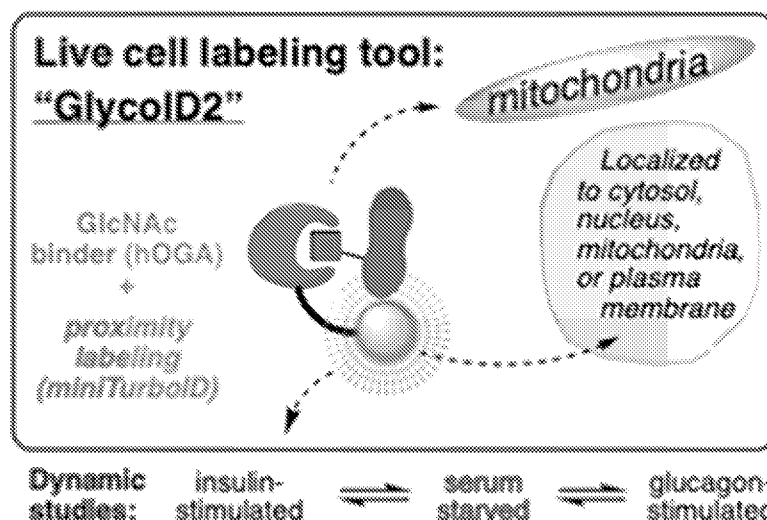


FIG. 1C

(57) **Abstract:** Fusion proteins are provided according to aspects of the present disclosure which include: a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein. Methods of detecting proteins proximal to a target protein are provided according to aspects of the present disclosure which include contacting a living cell with the fusion protein according to aspects of the present disclosure under compatible biological conditions; and detecting the biotinylated proteins, thereby detecting proteins proximal to the target protein.

SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*
- *with sequence listing part of description (Rule 5.2(a))*

GLYCAN PROXIMITY LABELING METHODS AND APPLICATIONS THEREOF

REFERENCE TO RELATED APPLICATION

[0001] This application claims priority from U.S. Provisional Patent Application Serial No. 63/524,354, filed June 30, 2023, the entire content of which is incorporated herein by reference.

GOVERNMENT FUNDING CLAUSE

[0002] This invention was made with government support under R35GM142637 awarded by the National Institute of General Medicine Sciences. The government has certain rights in the invention.

FIELD OF THE INVENTION

[0003] According to general aspects, the present disclosure relates to compositions and methods for labeling proteins. According to specific aspects, the present disclosure relates to fusion proteins including: a glycan binding component linked to a mutant *E. coli* biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant *E. coli* biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein; and methods for their use.

BACKGROUND OF THE INVENTION

[0004] A fundamental mechanism that all eukaryotic cells use to adapt to their environment is dynamic protein modification with monosaccharide sugars. In humans, O-linked N-acetylglucosamine (O-GlcNAc) is rapidly added to and removed from diverse protein sites as a response to fluctuating nutrient levels, stressors, and signaling cues.

[0005] The O-GlcNAc (O-linked *N*-acetylglucosamine) modification on proteins is a nutrient- and condition-sensing post-translational modification essential for all mammalian cells to adapt to their microenvironment. Thousands of O-GlcNAc sites regulate cell biology, including signaling and transcription, in both nutrient-driven and nutrient-independent roles. Protein O-GlcNAcylation is cycled by two proteins, O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA) (Figure 1A). The OGT gene can produce three isoforms, each of which is most active in a distinct cellular location: nucleocytoplasmic ncOGT is primarily found in the

nucleus; mitochondrial mOGT is found in mitochondria; and short sOGT, which lacks a nuclear localization signal and is therefore mainly cytosolic. During insulin signaling, OGT is known to move to the plasma membrane, where it is then active on membrane proteins. Therefore, a crucial facet of O-GlcNAc regulation depends on the spatial location of target proteins in the cell and which isoform(s) of OGT is produced at a given time (Figure 1B).

[0006] A second mechanism for O-GlcNAc regulation is time-based because O-GlcNAc modifications can be dynamically removed by OGA. In this vein, mammalian cells regulate the balance of OGT/OGA concerning overall O-GlcNAc levels, employing a variety of mechanisms including regulatory modifications, expression, as well as levels of OGT and OGA pre-mRNA transcripts. In particular, this mRNA regulation via alternative splicing enables cells to respond to O-GlcNAc perturbations within 30 min. During OGT/OGA rebalancing, O-GlcNAc events in this 30 min phase are increasingly recognized as critical for a wide range of cellular functions.

[0007] There is a continuing need for compositions and methods specific for O-GlcNAc sugar modifications which allow detection of changes in space and time.

SUMMARY OF THE INVENTION

[0008] Fusion proteins are provided according to aspects of the present disclosure which include: a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein. According to aspects of the present disclosure, the glycan binding component is linked to the mutant E. coli biotin ligase BirA by a linker disposed between the glycan binding component and the mutant E. coli biotin ligase BirA.

[0009] Fusion proteins are provided according to aspects of the present disclosure which include: a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein, wherein the glycan binding component is selected from the group consisting of: an enzyme, a lectin with the proviso that the lectin is not a GalD lectin, a collectin, a ficolin, a C-reactive protein, and a carbohydrate-binding domain of any thereof. According to aspects of the present disclosure, the glycan binding component is linked to the mutant E. coli biotin ligase BirA by a linker disposed between the glycan binding component and the mutant E. coli biotin ligase BirA.

[0010] Fusion proteins are provided according to aspects of the present disclosure which include: a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein, wherein the glycan binding component is selected from the group consisting of: an aptamer, an antibody, and an antigen-binding fragment of an antibody. According to aspects of the present disclosure, the glycan binding component is linked to the mutant E. coli biotin ligase BirA by a linker disposed between the glycan binding component and the mutant E. coli biotin ligase BirA.

[0011] Fusion proteins are provided according to aspects of the present disclosure which include: a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein, wherein the glycan binding component is a mutant O-GlcNAcase enzyme derived from a member of the GH84 family of glycosylhydrolases and lacking enzymatic glycosylhydrolase activity. According to aspects of the present disclosure, the glycan binding component is linked to the mutant E. coli biotin ligase BirA by a linker disposed between the glycan binding component and the mutant E. coli biotin ligase BirA.

[0012] Fusion proteins are provided according to aspects of the present disclosure which include: a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein, wherein the glycan binding component is a mutant O-GlcNAcase enzyme derived from a member of the GH84 family of glycosylhydrolases and lacking enzymatic glycosylhydrolase activity, wherein the mutant human O-GlcNAcase (hOGA) enzyme is a D174N mutant human O-GlcNAcase (hOGA) enzyme which includes the amino acid sequence:

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAPGEDNPAGAGGA AVAGAAGGAR
RFLCGVVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMYSVE
EAEQLMTLISAAREYEIEFIYAISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDI
DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNVSQSPYLRT
VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY

KGRSTELIPRLKGVLTNPNCFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
IQIKLENEGSDEDIETDVLVYSPQMALKLALTEWLQEFVPHQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATTVVTTVYQEPIMSQGAALSGEPTTLTKEEEKKQPDEEPMDMVVEK
QEETDHKNDNQILSEIVEAKMAEELKPMDDTKESIAESKSPMSMQEDCISDIAPMQTD
EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVVSVNCKGKDSEKIEEWSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
DIKSIMSMVKSFVQWLGCRRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ
PP (SEQ ID NO:4), or a variant thereof. According to aspects of the present disclosure, the glycan binding component is linked to the mutant E. coli biotin ligase BirA by a linker disposed between the glycan binding component and the mutant E. coli biotin ligase BirA.

[0013] Fusion proteins are provided according to aspects of the present disclosure which include: a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein, wherein the glycan binding component has a C-terminus and an N-terminus, the mutant E. coli biotin ligase BirA has a C-terminus and an N-terminus, and the C-terminus of the glycan binding component is linked to the N-terminus of the mutant E. coli biotin ligase BirA or the N-terminus of the glycan binding component is linked to the C-terminus of the mutant E. coli biotin ligase BirA. According to aspects of the present disclosure, the glycan binding component is linked to the mutant E. coli biotin ligase BirA by a linker disposed between the glycan binding component and the mutant E. coli biotin ligase BirA.

[0014] Fusion proteins are provided according to aspects of the present disclosure which include: a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein, wherein the glycan binding component is selected from the group consisting of: an enzyme, a lectin with the proviso that the lectin is not a GafD lectin, a collectin, a ficolin, a C-reactive protein, and a carbohydrate-binding domain of any thereof, wherein the glycan binding component has a C-terminus and an N-terminus, the mutant E. coli biotin ligase BirA has a C-terminus and an N-terminus, and the C-terminus of the glycan binding component is linked to the N-terminus of the mutant E. coli biotin ligase BirA or the N-terminus of the glycan binding component is linked to the C-terminus of the mutant E. coli biotin ligase BirA. According to aspects of the present disclosure, the glycan binding

component is linked to the mutant *E. coli* biotin ligase BirA by a linker disposed between the glycan binding component and the mutant *E. coli* biotin ligase BirA.

[0015] Fusion proteins are provided according to aspects of the present disclosure which include: a glycan binding component linked to a mutant *E. coli* biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant *E. coli* biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein, wherein the glycan binding component is selected from the group consisting of: an aptamer, an antibody, and an antigen-binding fragment of an antibody, wherein the glycan binding component has a C-terminus and an N-terminus, the mutant *E. coli* biotin ligase BirA has a C-terminus and an N-terminus, and the C-terminus of the glycan binding component is linked to the N-terminus of the mutant *E. coli* biotin ligase BirA or the N-terminus of the glycan binding component is linked to the C-terminus of the mutant *E. coli* biotin ligase BirA. According to aspects of the present disclosure, the glycan binding component is linked to the mutant *E. coli* biotin ligase BirA by a linker disposed between the glycan binding component and the mutant *E. coli* biotin ligase BirA.

[0016] Fusion proteins are provided according to aspects of the present disclosure which include: a glycan binding component linked to a mutant *E. coli* biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant *E. coli* biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein, wherein the glycan binding component is a mutant O-GlcNAcase enzyme derived from a member of the GH84 family of glycosylhydrolases and lacking enzymatic glycosylhydrolase activity, wherein the glycan binding component has a C-terminus and an N-terminus, the mutant *E. coli* biotin ligase BirA has a C-terminus and an N-terminus, and the C-terminus of the glycan binding component is linked to the N-terminus of the mutant *E. coli* biotin ligase BirA or the N-terminus of the glycan binding component is linked to the C-terminus of the mutant *E. coli* biotin ligase BirA. According to aspects of the present disclosure, the glycan binding component is linked to the mutant *E. coli* biotin ligase BirA by a linker disposed between the glycan binding component and the mutant *E. coli* biotin ligase BirA.

[0017] Fusion proteins are provided according to aspects of the present disclosure which include: a glycan binding component linked to a mutant *E. coli* biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant *E. coli* biotin ligase BirA having enzymatic activity to ligate

biotin to proteins proximal to the target protein, wherein the glycan binding component is a mutant O-GlcNAcase enzyme derived from a member of the GH84 family of glycosylhydrolases and lacking enzymatic glycosylhydrolase activity, wherein the mutant human O-GlcNAcase (hOGA) enzyme is a D174N mutant human O-GlcNAcase (hOGA) enzyme which includes the amino acid sequence:

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAPGEDNPAGAGGA AVAGAAGGAR
RFLCGVVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMYSVE
EAEQLMTLISAAREYEIEFIY AISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDI
DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNVSQSPYLRT
VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY
KGRSTELIPRLKGVL TNPNCEFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
IQIKLENEGSDEDIETDVL YSPQMALKLALTEWLQEF GVP HQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATT VVT TVYQEPIMSQGAALSGEPTTLTKEEEKKQPDEEPM DMVVEK
QEETDHKNDNQILSEIVEAKMAEELKPM DTDKESIAESKSP EMSMQEDCISDIAPMQTD
EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVVSVNCKGKDSEKIEE WRSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
DIKSIMSMVKS FVQWL GCRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ

PP (SEQ ID NO:4), or a variant thereof, wherein the glycan binding component has a C-terminus and an N-terminus, the mutant E. coli biotin ligase BirA has a C-terminus and an N-terminus, and the C-terminus of the glycan binding component is linked to the N-terminus of the mutant E. coli biotin ligase BirA or the N-terminus of the glycan binding component is linked to the C-terminus of the mutant E. coli biotin ligase BirA. According to aspects of the present disclosure, the glycan binding component is linked to the mutant E. coli biotin ligase BirA by a linker disposed between the glycan binding component and the mutant E. coli biotin ligase BirA.

[0018] According to aspects of the present disclosure, the fusion protein includes a localization signal peptide.

[0019] According to aspects of the present disclosure, the fusion protein includes a localization signal peptide capable of promoting localization of the fusion protein to a subcellular compartment selected from the group consisting of: nucleus, cytosol, mitochondria, endoplasmic reticulum, and plasma membrane.

[0020] According to aspects of the present disclosure, the fusion protein includes an exogenous detectable tag.

[0021] According to aspects of the present disclosure, the fusion protein includes: 1) SEQ ID NO: 3 or a variant thereof and SEQ ID NO: 2 or a variant thereof; 2) SEQ ID NO: 3 or a variant thereof and SEQ ID NO: 11 or a variant thereof; 3) SEQ ID NO: 34 or a variant thereof and SEQ ID NO: 2 or a variant thereof; or SEQ ID NO: 4 or a variant thereof and SEQ ID NO: 11 or a variant thereof.

[0022] A fusion protein provided according to aspects of the present disclosure includes: SEQ ID NO:15, or a variant thereof.

[0023] A fusion protein provided according to aspects of the present disclosure includes: SEQ ID NO:16, or a variant thereof.

[0024] A fusion protein provided according to aspects of the present disclosure includes: SEQ ID NO:28, or a variant thereof.

[0025] Fusion proteins are provided according to aspects of the present disclosure which include: SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, or a variant of any thereof.

[0026] Methods of detecting proteins proximal to a target protein are provided according to aspects of the present disclosure which include contacting a living cell with the fusion protein according to aspects of the present disclosure which include: a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein, under compatible biological conditions, whereby the fusion protein specifically binds to a glycosylation post-translational modification of a target protein of the cell; providing biotin to the living cell, whereby the mutant E. coli biotin ligase BirA ligates biotin to proteins proximal to the target protein; and detecting the biotinylated proteins, thereby detecting proteins proximal to the target protein.

[0027] According to aspects of methods of the present disclosure, detecting the biotinylated proteins includes purifying the biotinylated proteins and detecting the purified biotinylated proteins.

[0028] According to aspects of methods of the present disclosure, detecting the purified biotinylated proteins comprises mass spectrometry.

[0029] According to aspects of methods of the present disclosure, detecting the purified biotinylated proteins comprises chromatography. According to aspects of methods of the present disclosure, the chromatography comprises gel electrophoresis.

[0030] According to aspects of methods of the present disclosure, the chromatography comprises gel electrophoresis and transfer of the electrophoresed purified biotinylated proteins to a membrane.

[0031] According to aspects of methods of the present disclosure, contacting the living cell with the fusion protein includes introducing an expression construct encoding the fusion protein into the cell.

[0032] Expression constructs are provided according to aspects of the present disclosure which include a nucleic acid encoding a fusion protein of the present disclosure

[0033] Cells including an expression construct containing a nucleic acid encoding a fusion protein of the present disclosure are provided herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0034] Figure 1A is an illustration showing that O-GlcNAc modifications are cycled by two enzymes on 1000s of known substrates, occurring at varying timescales and driven by fluctuations in nutrients, cell stressors, or signaling cues.

[0035] Figure 1B is an illustration showing that OGT isoforms localize to and move between subcellular compartments, enabling spatiotemporal control of protein functions in cells.

[0036] Figure 1C is a diagrammatic representation of a chimeric protein according to aspects of the present disclosure for live-cell O-GlcNAc labeling and methods of use.

[0037] Figure 2 is a graph showing results of stable transfection of hOGA-miniTurbo-NES in U2OS cells, and incubation of the cells with biotin for 1 hour of labeling; biotin labeled proteins were enriched and the enriched proteins were identified with proteomics and then checked to determine whether the identified proteins were known to be O-GlcNAc proteins.

[0038] Figure 3 diagrammatically shows selected fusion protein constructs of the present disclosure including linkers.

[0039] Figure 4 shows a set of images of Western blots and a graph demonstrating that GlycoID2 vs. mCherry-miniTurbo labeling experiments in high and low glucose conditions did not make an O-GlcNAc-dependent difference in labeling.

[0040] Figure 5A is an image showing results of mapping technical replicates between the two biological replicates and demonstrating 80% and 67% overlap between the two samples, respectively, indicating moderate to good reproducibility for cyt-GlycoID2 labeling in stable expression cell lines.

[0041] Figure 5B is a graph showing O-GlcNAc overlap (cyt-GlycoID2 replicates).

[0042] Figure 5C is a graph showing primary subcellular locations of O-GlcNAc proteins from the cyt-GlycoID2 biological replicates.

[0043] Figure 6A is an image showing results of analysis of overlap between the O-GlcNAc proteins induced by insulin compared with the O-GlcNAc proteins induced by glucagon.

[0044] Figure 6B is an image showing results of analysis of overlap of O-GlcNAc proteins induced between 4 time points with insulin.

[0045] Figure 6C is an image showing results of analysis of overlap of O-GlcNAc proteins induced between 4 time points with glucagon.

[0046] Figure 6D is an image of a Western blot showing Akt-Serine473 phosphorylation as a marker for insulin induction of O-GlcNAc proteins.

[0047] Figure 6E is an image of a Western blot showing CREB phosphorylation as a marker of glucagon induction of O-GlcNAc proteins.

DETAILED DESCRIPTION

[0048] Scientific and technical terms used herein are intended to have the meanings commonly understood by those of ordinary skill in the art. Such terms are found defined and used in context in various standard references illustratively including J. Sambrook and D.W. Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press; 3rd Ed., 2001; F.M. Ausubel, Ed., *Short Protocols in Molecular Biology*, Current Protocols; 5th Ed., 2002; B. Alberts et al., *Molecular Biology of the Cell*, 4th Ed., Garland, 2002; CRISPR/Cas: A Laboratory Manual, Doudna and Mali (eds), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, 2016; D.L. Nelson and M.M. Cox, *Lehninger Principles of Biochemistry*, 4th Ed., W.H. Freeman & Company, 2004; J.-H. Fuhrhop et al. (Eds.), *Organic Synthesis, Concepts and Methods*, 3rd Ed., Wiley-VCH Verlag GmbH & Co. KGaA, 2003; Herdewijn, P. (Ed.), *Oligonucleotide Synthesis: Methods and Applications*, Methods in Molecular Biology, Humana Press, 2004; D. J. Taxman (ed.), *siRNA Design, Methods and Protocols*, Humana Press, 2012; Harlow, E. and Lane, D., *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1988; J. D. Pound (Ed.) *Immunochemical Protocols*, Methods in Molecular Biology, Humana Press, 2nd ed., 1998; Chu, E. and Devita, V.T., Eds., *Physicians' Cancer Chemotherapy Drug Manual*, Jones & Bartlett Publishers, 2021; J.M. Kirkwood et al., Eds., *Current Cancer Therapeutics*, 4th Ed., Current Medicine Group, 2001; A Adejare (Ed.), *Remington: The Science and Practice of Pharmacy*, Elsevier, 23rd Ed., 2021; L.V. Allen, Jr. et al., *Ansel's*

Pharmaceutical Dosage Forms and Drug Delivery Systems, 11th Ed., Wolters Kluwer, 2016; and L. Brunton et al., Goodman & Gilman's The Pharmacological Basis of Therapeutics, McGraw-Hill Education, 13th Ed., 2018.

[0049] The singular terms "a," "an," and "the" are not intended to be limiting and include plural referents unless explicitly stated otherwise or the context clearly indicates otherwise.

[0050] The terms "includes," "comprises," "including," "comprising," "has," "having," and grammatical variations thereof, when used in this specification, are not intended to be limiting, and specify the presence of stated features, elements, and/or components, but do not preclude the presence or addition of one or more other features, elements, components, and/or groups thereof.

[0051] The term "about" as used herein in reference to a number is used herein to include numbers which are greater, or less than, a stated or implied value by 1%, 5%, 10%, or 20%.

[0052] Particular combinations of features are recited in the claims and/or disclosed in the specification, and these combinations of features are not intended to limit the disclosure of various aspects. Combinations of such features not specifically recited in the claims and/or disclosed in the specification. Although each dependent claim listed below may directly depend on only one claim, the disclosure of various aspects includes each dependent claim in combination with every other claim in the claim set. As used herein, a phrase referring to "at least one of" a list of items refers to any combination of those items, including single members. As an example, "at least one of: a, b, or c" is intended to cover a alone; b alone; c alone, a and b, a, b, and c, b and c, a and c, as well as any combination with multiples of the same element, such as a and a; a, a, and a; a, a, and b; a, a, and c; a, b, and b; a, c, and c; and any other combination or ordering of a, b, and c).

[0053] The terms "first," "second," and the like are used herein to describe various features or elements, but these features or elements are not intended to be limited by these terms, but are only used to distinguish one feature or element from another feature or element. Thus, a first feature or element could be termed a second feature or element, and vice versa, without departing from the teachings of the present disclosure.

[0054] Fusion proteins are provided according to aspects of the present disclosure which include: a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein.

[0055] Figure 1C shows an overview of compositions and methods according to aspects of the present disclosure.

[0056] The term "glycan binding component" as used herein refers to a binding agent characterized by specific binding to a specified glycan target, but excludes GafD lectin. The phrase "specific binding" and grammatical equivalents as used herein in reference to binding of a binding agent to a specified glycan target refers to binding of the binding agent to the specified glycan target without substantial binding to other substances present in a cell which include a fusion protein according to aspects of the present disclosure. The term "binding" refers to a physical or chemical interaction between a binding agent and its target. Binding includes, but is not limited to, ionic bonding, non-ionic bonding, covalent bonding, hydrogen bonding, hydrophobic interaction, hydrophilic interaction, and Van der Waals interaction.

[0057] Specific binding refers to a binding agent that binds to a specified glycan target with greater affinity, greater avidity, and/or greater duration, than to other substances. According to aspects of the present disclosure, a binding agent specifically binds to its glycan target when it has an equilibrium dissociation constant, K_D , for its target in the range of about 10^{-4} to about 10^{-12} , i.e. a K_D of about 10^{-4} , about 10^{-5} , about 10^{-6} , about 10^{-7} , about 10^{-8} , about 10^{-9} , about 10^{-10} , about 10^{-11} , or about 10^{-12} . Binding affinity of a binding agent can be determined by Scatchard analysis such as described in P.J. Munson and D. Rodbard, *Anal. Biochem.*, 107:220-239, 1980 or by other methods such as Biomolecular Interaction Analysis using plasmon resonance.

[0058] Binding agents specific for a specified glycan target may be obtained from commercial sources or generated for use in methods of the present disclosure according to well-known methodologies.

[0059] According to aspects of the present disclosure, the glycan binding component is a binding agent which is, or includes, an enzyme, a lectin other than GafD lectin, a collectin, a ficolin, a C-reactive protein, or a carbohydrate-binding domain of any thereof. According to aspects of the present disclosure, the glycan binding component is a binding agent which is or includes, an aptamer, antibody, or an antigen-binding fragment of an antibody.

[0060] The term "antibody" is used herein in its broadest sense and includes antibodies, and antigen-binding fragments, characterized by specific binding to an antigen. An antibody included in methods according to aspects of the present disclosure may be a polyclonal antibody, a monoclonal antibody, a chimeric antibody, a humanized antibody, and/or an antigen-binding antibody fragment of any thereof. An antibody included in methods in particular aspects of the present disclosure includes a standard intact immunoglobulin having four polypeptide chains

including two heavy chains (H) and two light chains (L) linked by disulfide bonds. An antibody included in methods in particular aspects of the present disclosure includes an antigen-binding antibody fragments illustratively include an Fab fragment, an Fab' fragment, an F(ab')₂ fragment, an Fd fragment, an Fv fragment, an scFv fragment and a domain antibody (dAb), for example. In addition, the term antibody refers to antibodies of various classes including IgG, IgM, IgA, IgD and IgE, as well as subclasses, illustratively including for example human subclasses IgG1, IgG2, IgG3 and IgG4 and marine subclasses IgG1, IgG2, IgG2a, IgG2b, IgG3 and IgGM, for example.

[0061] In particular embodiments, an antibody which is characterized by specific binding to its target has a dissociation constant in the range of about 10^{-4} to about 10^{-12} , i.e. a K_D of about 10^{-4} , about 10^{-5} , about 10^{-6} , about 10^{-7} , about 10^{-8} , about 10^{-9} , about 10^{-10} , about 10^{-11} , or about 10^{-12} . Binding affinity of an antibody can be determined by Scatchard analysis such as described in P.J. Munson and D. Rodbard, *Anal. Biochem.*, 107:220-239, 1980 or by other methods such as Biomolecular Interaction Analysis using plasmon resonance. Antibodies may be tested for specific binding to the target by methods illustratively including ELISA, Western blot, and immunocytochemistry.

[0062] Antibodies, antigen-binding fragments, and methods for their generation are known in the art, for instance, as described in *Antibody Engineering*, Kontemann, R. and Dubel, S. (Eds.), Springer, 2001; Harlow, E. and Lane, D., *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1988; Ausubel, F. et al., (Eds.), *Short Protocols in Molecular Biology*, Wiley, 2002; J. D. Pound (Ed.) *Immunochemical Protocols*, Methods in Molecular Biology, Humana Press, 2nd ed., 1998; B. K. C. Lo (Ed.), *Antibody Engineering: Methods and Protocols*, Methods in Molecular Biology, Humana Press, 2003; and Kohler, G. and Milstein, C., *Nature*, 256:495-497 (1975).

[0063] A binding agent according to aspects of the present disclosure may be an aptamer. The term "aptamer" refers to a nucleic acid or peptide that substantially specifically binds to a specified substance. In the case of a nucleic acid aptamer, the aptamer is characterized by binding interaction with a target other than Watson/Crick base pairing or triple helix binding with a second and/or third nucleic acid. Such binding interaction may include Van der Waals interaction, hydrophobic interaction, hydrogen bonding and/or electrostatic interactions, for example. Techniques for identification and generation of aptamers is known in the art as described, for example, in F. M. Ausubel et al., Eds., *Short Protocols in Molecular Biology*, Current Protocols, Wiley, 2002; S. Klussman, Ed., *The Aptamer Handbook: Functional*

Oligonucleotides and Their Applications, Wiley, 2006; and J. Sambrook and D. W. Russell, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 3rd Ed., 2001.

[0064] According to aspects of the present disclosure, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein is a lectin.

[0065] According to aspects of the present disclosure, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein is a glycan binding enzyme, or a glycan specific binding fragment thereof.

[0066] According to aspects of the present disclosure, the glycan binding component is referred to as an O-GlcNAc binding protein.

[0067] According to aspects of the present disclosure, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein is a O-linked N-acetylglucosamine (O-GlcNAc) binding enzyme from the GH84 family of glycosylhydrolases, a glycan specific binding fragment thereof, or a variant of either thereof.

[0068] According to aspects of the present disclosure, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein is a O-GlcNAcase enzyme from the GH84 family of glycosylhydrolases, a glycan specific binding fragment thereof, or a variant of either thereof.

[0069] According to aspects of the present disclosure, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein is a human O-GlcNAcase (hOGA) enzyme, a glycan specific binding fragment thereof, or a variant of either thereof.

[0070] According to aspects of the present disclosure, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein is a human O-GlcNAcase (hOGA) enzyme includes the amino acid sequence:

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAGEDNPAGAGGA AVAGAAGGAR
RFLCGVVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMY SVE
EAEQLMTLISAAREYEIEFIY AISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFDDI
DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNVSQSPYLRT
VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY
KGRSTELIPRLKGVLTNP NCFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
IQIKLENEGSDEDIETDVL YSPQMALKLALTEWLQEF GVP HQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATT VTTVYQEPIMSQGAALS GEPTTLTKEEEKKQPDEEPM DMVVEK

QEETDHKNDNQILSEIVEAKMAEELKPMDDTKESIAESKSPMSMQEDCISDIAPMQTD
EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVVSVNCKGKDSEKIEEWSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
DIKSIMSMVKSFVQWLGCRRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ
PP (SEQ ID NO: 3)

or a variant thereof.

[0071] According to aspects of the present disclosure, the glycan binding component capable of specific binding to O-GlcNAc is a mutant O-GlcNAcase enzyme derived from a member of the GH84 family of glycosylhydrolases and lacking enzymatic glycosylhydrolase activity.

[0072] According to aspects of the present disclosure, the mutant O-GlcNAcase enzyme derived from a member of the GH84 family of glycosylhydrolases and lacking enzymatic glycosylhydrolase activity is a D174N mutant human O-GlcNAcase (hOGA) enzyme which includes the amino acid sequence:

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAGEDNPAGAGGA AVAGAAGGAR
RFLCGVVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMF WREMYSVE
EAEQLMTLISAAREYEIEFIY AISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDI
DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNVSQSPYLRT
VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY
KGRSTELIPRLKGVLTNP NCFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
IQIKLENEGSDEDIETDVL YSPQMALKLALTEWLQEFGVPHQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATTVVTTVYQEPIMSQGAALS GEPTTLTKEEEKKQPDEEPM DMVVEK
QEETDHKNDNQILSEIVEAKMAEELKPMDDTKESIAESKSPMSMQEDCISDIAPMQTD
EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVVSVNCKGKDSEKIEEWSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
DIKSIMSMVKSFVQWLGCRRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ
PP (SEQ ID NO: 4)

or a variant thereof.

[0073] Amino acid sequences and nucleic acid sequences are shown or described herein. Methods and compositions of the present invention are not limited to particular amino acid sequences and nucleic acid sequences identified herein and variants of a reference amino acid or nucleic acid sequence are encompassed.

[0074] As used herein, the term "variant" defines either an isolated naturally occurring mutant of a protein or nucleic acid, or a recombinantly prepared mutant of a protein or nucleic acid, each of which contain one or more mutations compared to a corresponding reference sequence, such as a wild-type sequence. For example, such mutations in a protein sequence can be one or more amino acid substitutions, additions, and/or deletions. In a further example, For example, such mutations in a nucleic acid sequence can be one or more nucleotide substitutions, additions, and/or deletions. The term "variant" further refers to orthologues.

[0075] The term "wild-type" refers to a naturally occurring, or unmutated, protein or nucleic acid.

[0076] According to aspects of the present disclosure, a variant protein includes an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or greater than 99%, identity with the reference amino acid sequence, and retains at least a substantial proportion (at least about 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99% or more) of the functional characteristics of the reference protein.

[0077] According to aspects of the present disclosure, a variant protein includes an amino acid sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identity, or greater than 99%, identity with the reference amino acid sequence, and retains at least a substantial proportion (at least about 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99% or more) of the functional characteristics of the reference protein.

[0078] To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino acid or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity=number of identical overlapping positions/total number of positions X100%). The two sequences compared are generally the same length or nearly the same length. Optionally, the two sequences are natural variants of a structural domain of a protein or two related proteins.

[0079] The determination of percent identity between two sequences can also be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul, 1990, PNAS 87:2264 2268, modified as in Karlin and Altschul, 1993, PNAS. 90:5873 5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul et al., 1990, J. Mol. Biol. 215:403. BLAST nucleotide searches are performed with the NBLAST nucleotide program parameters set, e.g., for score=100, wordlength=12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the present invention. BLAST protein searches are performed with the XBLAST program parameters set, e.g., to score 50, wordlength=3 to obtain amino acid sequences homologous to a protein molecule of the present invention. To obtain gapped alignments for comparison purposes, Gapped BLAST are utilized as described in Altschul et al., 1997, Nucleic Acids Res. 25:3389 3402. Alternatively, PSI BLAST is used to perform an iterated search which detects distant relationships between molecules (Id.). When utilizing BLAST, Gapped BLAST, and PSI Blast programs, the default parameters of the respective programs (e.g., of XBLAST and NBLAST) are used (see, e.g., the NCBI website). Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, 1988, CABIOS 4:11 17. Such an algorithm is incorporated in the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 is used.

[0080] The percent identity between two sequences is determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically only exact matches are counted.

[0081] One of skill in the art will recognize that one or more nucleotide or amino acid mutations can be introduced without altering the functional properties of a given nucleic acid or protein, respectively.

[0082] Mutations can be introduced using standard molecular biology techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis, to produce variants. For example, one or more amino acid substitutions, additions, or deletions can be made without altering the functional properties of a reference protein. When comparing a reference protein to a putative variant, amino acid similarity may be considered in addition to identity of amino acids at corresponding positions in an amino acid sequence. "Amino acid similarity" refers to amino acid

identity and conservative amino acid substitutions in a putative variant compared to the corresponding amino acid positions in a reference protein.

[0083] Conservative amino acid substitutions can be made or may be present in reference proteins to produce or identify variants.

[0084] Conservative amino acid substitutions are art recognized substitutions of one amino acid for another amino acid having similar characteristics. For example, each amino acid may be described as having one or more of the following characteristics: electropositive, electronegative, aliphatic, aromatic, polar/nonpolar, hydrophobic and hydrophilic. A conservative substitution is a substitution of one amino acid having a specified structural or functional characteristic for another amino acid having the same characteristic. Acidic amino acids include aspartate, glutamate; basic amino acids include histidine, lysine, arginine; aliphatic amino acids include isoleucine, leucine and valine; aromatic amino acids include phenylalanine, tyrosine and tryptophan; polar amino acids include aspartate, glutamate, histidine, lysine, asparagine, glutamine, arginine, serine, threonine and tyrosine; and hydrophobic amino acids include alanine, cysteine, phenylalanine, glycine, isoleucine, leucine, methionine, proline, valine and tryptophan; and conservative substitutions include substitution among amino acids within each group. Amino acids may also be described in terms of relative size; alanine, cysteine, aspartate, glycine, asparagine, proline, threonine, serine, valine are all typically considered to be small.

[0085] A variant can include synthetic amino acid analogs, amino acid derivatives and/or non-standard amino acids, illustratively including, without limitation, alphaaminobutyric acid, citrulline, canavanine, cyanoalanine, diaminobutyric acid, diaminopimelic acid, dihydroxy-phenylalanine, djenkolic acid, homoarginine, - 18 - 18 hydroxyproline, norleucine, norvaline, 3-phosphoserine, homoserine, 5-hydroxytryptophan, 1-methylhistidine, 3-methylhistidine, and ornithine.

[0086] It will be appreciated by those of ordinary skill in the art that, due to the degenerate nature of the genetic code, alternate nucleic acid sequences encode a specified protein such variant nucleic acid sequences may be used in compositions and methods described herein.

[0087] Protein variants are encoded by nucleic acids having a high degree of identity with a nucleic acid encoding a corresponding reference protein, such as a wild-type protein, or a corresponding portion thereof. The complement of a nucleic acid encoding a variant specifically hybridizes with a nucleic acid encoding a corresponding reference protein, such as a wild-type protein, under high stringency conditions.

[0088] The term “nucleic acid” refers to RNA or DNA molecules having more than one nucleotide in any form including single-stranded, double-stranded, oligonucleotide or polynucleotide. The term “nucleotide sequence” refers to the ordering of nucleotides in an oligonucleotide or polynucleotide in a single-stranded form of nucleic acid.

[0089] The term “complementary” refers to Watson-Crick base pairing between nucleotides and specifically refers to nucleotides hydrogen bonded to one another with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds. In general, a nucleic acid includes a nucleotide sequence described as having a “percent complementarity” to a specified second nucleotide sequence. For example, a nucleotide sequence may have 80%, 90%, or 100% complementarity to a specified second nucleotide sequence, indicating that 8 of 10, 9 of 10 or 10 of 10 nucleotides of a sequence are complementary to the specified second nucleotide sequence. For instance, the nucleotide sequence 3'-TCGA-5' is 100% complementary to the nucleotide sequence 5'-AGCT-3'. Further, the nucleotide sequence 3'-TCGA- is 100% complementary to a region of the nucleotide sequence 5'-TTAGCTGG-3'.

[0090] The terms “hybridization” and “hybridizes” refer to pairing and binding of complementary nucleic acids. Hybridization occurs to varying extents between two nucleic acids depending on factors such as the degree of complementarity of the nucleic acids, the melting temperature, T_m , of the nucleic acids and the stringency of hybridization conditions, as is well known in the art. The term “stringency of hybridization conditions” refers to conditions of temperature, ionic strength, and composition of a hybridization medium with respect to particular common additives such as formamide and Denhardt's solution. Determination of particular hybridization conditions relating to a specified nucleic acid is routine and is well known in the art, for instance, as described in J. Sambrook and D.W. Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press; 3rd Ed., 2001; and F.M. Ausubel, Ed., *Short Protocols in Molecular Biology*, Current Protocols; 5th Ed., 2002. High stringency hybridization conditions are those which only allow hybridization of substantially complementary nucleic acids. Typically, nucleic acids having about 85-100% complementarity are considered highly complementary and hybridize under high stringency conditions. Intermediate stringency conditions are exemplified by conditions under which nucleic acids having intermediate complementarity, about 50-84% complementarity, as well as those having a high degree of complementarity, hybridize. In contrast, low stringency hybridization conditions are those in which nucleic acids having a low degree of complementarity hybridize.

[0091] The terms “specific hybridization” and “specifically hybridizes” refer to hybridization of a particular nucleic acid to a target nucleic acid without substantial hybridization to nucleic acids other than the target nucleic acid in a sample.

[0092] Stringency of hybridization and washing conditions depends on several factors, including the T_m of the probe and target and ionic strength of the hybridization and wash conditions, as is well-known to the skilled artisan. Hybridization and conditions to achieve a desired hybridization stringency are described, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 2001; and Ausubel, F. et al., (Eds.), *Short Protocols in Molecular Biology*, Wiley, 2002.

[0093] An example of high stringency hybridization conditions is hybridization of nucleic acids over about 100 nucleotides in length in a solution containing 6X SSC, 5X Denhardt's solution, 30% formamide, and 100 micrograms/ml denatured salmon sperm at 37°C overnight followed by washing in a solution of 0.1X SSC and 0.1% SDS at 60°C for 15 minutes. SSC is 0.15M NaCl/0.015M Na citrate. Denhardt's solution is 0.02% bovine serum albumin/0.02% FICOLL/0.02% polyvinylpyrrolidone.

[0094] Nucleic acids encoding a protein, or a variant thereof, can be isolated or generated recombinantly or synthetically using well-known methodology.

[0095] The term “wild-type E. coli biotin ligase BirA” refers to the protein of SEQ ID NO:1.

MKDNTVPLKLIALLANGFHSGEQLGETLGMSRAAINKHQTLRDWGVDFVFTVPGKGY
SLPEPIQLLNAKQILGQLDGGSVAVLPVIDSTNQYLLDRIGELKSGDACIAEYQQAGRGR
RGRKWFSPFGANLYLSMFWRLEQGPAAAGLSLVIGIVMAEVLRLKLGADKVRVKWPN
DLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAMRRVEESVVNQGWITLQEAGINL
DRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFNRPVKLIIGDKEIFGISRGIDK
QGALLLEQDGIKPWMGGEISLRSAEK (SEQ ID NO:1)

[0096] The term “mutant E. coli biotin ligase BirA” refers to a mutant version of the protein of SEQ ID NO:1, the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein.

[0097] The protein of SEQ ID NO:2, or a variant thereof, is a “mutant E. coli biotin ligase BirA” having enzymatic activity to ligate biotin to proteins proximal to the target protein (also called miniTurboID, mTurbo, and mTurboID).

MIPLLNAKQILGQLDGGSVAVLPVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRSRGR
KWFSPFGANLYLSMFWRKRGPAAGLGPVIGIVMAEALRKLKLGADKVRVKWPNDLYL
QDRKLAGILVELAGITGDAAQIVIGAGINVAMRRVEESVVNQGWITLQEAGINLDRNTL
AAMLIRELRAALELFEQEGLAPYLSRWEKLDNFNRPVKLIIGDKEIFGISRGIDKQGALL
LEQDGVIPKPWMGGEISLRSAEK (SEQ ID NO:2)

[0098] The protein of SEQ ID NO:2, or a variant thereof, is a “mutant E. coli biotin ligase BirA” having enzymatic activity to ligate biotin to proteins proximal to the target protein.

IPLLNAKQILGQLDGGSVAVLPVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRGRK
WFSPFGANLYLSMFWRLKRGPA AIGLGPVIGIVMAEALRKLGADKVRVKWPNDLYLQ
DRKLAGILVELAGITGDAAQIVIGAGINVAMRRVEESVVNQGWITLQEAGINLDRNTLA
AMLIRELRAALELFEQEGLAPYLSRWEKLDNFNRPVKLIIGDKEIFGISRGIDKQGALLL
EQDGVIKPWMGGEISLRSAEK SEQ ID NO:11

[0099] According to aspects of the present disclosure, the mutant E. coli biotin ligase BirA comprises a variant of SEQ ID NO:1, having enzymatic activity to ligate biotin to proteins proximal to the target protein.

[00100] According to aspects of the present disclosure, the mutant E. coli biotin ligase BirA comprises SEQ ID NO:2, or a variant thereof having enzymatic activity to ligate biotin to proteins proximal to the target protein.

[00101] According to aspects of the present disclosure, the mutant E. coli biotin ligase BirA comprises SEQ ID NO:11, or a variant thereof having enzymatic activity to ligate biotin to proteins proximal to the target protein.

[00102] According to aspects of the present disclosure, the mutant E. coli biotin ligase BirA comprises at least a R118S mutation compared to wild-type E. coli biotin ligase BirA.

[00103] According to aspects of the present disclosure, the mutant E. coli biotin ligase BirA comprises at least a R118S mutation and deletion of 62 amino acids from the N-terminus compared to wild-type E. coli biotin ligase BirA.

[00104] According to aspects of the present disclosure, the mutant E. coli biotin ligase BirA comprises at least Q65P, I87V, R118S, Q141R, E149K, S150G, L151P, V160A, T192A, K194I, M209V, S236P, M241T, and I305V mutations compared to wild-type E. coli biotin ligase BirA.

[00105] According to aspects of the present disclosure, the mutant E. coli biotin ligase BirA comprises at least Q65P, I87V, R118S, Q141R, E149K, S150G, L151P, V160A, T192A, K194I, M209V, S236P, M241T, and I305V mutations and deletion of 62 amino acids from the N-terminus compared to wild-type E. coli biotin ligase BirA.

[00106] According to aspects of the present disclosure, the glycan binding component has a C-terminus and an N-terminus, the mutant E. coli biotin ligase BirA has a C-terminus and an N-terminus, and the C-terminus of the glycan binding component is linked to the N-terminus of the mutant E. coli biotin ligase BirA.

[00107] A linker is disposed between, and linked to each of, two components of a fusion protein according to aspects of the present disclosure, thereby linking the two components through the linker.

[00108] According to aspects of the present disclosure, an included linker is, or includes about 1 to 100 amino acids, such as about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 amino acids. According to aspects of the present disclosure, is a peptide including about 1 to 100 amino acids, such as about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 amino acids.

[00109] According to aspects of the present disclosure, the linker can be , or include, a bond, an atom, a multi-atom group, or a chain of atoms. Non-limiting examples of a linker which is an atom are oxygen, and sulfur. A non-limiting example of a linker which is a multi-atom group is C(O).

[00110] According to aspects of the present disclosure, an included linker is, or includes, a chain of atoms such as, branched or linear chain of 2- 20, or more, atoms. According to aspects of the present disclosure, a linker is, or includes, a linear chain of 3- 20, or more, atoms. According to aspects of the present disclosure, a linker is, or includes, a linear chain of 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 atoms. According to aspects of the present disclosure, a linker is, or includes, a linear chain of 6, 7, 8, 9, 10, 11, or 12 atoms. According to aspects of the present disclosure, a linker is, or includes, a linear chain of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 atoms. According to aspects of the present disclosure, a linker is, or includes, a linear chain of 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 atoms. According to aspects of the present disclosure, a linker is, or includes, a linear chain of 10, 11, 12, 13, 14, or 15 atoms.

[00111] According to aspects of the present disclosure, an included linker is, or includes, a chain of atoms such as, but not limited to, substituted or unsubstituted C1-C20 alkyl, substituted or unsubstituted C2-C20 alkenyl, substituted or unsubstituted C2-C20 alkynyl, substituted or unsubstituted C6-C12 aryl, substituted or unsubstituted C3-C12 cycloalkyl, substituted or unsubstituted C5-C12 heteroaryl, or substituted or unsubstituted C5-C12 heterocyclyl. According to aspects of the present disclosure, the chain includes at least one hydroxyl functional group, and preferably at least one terminal hydroxyl functional group. The term “terminal hydroxyl functional group” as used herein refers to a hydroxyl group on the final atom

of a chain of atoms in a linker, i.e. the atom of the chain of atoms which is furthest from the magnetic particle, or within no more than 2, 3, or 4 atoms from the final atom of the chain of atoms in a linker.

[00112] According to aspects of the present disclosure, a linker is disposed between, and linked to each of, a glycan binding component and a mutant E. coli biotin ligase BirA, thereby linking the glycan binding component and the mutant E. coli biotin ligase BirA through the linker.

[00113] According to aspects of the present disclosure, the glycan binding component is linked to the mutant E. coli biotin ligase BirA by a linker disposed between the glycan binding component and the mutant E. coli biotin ligase BirA.

[00114] According to aspects of the present disclosure, the fusion protein includes a localization signal peptide. According to aspects of the present disclosure, the localization signal peptide is capable of promoting localization of the fusion protein to a subcellular compartment selected from the group consisting of: nucleus, cytosol, mitochondria, endoplasmic reticulum, and plasma membrane. One or more localization signal peptides are included in a fusion protein according to aspects of the present disclosure. Optionally, a linker is disposed between a localization signal peptide and the glycan binding component and/or linker is disposed between the localization signal peptide and the mutant E. coli biotin ligase BirA. When two or more localization signal peptides are included in a fusion protein according to aspects of the present disclosure, a linker is optionally disposed between the two or more localization signal peptides.

[00115] According to aspects of the present disclosure, the localization signal peptide is a mitochondrial localization signal. An exemplary mitochondrial localization signal is MLGFVGRVAAAPASGALRRLTPSASLPPAQLLLRAAPTAVHPVRDYA (SEQ ID NO:5), or a variant thereof.

[00116] According to aspects of the present disclosure, the localization signal peptide is a plasma membrane localization signal. An exemplary plasma membrane localization signal is MGCINSKRK, (SEQ ID NO:6), or a variant thereof.

[00117] According to aspects of the present disclosure, the localization signal peptide is a nuclear localization signal. An exemplary nuclear localization signal (NLS) is PKKKRKV (SEQ ID NO:7), or a variant thereof.

[00118] According to aspects of the present disclosure, the localization signal peptide is a cytoplasm localization signal. An exemplary cytoplasm localization signal (NES) is LPPLERTL, (SEQ ID NO:8), or a variant thereof.

[00119] According to aspects of the present disclosure, the fusion protein includes an exogenous detectable tag, such as a V5 tag or HA tag, or a variant of either thereof. An exemplary V5 tag is GKIPNPLLGLDST, (SEQ ID NO:9), or a variant thereof. An exemplary HA tag is YPYDVDPDYA (SEQ ID NO:10), or a variant thereof.

[00120] An expression construct is provided according to aspects of the present disclosure which includes a nucleic acid encoding a fusion protein including a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein.

[00121] The term “expression construct” is used herein to refer to a double-stranded recombinant DNA molecule containing a nucleic acid desired to be expressed and containing appropriate regulatory elements necessary or desirable for the transcription of the operably linked nucleic acid sequence in vitro or in vivo. The term “recombinant” is used to indicate a nucleic acid construct in which two or more nucleic acids are linked and which are not found linked in nature. The term “expressed” refers to transcription of a nucleic acid to produce a corresponding mRNA and/or translation of the mRNA to produce the corresponding protein. Expression constructs can be generated recombinantly or synthetically or by DNA synthesis using well-known methodology.

[00122] An expression construct is introduced into a cell using well-known methodology, such as, but not limited to, by introduction of a vector containing the expression construct into the cell. A “vector” is a nucleic acid that transfers an inserted nucleic acid into a host cell and/or between host cells becoming self-replicating. The term includes vectors that function primarily for insertion of a nucleic acid into a cell, replication of vectors that function primarily for the replication of nucleic acid, and expression vectors that function for transcription and/or translation of a nucleic acid. Also included are vectors that provide more than one of the above functions. Such cells include, but are not limited to cells of a microorganism, such as a bacterial cell, yeast, a human cell, a cell of a non-human mammal, a vertebrate cell, an invertebrate cell, a microorganism, or a plant cell. Such cells may be cell lines or primary cells.

[00123] Vectors include plasmids, viruses, BACs, YACs, and the like. Particular viral vectors illustratively include those derived from adenovirus, adeno-associated virus and lentivirus.

[00124] The term “regulatory element” as used herein refers to a nucleotide sequence which controls some aspect of the expression of an operably linked nucleic acid. Exemplary regulatory elements illustratively include an enhancer, an internal ribosome entry site (IRES), an intron; an origin of replication, a polyadenylation signal (pA), a promoter, a transcription termination sequence, and an upstream regulatory domain, which contribute to the replication, transcription, post-transcriptional processing of a nucleic acid. Those of ordinary skill in the art are capable of selecting and using these and other regulatory elements in an expression construct with no more than routine experimentation.

[00125] The term “promoter” as used herein refers to a DNA sequence operably linked to a nucleic acid to be transcribed such as a nucleic acid encoding a desired molecule. A promoter is generally positioned upstream of a nucleic acid sequence to be transcribed and provides a site for specific binding by RNA polymerase and other transcription factors.

[00126] In addition to a promoter, one or more enhancer sequences may be included such as, but not limited to, cytomegalovirus (CMV) early enhancer element and an SV40 enhancer element. Additional included sequences are an intron sequence such as the beta globin intron or a generic intron, a transcription termination sequence, and an mRNA polyadenylation (pA) sequence such as, but not limited to SV40-pA, beta-globin-pA, the human growth hormone (hGH) pA and SCF-pA. The term “polyA” or “p(A)” or “pA” refers to nucleic acid sequences that signal for transcription termination and mRNA polyadenylation. The polyA sequence is characterized by the hexanucleotide motif AAUAAA. Commonly used polyadenylation signals are the SV40 pA, the human growth hormone (hGH) pA, the beta-actin pA, and beta-globin pA. The sequences can range in length from 32 to 450 bp. Multiple pA signals may be used.

[00127] The term “operably linked” as used herein refers to a nucleic acid in functional relationship with a second nucleic acid. The term “operably linked” encompasses functional connection of two or more nucleic acids, such as an oligonucleotide or polynucleotide to be transcribed and a regulatory element such as a promoter or an enhancer element, which allows transcription of the nucleic acid to be transcribed.

[00128] A Kozak consensus sequence can be included in an expression construct. A Kozak consensus sequence is well-known as a nucleic acid motif that functions as the protein translation initiation site in most eukaryotic mRNA transcripts and can be included in an expression construct.” An example of a Kozak consensus sequence is GGATCCGCCACC (SEQ ID NO:34).

[00129] An expression construct provided according to aspects of the present disclosure includes a nucleic acid encoding a fusion protein including a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein is or includes a nucleic acid sequence encoding:

(A)

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAPGEDNPAGAGGA AVAGAAGGAR
RFLCGVVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMY SVE
EAEQLMTLISAAREYEIEFIY AISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDI
DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNVSQSPYLRT
VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY
KGRSTELIPRLKGVLTNP NCEFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
IQIKLENEGSDEDIETDVL YSPQMALKLALTEWLQEF GVP HQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATTVVTTVYQEPIMSQGAALS GEPTTLTKEEEKKQPDEEPM DMVVEK
QEETDHKNDNQILSEIVEAKMAEELKPM DTDKESIAESKSP EMSMQEDCISDIAPMQTD
EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVSVVNCKGKDSEKIEE WRSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
DIKSIMSMVKSFVQWLGCRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ
PP (SEQ ID NO:4),

or a variant thereof
and

(B)

IPLLNAKQILGQLDGGSVAVLPVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRGRK
WFSPFGANLYLSMFWR LKRGPAAIGLGPVIGIVMAEALRKL GADKVRVKWPNDLYLQ
DRKLAGILVELAGITGDAAQIVIGAGIN VAMRRVEESVVNQGWITLQEAGINLDRNTLA
AMLIRELRAALELFEQEGLAPYLSRWEKLDNF INRPVKLIIGDKEIFGISRGIDKQGALLL
EQDGVIKPWMGGEISLRSAEK (SEQ ID NO:11),

or a variant thereof

wherein the two component proteins are linked directly or through a linker as (A)-(B), and wherein the expression construct optionally further encodes a localization signal peptide and/or exogenous tag in operable linkage with one or both component proteins.

[00130] An expression construct provided according to aspects of the present disclosure includes a nucleic acid encoding a fusion protein including a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein is or includes a nucleic acid sequence encoding:

(A)

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAGEDNPAGAGGA AVAGAAGGAR
 RFLCGVVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMYSVE
 EAEQLMTLISAAREYEIEFIYAISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDI
 DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNV SQSPYLRT
 VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY
 KGRSTELIPRLKGVLTNPNCEFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
 IQIKLENEGSDIEDITDVL YSPQMALKLALTEWLQEFVPHQYSSRQVAHSGAKASVVD
 GTPLVAAPSLNATTVTTVYQEPIMSQGAALSGEPTTLTKEEEKKQPDEEPM DMVVEK
 QEETDHKNDNQILSEIVEAKMAEELKPM DTDKESIAESKSP EMSMQEDCISDIAPMQTD
 EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWL RAN
 SSVVSVNCKGKDSEKIEE WRSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
 DIKSIMSMVKS FVQWLGC RSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ
 PP (SEQ ID NO:4),

or a variant thereof
 and

(B)

MIPLLNAKQILGQLDGGSVAVLPVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRGR
 KWFSPFGANLYLSMFWRLKRGPA AIGLPVIGIVMAEALRKL GADKVRVKWPNDLYL
 QDRKLAGILVELAGITGDAAQIVIGAGINVAMRRVEESVVNQGWITLQEAGINLDRNTL
 AAMLIRELRAALELFEQEGLAPYLSRWEKLDNF INRPVKLIIGDKEIFGISRGIDKQGALL
 LEQDGVIPKPMWGGEISLRS AEK (SEQ ID NO:2),

or a variant thereof

wherein the two component proteins are linked directly or through a linker as (B)-(A), and
 wherein the expression construct optionally further encodes a localization signal peptide and/or
 exogenous tag in operable linkage with one or both component proteins.

[00131] An expression construct provided according to aspects of the present disclosure
 includes a nucleic acid encoding a fusion protein including a glycan binding component linked
 to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding
 to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin
 ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein is
 or includes a nucleic acid sequence encoding:

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAGEDNPAGAGGA AVAGAAGGAR
 RFLCGVVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMYSVE
 EAEQLMTLISAAREYEIEFIYAISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDI
 DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNV SQSPYLRT
 VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY
 KGRSTELIPRLKGVLTNPNCEFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
 IQIKLENEGSDIEDITDVL YSPQMALKLALTEWLQEFVPHQYSSRQVAHSGAKASVVD
 GTPLVAAPSLNATTVTTVYQEPIMSQGAALSGEPTTLTKEEEKKQPDEEPM DMVVEK
 QEETDHKNDNQILSEIVEAKMAEELKPM DTDKESIAESKSP EMSMQEDCISDIAPMQTD
 EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWL RAN

SSVSVNCKGKDSEKIEEWSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
 DIKSIMSMVKSFVQWLGCRRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ
 PPMIPLLNQILGQLDGGSVAVLPVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRSR
 GRKWSPFGANLYLSMFWRLKRGPAIIGLGPVIGIVMAEALRKLGADKVRVKWPNDL
 YLQDRKLAGILVELAGITGDAAQIVIGAGINVAMRRVEESVVNQGWITLQEAGINLDRN
 TLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFNRPVKLIIGDKEIFGISRGIDKQGA
 LLEQDGVIPKPMGGEISLRSAEK (SEQ ID NO:12),

or a variant thereof.

[00132] An expression construct provided according to aspects of the present disclosure includes a nucleic acid encoding a fusion protein including a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein is or includes a nucleic acid sequence encoding:

MIPLLNQILGQLDGGSVAVLPVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRSRGR
 KWSPFGANLYLSMFWRLKRGPAIIGLGPVIGIVMAEALRKLGADKVRVKWPNDLYL
 QDRKLAGILVELAGITGDAAQIVIGAGINVAMRRVEESVVNQGWITLQEAGINLDRNTL
 AAMLIRELRAALELFEQEGLAPYLSRWEKLDNFNRPVKLIIGDKEIFGISRGIDKQGALL
 LEQDGVIPKPMGGEISLRSAEKMVQKESQATLEERESELSSNPAASAGASLEPPAAPAP
 GEDNPAGAGGAAGVAGAAGGARRFLCGVVEGFYGRPVMQQRKELFRRLQKWELNTY
 LYAPKDDYKHRMFWRMYSVEEAEQLMTLISAAREYEIEFIYAI SPGLDITFSNPKEVST
 LKRKLDQVSQFGCRSFALLFNDIDHNMCAADKEVFSSFAHAQVSITNEIYQYLGEPEFTL
 FCPTEYCGTFCYPNVQSPLYLRTVGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKIIRAP
 VIWDNIHANDYDQKRLFLGPYKGRSTELIPRLKGVLTNPNCFEANYVAIHTLATWYKS
 NMNGVRKDVVMTDSEDSTVSIQIKLENEGSDIEDTDLVLYSPQMALKLALTEWLQEFQ
 VPHQYSSRQVAHSGAKASVVDGTPVLAAPSLNATTVVTTVYQEPIMSQGAALSGETT
 LTKEEEKKQPDDEPMVMVVEKQEETDHKNDNQILSEIVEAKMAEELKPMDDTKESIAE
 SKSPEMSMQEDCISDIAPMQTDEQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYL
 PYEHGPKGAQMLREFQWLRANSSVSVNCKGKDSEKIEEWSRAAKFEEMCGLVMG
 MTRLSNCANRTILYDMYSYVWDIKSIMSMVKSFVQWLGCRRSHSSAQFLIGDQEPWAF
 RGGLAGEFQRLLPIDGANDLFFQPP (SEQ ID NO:13), or a variant thereof.

[00133] An expression construct provided according to aspects of the present disclosure includes a nucleic acid encoding: hOGA*-6X-mTurbo-flag, D174N mutant:

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAPGEDNPAGAGGAAGVAGAAGGAR
 RFLCGVVEGFYGRPVMQQRKELFRRLQKWELNTYLYAPKDDYKHRMFWRMYSVE

DIKSIMSMVKSFVQWLGCRRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQPP-GGGGS-DYKDDDDK (SEQ ID NO:15), or a variant thereof.

[00135] An expression construct provided according to aspects of the present disclosure includes a nucleic acid encoding: hOGA*-6X-mTurbo-flag, D174A mutant:

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAGEDNPAGAGGA AVAGAAGGAR
RFLCGVVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMYSVE
EAEQLMTLISAAREYEIEFIY AISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFADI
DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNV SQSPYLRT
VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY
KGRSTELIPRLKGVLTNP NCEFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
IQIKLENEGSDEDIETDVL YSPQMALKLALTEWLQEF GVP HQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATT VTTVYQEPIMSQGAALSGEPTTLTKEEEKKQPDEEPM DMVVEK
QEETHDKNDNQILSEIVEAKMAEELKPM DTDKESIAESKSP EMSMQEDCISDIAPMQTD
EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVVS VNCKGKDSEKIEE WRSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
DIKSIMSMVKSFVQWLGCRRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQPP-GGGGS-

MIPLLNAKQILGQLDGGSVAVLPVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRGR
KWFSPFGANLYLSMFWRLKRGPA AIGLGPVIGIVMAEALRKLGADKVRVKWPNDLYL
QDRKLAGILVELAGITGDAAQIVIGAGIN VAMRRVEESV VNQGWITLQEAGINLDRNTL
AAMLIRELRAALELFEQEGLAPYLSRWEKLDNF INRPVKLIIGDKEIFGISRGIDKQGALL
LEQDGVIPKPMWGGEISLRSAEK-GGGGS-DYKDDDDK (SEQ ID NO:16), or a variant thereof.

[00136] An expression construct provided according to aspects of the present disclosure includes a nucleic acid encoding: mCherry-TurboID (control construct)

MVSKGEEDNMAIIEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGP
LPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDS
SLQDGEFIYKVKLRGTNFP SDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLK LK
DGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGM
DELYK-GGGGS-

MIPLLNAKQILGQLDGGSVAVLPVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRGR
KWFSPFGANLYLSMFWRLKRGPA AIGLGPVIGIVMAEALRKLGADKVRVKWPNDLYL

CCTGGACAGGAATACTCTGGCCGCTATGCTGATCCGAGAGCTGCGGGCAGCCCTGG
AACTGTTTCGAGCAGGAAGGCCTGGCTCCATATCTGTACGGTGGGAGAAGCTGGAT
AACTTCATCAATAGACCCGTGAAGCTGATCATTGGGGACAAAGAGATTTTCGGGAT
TAGCCGGGGGATTGATAAACAGGGAGCCCTGCTGCTGGAACAGGACGGAGTTATCA
AACCCTGGATGGGCGGAGAAATCAGTCTGCGGTCTGCCGAAAAG (SEQ ID NO:19).

[00139] A nucleic acid encoding a NLS is:

GACCCCAAGAAGAAGAGGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGGACCCC
AAGAAGAAGAGGAAGGTG (SEQ ID NO:20).

[00140] A nucleic acid encoding an NES is:
CTGCCTCCCCTGGAGCGCCTGACCCTGGAC (SEQ ID NO: 21).

[00141] Nucleic acids encoding a linker are:
GCGGCCGCCACCATGTACCCGTATGATGTTCCGGATTACGCTGGCTATCCCTACGAC
GTGCCCCGACTATGCCGGGTACCCCTATGACGTCCCAGACTACGCAGCTAGC (SEQ ID
NO:22), CTGCAG, and
AAGCTTGCGGCCGCCACCATGGGCAAGCCCATCCCCAACCCCTGCTGGGCCTGGA
CAGCACCGCTAGC (SEQ ID NO: 23).

[00142] A tag can be included such as one or more copies of GGGGS (SEQ ID NO: 35),
such as 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, or more, copies. For example, a 6X tag is:
GGGGS GGGGS GGGGS GGGGS GGGGS GGGGS (SEQ ID NO: 24). A tag can have various
functions, including as a linker, for example providing flexibility, and/or for detection of a
fusion protein, e.g. using an anti-GGGGS antibody

[00143] A flag tag can be included such as DYKDDDDK (SEQ ID NO: 25). A FLAG tag
can be used for such functions as to facilitate protein purification, detection, and localization.
For example a FLAG tag can be fused to the N- or C-terminus of a protein of interest, allowing
for easy identification and isolation of the protein using an anti-FLAG antibody.

[00144] Nucleic acids encoding a protein, or a variant thereof, can be isolated or generated
recombinantly or synthetically using well-known methodology.

[00145] According to aspects of the present disclosure, contacting the living cell with the
fusion protein comprises introducing an expression construct encoding the fusion protein into
the cell.

[00146] The expression construct may be transfected into cells using well-known methods,
such as electroporation, calcium-phosphate precipitation transfection, and lipofection. The cells
are screened for presence and/or integration by DNA analysis, such as PCR, Southern blot or

sequencing. Cells with the expression construct can be screened for functional expression, for example ELISA or Western blot analysis.

[00147] Cells are provided according to aspects of the present disclosure which include the expression construct which includes a nucleic acid encoding a fusion protein including a glycan binding component linked to a mutant *E. coli* biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant *E. coli* biotin.

[00148] Methods of detecting proteins proximal to a target protein according to aspects of the present disclosure include: providing a fusion protein including a glycan binding component linked to a mutant *E. coli* biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant *E. coli* biotin; contacting the fusion protein with a living cell under compatible biological conditions, whereby the fusion protein specifically binds to a glycosylation post-translational modification of a target protein of the cell; providing biotin to the living cell, whereby the mutant *E. coli* biotin ligase BirA ligates biotin to proteins proximal to the target protein; and detecting the biotinylated proteins, thereby detecting proteins proximal to the target protein.

[00149] The term “compatible biological conditions” refers to conditions which are compatible with living cells and which do not interfere with the desired function and localization of the fusion protein. Physiological conditions is a general term signifying compatible biological conditions and such conditions are well-known.

[00150] According to aspects of the present disclosure, detecting the biotinylated proteins includes purifying the biotinylated proteins and detecting the purified biotinylated proteins.

[00151] The term “purifying” in the context of purifying the biotinylated proteins for detection refers to separation of the biotinylated proteins from at least one other component present in the system in which the biotinylated proteins were produced. For example, biotinylated proteins are separated from cells in which they are produced, generating purified biotinylated proteins.

[00152] According to aspects, the purified biotinylated proteins make up at least about 0.01 - 100% of the mass, by weight, such as about 0.01%, 0.1%, 1%, 5%, 10%, 25%, 50%, 75%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or greater than about 99% of the mass, by weight, of material in a sample of purified biotinylated proteins. Such purification is achieved by techniques illustratively including salt, pH, hydrophobic or affinity precipitation, electrophoretic methods such as gel electrophoresis and 2-D gel electrophoresis; chromatography methods such as

HPLC, ion exchange chromatography, affinity chromatography, size exclusion chromatography, thin layer and paper chromatography.

[00153] According to aspects of the present disclosure, detecting the purified biotinylated proteins includes mass spectrometry.

[00154] According to aspects of the present disclosure, mass spectrometry is used in a method for detecting the purified biotinylated proteins. A variety of configurations of mass spectrometers can be used in a method of the present disclosure. In general, a mass spectrometer has the following major components: a sample inlet, an ion source, a mass analyzer, a detector, a vacuum system, and instrument-control system, and a data system. Common mass analyzers include a quadrupole mass filter, ion trap mass analyzer and time-of-flight mass analyzer.

[00155] The ion formation process is a starting point for mass spectrum analysis and several ionization methods are available. For example, electrospray ionization (ESI) can be used. Generally described, in ESI a solution containing the material to be analyzed is passed through a fine needle at high potential which creates a strong electrical field resulting in a fine spray of highly charged droplets that is directed into the mass spectrometer. Other ionization procedures include, for example, fast-atom bombardment (FAB) which uses a high-energy beam of neutral atoms to strike a solid sample causing desorption and ionization. Matrix-assisted laser desorption ionization (MALDI) is a method in which a laser pulse is used to strike a sample that has been crystallized in an UV-absorbing compound matrix. Other ionization procedures known in the art include, for example, plasma and glow discharge, plasma desorption ionization, resonance ionization, and secondary ionization.

[00156] Electrospray ionization (ESI) has several properties that are useful for methods of assessing an analyte of the present disclosure. For example, the efficiency of ESI can be very high which provides the basis for highly sensitive measurements. Furthermore, ESI produces charged molecules from solution, which is convenient for analyzing analytes and standards that are in solution. In contrast, ionization procedures such as MALDI require crystallization of the material to be analyzed prior to ionization.

[00157] Since ESI can produce charged molecules directly from solution, it is compatible with samples from liquid chromatography systems. In liquid chromatography with tandem mass spectrometry (LC-MS-MS), the inlet can be a capillary-column liquid chromatography source. For example, a mass spectrometer can have an inlet for a liquid chromatography system, such as an HPLC, so that fractions flow from the chromatography column into the mass spectrometer. This in-line arrangement of a liquid chromatography system and mass spectrometer is

sometimes referred to as LC-MS. An LC-MS system can be used, for example, to separate analytes and standards from complex mixtures before mass spectrometry analysis. In addition, chromatography can be used to remove salts or other buffer components from the sample before mass spectrometry analysis. For example, desalting of a sample using a reversed-phase HPLC column, in-line or off-line, can be used to increase the efficiency of the ionization process and thus improve sensitivity of detection by mass spectrometry.

[00158] A variety of mass analyzers are available that can be paired with different ion sources. Different mass analyzers have different advantages as known to one skilled in the art and as described herein. The mass spectrometer and methods chosen for detection depends on the particular assay, for example, a more sensitive mass analyzer can be used when a small amount of ions are generated for detection. Several types of mass analyzers and mass spectrometry methods are described below.

[00159] Quadrupole mass spectrometry utilizes a quadrupole mass filter or analyzer. This type of mass analyzer is composed of four rods arranged as two sets of two electrically connected rods. A combination of rf and dc voltages are applied to each pair of rods which produces fields that cause an oscillating movement of the ions as they move from the beginning of the mass filter to the end. The result of these fields is the production of a high-pass mass filter in one pair of rods and a low-pass filter in the other pair of rods. Overlap between the high-pass and low-pass filter leaves a defined m/z that can pass both filters and traverse the length of the quadrupole. This m/z is selected and remains stable in the quadrupole mass filter while all other m/z have unstable trajectories and do not remain in the mass filter. A mass spectrum results by ramping the applied fields such that an increasing m/z is selected to pass through the mass filter and reach the detector. In addition, quadrupoles can also be set up to contain and transmit ions of all m/z by applying a rf-only field. This allows quadrupoles to function as a lens or focusing system in regions of the mass spectrometer where ion transmission is needed without mass filtering. This will be of use in tandem mass spectrometry as described further below.

[00160] A quadrupole mass analyzer, as well as the other mass analyzers described herein, can be programmed to analyze a defined m/z or mass range. Since the mass range of analytes and standards will be known prior to an assay, a mass spectrometer can be programmed to transmit ions of the projected correct mass range while excluding ions of a higher or lower mass range. The ability to select a mass range can decrease the background noise in the assay and thus increase the signal-to-noise ratio as well as increasing the specificity of the assay. Therefore, the

mass spectrometer can accomplish an inherent separation step as well as detection and identification of analytes and standards.

[00161] Ion trap mass spectrometry utilizes an ion trap mass analyzer. In these mass analyzers, fields are applied so that ions of all m/z are initially trapped and oscillate in the mass analyzer. Ions enter the ion trap from the ion source through a focusing device such as an octapole lens system. Ion trapping takes place in the trapping region before excitation and ejection through an electrode to the detector. Mass analysis is accomplished by sequentially applying voltages that increase the amplitude of the oscillations in a way that ejects ions of increasing m/z out of the trap and into the detector. In contrast to quadrupole mass spectrometry, all ions are retained in the fields of the mass analyzer except those with the selected m/z . One advantage to ion traps is that they have very high sensitivity, as long as one is careful to limit the number of ions being tapped at one time. Control of the number of ions can be accomplished by varying the time over which ions are injected into the trap. The mass resolution of ion traps is similar to that of quadrupole mass filters, although ion traps do have low m/z limitations.

[00162] Time-of-flight mass spectrometry utilizes a time-of-flight mass analyzer. For this method of m/z analysis, an ion is first given a fixed amount of kinetic energy by acceleration in an electric field (generated by high voltage). Following acceleration, the ion enters a field-free or "drift" region where it travels at a velocity that is inversely proportional to its m/z . Therefore, ions with low m/z travel more rapidly than ions with high m/z . The time required for ions to travel the length of the field-free region is measured and used to calculate the m/z of the ion. One consideration in this type of mass analysis is that the set of ions being studied be introduced into the analyzer at the same time. For example, this type of mass analysis is well suited to ionization techniques like MALDI which produces ions in short well-defined pulses. Another consideration is to control velocity spread produced by ions that have variations in their amounts of kinetic energy. The use of longer flight tubes, ion reflectors, or higher accelerating voltages can help minimize the effects of velocity spread. Time-of-flight mass analyzers have a high level of sensitivity and a wider m/z range than quadrupole or ion trap mass analyzers. Also data can be acquired quickly with this type of mass analyzer because no scanning of the mass analyzer is necessary.

[00163] Tandem mass spectrometry can utilize combinations of the mass analyzers described above.

[00164] Tandem mass spectrometers can use a first mass analyzer to separate ions according to their m/z in order to isolate an ion of interest for further analysis. The isolated ion of interest

is then broken into fragment ions, called collisionally activated dissociation or collisionally induced dissociation, and the fragment ions are analyzed by the second mass analyzer. These types of tandem mass spectrometer systems are called tandem in space systems because the two mass analyzers are separated in space, usually by a collision cell. Tandem mass spectrometer systems also include tandem in time systems where one mass analyzer is used, however the mass analyzer is used sequentially to isolate an ion, induce fragmentation, and then perform mass analysis.

[00165] Mass spectrometers in the tandem in space category have more than one mass analyzer. For example, a tandem quadrupole mass spectrometer system can have a first quadrupole mass filter, followed by a collision cell, followed by a second quadrupole mass filter and then the detector. Another arrangement is to use a quadrupole mass filter for the first mass analyzer and a time-of-flight mass analyzer for the second mass analyzer with a collision cell separating the two mass analyzers.

[00166] Other tandem systems are known in the art including reflection-time-of-flight, tandem sector and sector-quadrupole mass spectrometry.

[00167] Mass spectrometers in the tandem in time category have one mass analyzer that performs different functions at different times. For example, an ion trap mass spectrometer can be used to trap ions of all m/z . A series of rf scan functions are applied which ejects ions of all m/z from the trap except the m/z of ions of interest. After the m/z of interest has been isolated, an rf pulse is applied to produce collisions with gas molecules in the trap to induce fragmentation of the ions. Then the m/z values of the fragmented ions are measured by the mass analyzer. Ion cyclotron resonance instruments, also known as Fourier transform mass spectrometers, are an example of tandem-in-time systems.

[00168] Several types of tandem mass spectrometry experiments can be performed by controlling the ions that are selected in each stage of the experiment. The different types of experiments utilize different modes of operation, sometimes called "scans," of the mass analyzers. In a first example, called a mass spectrum scan, the first mass analyzer and the collision cell transmit all ions for mass analysis into the second mass analyzer. In a second example, called a product ion scan, the ions of interest are mass-selected in the first mass analyzer and then fragmented in the collision cell. The ions formed are then mass analyzed by scanning the second mass analyzer. In a third example, called a precursor ion scan, the first mass analyzer is scanned to sequentially transmit the mass analyzed ions into the collision cell for fragmentation. The second mass analyzer mass-selects the product ion of interest for

transmission to the detector. Therefore, the detector signal is the result of all precursor ions that can be fragmented into a common product ion. Other experimental formats include neutral loss scans where a constant mass difference is accounted for in the mass scans. The use of these different tandem mass spectrometry scan procedures can be advantageous when large sets of analytes are measured in a single experiment.

[00169] In view of the above, those skilled in the art recognize that different mass spectrometry methods, for example, quadrupole mass spectrometry, ion trap mass spectrometry, time-of-flight mass spectrometry and tandem mass spectrometry, can use various combinations of ion sources and mass analyzers which allows for flexibility in designing customized detection protocols. In addition, mass spectrometers can be programmed to transmit all ions from the ion source into the mass spectrometer either sequentially or at the same time. Furthermore, a mass spectrometer can be programmed to select ions of a particular mass for transmission into the mass spectrometer while blocking other ions. The ability to precisely control the movement of ions in a mass spectrometer allows for greater options in detection protocols which can be advantageous when a large number of analytes are being analyzed.

[00170] Different mass spectrometers have different levels of resolution, that is, the ability to resolve peaks between ions closely related in mass. The resolution is defined as $R=m/\Delta m$, where m is the ion mass and Δm is the difference in mass between two peaks in a mass spectrum. For example, a mass spectrometer with a resolution of 1000 can resolve an ion with a m/z of 100.0 from an ion with a m/z of 100.1. Those skilled in the art will therefore select a mass spectrometer having a resolution appropriate for the analyte(s) to be detected.

[00171] Mass spectrometers can resolve ions with small mass differences and measure the mass of ions with a high degree of accuracy. Therefore, analytes of similar masses can be used together in the same experiment since the mass spectrometer can differentiate the mass of even closely related molecules. The high degree of resolution and mass accuracy achieved using mass spectrometry methods allows the use of large sets of analytes because they can be distinguished from each other.

[00172] Mass spectrometry devices and general methods of their use are well known in the art as exemplified in McMaster, M., *LC/MS A Practical User's Guide*, 2005, John Wiley & Sons, USA; and Hoffmann and Stroobant, *Mass Spectrometry Principles and Applications*, 2007, John Wiley & Sons, England.

[00173] According to aspects of the present disclosure, detecting the purified biotinylated proteins includes chromatography. According to aspects of the present disclosure, detecting the

purified biotinylated proteins includes gel electrophoresis. According to aspects of the present disclosure, detecting the purified biotinylated proteins includes gel electrophoresis and transfer of the electrophoresed purified biotinylated proteins to a membrane.

[00174] One or more controls or standards can be used to detect biotinylated proteins and/or compare one or more biotinylated proteins obtained under different conditions, e.g. before and after treatment of cells with a test substance.

[00175] A test substance may be a natural or synthetic chemical compound, nucleic acid, peptide, protein, saccharide, oligosaccharide, polysaccharide, lipid, or combination of any two or more thereof. Extracts of plants which contain several characterized or uncharacterized components may be a test substance. According to aspects, the test substance is an antisense molecule, an aptamer, siRNA, shRNA, miRNA, a DNzyme, or a ribozyme.

[00176] Embodiments of inventive compositions and methods are illustrated in the following examples. These examples are provided for illustrative purposes and are not considered limitations on the scope of inventive compositions and methods.

[00177] Examples

[00178] Example 1

[00179] This example describes a selective O-GlcNAc glycoprotein labeling platform (“GlycoID2”). A Novel mutants of a human GH84 enzyme, human O-GlcNAcase (hOGA*) were made and used in an O-GlcNAc labeling platform that expresses in human cells. The hOGA mutants were coupled to a “TurboID” protein. The hOGA*--TurboID fusion protein constructs bind O-GlcNAc on intracellular proteins for labeling by the TurboID domain. Proteomics comparisons show ca. 90% selectivity for known O-GlcNAc proteins during labeling, which is extremely high for specific sugar labeling experiments in live cells. The activity and selectivity of the inventive fusion proteins in living cells was surprising and unexpected.

[00180] The hOGA*--TurboID fusion protein constructs allow comparison of O-GlcNAc changes between the two opposing signaling systems, such as insulin vs. glucagon over time. Further, these fusion protein constructs can be targeted to specific places in the cell, including the nucleus, cytosol, mitochondria, or plasma membrane. Therefore, these fusion proteins can be used in methods for spatiotemporal studies of O-GlcNAc effects during cellular events such as, but not limited to, disease processes, any type of signaling, nutrient fluctuation, and stress. Advantageously, the compositions and methods according to aspects of the present disclosure can be used in vivo.

[00181] Construct design: The following constructs were made and construction included codon optimization, DNA synthesis, and cloning. Plasmids were amplified and sequence verified to characterize the material.

[00182] hOGA*-6X-mTurbo-flag, D174N mutant:

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAGEDNPAGAGGA AVAGAAGGAR
RFLCGVVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMYSVE
EAEQLMTLISAAREYEIEFIYAISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDI
DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNVSQSPYLRT
VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY
KGRSTELIPRLKGVLTNPNCEFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
IQIKLENEGSDIEDIETDVL YSPQMALKLALTEWLQEF GVP HQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATTVVTTVYQEPIMSQGAALSGEPTTLTKEEEKKQPDEEPM DMVVEK
QEETHKNDNQILSEIVEAKMAEELKPM DTDKESIAESKSPMSMQEDCISDIAPMQTD
EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVSVNCKGKDSEKIEE WRSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
DIKSIMSMVKS FVQWLGC RSHSSAQFLIGDQEPWAFRGG LAGEFQRLLPIDGANDLFFQ
PPGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSMIPLLNAKQILGQLDGGSVAVLPVV
DSTNQYLLDRIGELKSGDACIAEYQQAGRGS RGRKWFSPFGANLYLSMFWR LKRGPA
IGLGPVIGIVMAEALRKL GADKVRVKWPNDLYLQDRKLAGILVELAGITGDAAQIVIGA
GINVAMRRVEESV VNQGWITLQEAGINLDRNTLAAMLIRELRAALELFEQEGLAPYLSR
WEKLDNFINRPVKLIIGDKEIFGISRGIDKQGALLLEQDGV IKPWMGGEISLRS AEKGGG
GSDYKDDDDK (SEQ ID NO:14)

[00183] mTurbo-6x-hOGA*-flag

MIPLLNAKQILGQLDGGSVAVLPVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGS RGR
KWFSPFGANLYLSMFWR LKRGPA AIGLGPVIGIVMAEALRKL GADKVRVKWPNDLYL
QDRKLAGILVELAGITGDAAQIVIGAGIN VAMRRVEESV VNQGWITLQEAGINLDRNTL
AAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIFGISRGIDKQGALL
LEQDGV IKPWMGGEISLRS AEKGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSMVQKE
SQATLEERESELSSNPAASAGASLEPPAAPAGEDNPAGAGGA AVAGAAGGARRFLCG
VVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMYSVEEAEQL
MTLISAAREYEIEFIYAISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDIDHNM
CAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNVSQSPYLRTVGEK
LLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPYKGRST
ELIPRLKGVLTNPNCEFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVSIQIKLE
NEGSDIEDIETDVL YSPQMALKLALTEWLQEF GVP HQYSSRQVAHSGAKASVVDGTPLV
AAPSLNATTVVTTVYQEPIMSQGAALSGEPTTLTKEEEKKQPDEEPM DMVVEKQEETH
HKNDNQILSEIVEAKMAEELKPM DTDKESIAESKSPMSMQEDCISDIAPMQTDEQTNK

EQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRANSSVVS
VNCKGKDSEKIEEWRSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVWDIKSI
MSMVKSFVQWLGCRRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQPPGG
GGSDYKDDDDK (SEQ ID NO:15)

[00184] hOGA*-6X-mTurbo-flag, D174A mutant:

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAPGEDNPAGAGGA AVAGAAGGAR
RFLCGVVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMF WREMYSVE
EAEQLMTLISAAREYEIEFIY AISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFADI
DHNMC AADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNV SQSPYLRT
VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY
KGRSTELIPRLKGVLTNP NCEFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
IQIKLENEGSDEDIETDVL YSPQMALKLALTEWLQEFGVPHQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATT VTTVYQEPIMSQGAALSGEPTTLTKEEEKKQPDEEPM DMVVEK
QEETD HKNDNQILSEIVEAKMAEELKPM DTDKESIAESKSPMSMQEDCISDIAPMQTD
EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVSVVNCKGKDSEKIEEWRSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
DIKSIMSMVKS FVQWLGCRRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ
PPGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSMIPLLNAKQILGQLDGGSVAVLPVV
DSTNQYLLDRIGELKSGDACIAEYQQAGRGSRGRKWFSPFGANLYLSMFWRLKRG PAA
IGLGPVIGIVMAEALRKL GADKVRVKWPNDLYLQDRKLAGILVELAGITGDAAQIVIGA
GINVAMRRVEESV VNQG WITLQEAGINLDRNTLAAMLIRELRAALELFEQEGLAPYLSR
WEKLDNF INRPVKLIIGDKEIFGISRGIDKQGALLLEQDGVIPKPMGGEISLRS AEKGGG
GSDYKDDDDK (SEQ ID NO:16)

[00185] mCherry-TurboID (control construct)

MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPPYEGTQTAKLKVTKGGP
LPFAWDILSPQFMYGSKAYVKHPADIPDY LKLSFPEGFKWERVMNFEDGGVVTVTQDS
SLQDGEFIYKVKLRGTNFP SDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLK
DGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGM
DELYKGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSMIPLLNAKQILGQLDGGSVAVL
PVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGSRGRKWFSPFGANLYLSMFWRLKRG
PAAIGLGPVIGIVMAEALRKL GADKVRVKWPNDLYLQDRKLAGILVELAGITGDAAQIV
IGAGINVAMRRVEESV VNQG WITLQEAGINLDRNTLAAMLIRELRAALELFEQEGLAPY
LSRWEKLDNF INRPVKLIIGDKEIFGISRGIDKQGALLLEQDGVIPKPMGGEISLRS AEKGG
GGGSDYKDDDDK (SEQ ID NO:17)

[00186] cpOGA*-6X-mTurbo-flag:

MVGPKTGEENQVLVPNLNPTPENLEVVG DGFKITSSINLVGEE EADENAVNALREFLTA
NNIEINSENDPNSTTLIIGE VDDDIPELDEALNGTTAENLKEEGYALVSNDGKIAIEGKDG

DGTFYGVQTFKQLVKESNIPEVNITDYPTVSARGIVEGFYGTPWTHKDRLDQIKFYGEN
KLNTYIYAPKDDPYHREKWREYPENEMQRMQELIDASAENKVDFVFGISPGIDIRFDG
EAGEEDFNHLIAKAESLYDMGVRSAFYWDNIQDKSAAKHAQVLNRFNEEFVKAKGD
VKPLITVPTEYDTGAMVSNGQPRTYTRIFAETVDPSIEVMWTGPGVVTNEIPLSDAQLIS
GIYNRNMAVWWNYPVTDYFKGKLALGPMHGLDKGLNQYVDFFTVNPMEHAELSKISI
HTAADYSWNMDNYDYDKAWNRAIDMLYGDLAEDMKVFANHSTRMDNKTWAKSGR
EDAPELRKMDDELWNKLSSKEDASALIEELYGEFARMEEACNNLKANLPEVALEECSR
QLDELITLAQGDKASLDMIVAQLNEDTEAYESAKEIAQNKLNTALSSFAVISEKVAQSFI
QEALSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSMIPLLNQILGQLDGGSVAVL
PVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRGRKWFSPFGANLYLSMFWRLKRG
PAAIGLGPVIGIVMAEALRKLGADKVRVKWPNDLYLQDRKLAGILVELAGITGDAAQIV
IGAGINVAMRRVEESVVNQGWITLQEAGINLDRNTLAAMLIRELRAALELFEQGLAPY
LSRWEKLDNFNRPVKLIIGDKEIFGISRGIDKQGALLLEQDGVIPKPMGGEISLRSAEKG
GGGSDYKDDDDK (SEQ ID NO:18)

[00187] miniTurbo-GafD

MIPLLNQILGQLDGGSVAVLPVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRGR
KWFSPFGANLYLSMFWRLKRGPAAGLGPVIGIVMAEALRKLGADKVRVKWPNDLYL
QDRKLAGILVELAGITGDAAQIVIGAGINVAMRRVEESVVNQGWITLQEAGINLDRNTL
AAMLIRELRAALELFEQGLAPYLSRWEKLDNFNRPVKLIIGDKEIFGISRGIDKQGALL
LEQDGVIPKPMGGEISLRSAEKGGGGSGGGGSGGGGSMASVFIGSTENDVGPSQGSYS
STHAMDNLPFVYNTGYNIGYQNANVWRISGGFCVGLDGKVDLPVVGSLDGQSIYGLTE
EVGLLIWMGDTNYSRGTAMSGNSWENVFSGWCVGNYVSTQGLSVHVRPVILKRNSSA
QYSVQKTSIGSIRMRPYNGSSGGGGSDYKDDDDK (SEQ ID NO:26)

[00188] GafD-miniTurbo:

MAVFIGSTENDVGPSQGSYSSTHAMDNLPFVYNTGYNIGYQNANVWRISGGFCVGLD
GKVDLPVVGSLDGQSIYGLTEEVGLLIWMGDTNYSRGTAMSGNSWENVFSGWCVGNY
VSTQGLSVHVRPVILKRNSSAQYSVQKTSIGSIRMRPYNGSSGGGGSGGGGSGGGGSMI
PLLNQILGQLDGGSVAVLPVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRGRK
WFSPFGANLYLSMFWRLKRGPAAGLGPVIGIVMAEALRKLGADKVRVKWPNDLYLQ
DRKLAGILVELAGITGDAAQIVIGAGINVAMRRVEESVVNQGWITLQEAGINLDRNTLA
AMLIRELRAALELFEQGLAPYLSRWEKLDNFNRPVKLIIGDKEIFGISRGIDKQGALL
EQDGVIPKPMGGEISLRSAEKGGGGSDYKDDDDK (SEQ ID NO:27)

[00189] Mammalian Cell Culture and Transfection: Cells were obtained from ATCC. All consumables (pipette tips, glass Pasteur pipettes, Eppendorf tubes) were sterilized via autoclave. HeLa cells were cultured in DMEM (Sigma Aldrich, D6429) supplemented with 10% (v/v) HyClone Fetal Bovine Serum (Cytiva, SH30396.03) and 1% HyClone Penicillin-Streptomycin solution (Cytiva, SNUC-TUBOID0010) at 37 °C under 5% CO₂. All stable cell lines are supplemented with puromycin. All mammalian cell manipulations were done inside a laminar flow hood sterilized with 70% ethanol. To seed cells, the cells were carefully washed with sterile 7 mL of PBS pH 7.4 (1X) (ThermoFisher, 10010-023). The 1.5 mL of Trypsin 0.25% (1X) solution (Cytiva, SNUC-TUBOID0031.01) was added to the flask and incubated for 5 minutes at 37 °C under 5% CO₂. The trypsin was then neutralized with serum-containing growth media, where the trypsin can be further removed via centrifugation (300 x g for 3 minutes) in a sterile 15 mL centrifuge tube (FisherScientific, 14-955-237). The cell pellet was washed with PBS and recentrifuged. The cells can then be seeded into the desired flask (2.1x10⁶ cells for T-75 flask [USAScientific, CC7682-4875], 2.2x10⁶ cells for 100 mm dish [FisherScientific, FB0875713], 0.3x10⁶ cells for 6-well dish [FisherScientific, 07-200-83], 0.1 x10⁶ cells for 12-well dish [Corning, 3512]). According to the manufacturer's protocol, cells were transfected using TransIT-LT1 transfection reagent (Mirus Bio LLC, MIR 2304) for transient transfection. The transfected cells were incubated for 48-72 hours before use. Reagent amounts shown in Table 1.

Table 1: Transfection amounts

Culture Vessel	24-well	12-well	6-well	100 mm
Surface Area	1.9 cm ²	3.8 cm ²	9.6 cm ²	59 cm ²
Complete Growth Media	0.5 mL	1.0 mL	2.5 mL	15.5 mL
Serum-free Media	50 µL	100 µL	250 µL	1.5 mL
DNA (1 µg/µL)	0.5 µL	1 µL	2.5 µL	15 µL
<i>TransIT</i> -LT1 Reagent	1.5 µL	3 µL	7.5 µL	45 µL

[00190] Mammalian Cell Stable Selection: A kill curve is generated by splitting a confluent plate into varying concentrations of antibiotic and incubate cells for around 2 weeks, replacing the media as needed while monitoring cell death. 0, 50, 100, 200, 300, 400, 500, 600,

700, 800, 900, and 1000 ug/ml G418 is added to duplicate wells of cells in complete growth media, replacing the antibiotic every 2-3 days. The optimal concentration kills 100% of cells in one week. The cells are transfected with the desired plasmid according to the aforementioned protocol (6-well). One well is left untransfected for control, and transfected cells are allowed to incubate at least 48 hours before selection. After adding the antibiotics at optimal concentration, all untransfected cells are allowed to die (usually past nine days), and selected cells were brought to confluence so they can be banked and frozen.

[00191] Western Blot: To analyze cell lysates via immunoblot, cells were collected with a cell scraper from plates in RIPA buffer containing protease inhibitors (150 mM NaCl, 1% Nonidet P-40, 0.5% Na-deoxycholate, 0.1% sodium dodecyl sulfate [SDS (Sodium Dodecyl Sulfate)], and 50 mM Tris-pH 7.4). Cell lysates were briefly sonicated and centrifuged (12,000 x g for 10 minutes at 4 0C) to collect the soluble protein fraction. Protein concentration was determined via Pierce Rapid Gold BCA Protein Assay Kit (ThermoFisher, A53225). Samples were boiled in an SDS gel loading buffer for 5 minutes. Proteins were separated on a 4-12% gradient gel (NuPAGE™ 4 to 12%, Bis-Tris, 1.0–1.5 mm, Mini Protein Gels; ThermoFisher, NP0321BOX) and transferred to a nitrocellulose membrane (iBlot™ 2 Transfer Stacks, nitrocellulose, mini; ThermoFisher, IB23002) using an iBlot 2 dry blotting system (ThermoFisher, IB21001). After blocking with 5% w/v bovine serum albumin (Research Products International, A30075) in TBST buffer (10 mM Tris-pH8, 150 mM NaCl, 0.05% Tween 20) for 1 hour, the membrane was incubated with the appropriate antibody following the manufacturer's protocol. The signals from the antibodies were detected via the iBright™ FL1500 instrument (ThermoFisher, A44115). The membranes were incubated with Cy5- or HRP-conjugated streptavidin to detect biotinylated proteins. Antibodies used are supplied in Table 2.

Table 2: Antibody list

Target	Conjugate	Host	Supplier	Dilution
Biotin	HRP	Goat	CST – 7075S	1:1000-1:3000
Flag	N/A	Mouse	Sigma F3165	1:1000
Mouse IgG	HRP	Goat	Thermo A28175	1:10000-1:200000
OGT	N/A	Rabbit	CST 24083S	1:1000
Rabbit IgG	HRP	Donkey	Thermo 31458	1:10000-1:200000

Biotin	Cy5	-	Thermo SA1011	-	1:1000
HA	N/A	Rabbit	CST- C29F4		1:1000
V5	N/A	Rabbit	CST	-	1:1000
			D3H8		
			Q		

[00192] Biotin Labeling with Fusion Constructs: For biotin labeling experiments of stably transfected cells, biotin was added after induction of insulin/glucagon or the addition of inhibitors. Biotin (Carbosynth, 58-85-5) was diluted to 100 μ M (or desired concentration) in complete growth media and added directly to cells. The cells were incubated at 37 °C for the desired amount of time. For western blots and proteomics, labeling was stopped by washing with cold PBS and freezing at -80 °C.

[00193] Insulin/Glucagon Induction to track GlycoID2 activity in cellular signaling: For insulin induction of desired cells, the cells are serum-starved overnight (using media lacking FBS and then only using media with dialyzed FBS to reduce background labeling). Once ready, insulin is directly diluted into the cell plate's well for the desired concentration (see below table for examples) and mixed gently by swirling.

Insulin	10000	μ g/mL
Stock		

Insulin	0.5	1	5	10	μ g/mL	Final Concentration
6 well	0.1	0.2	1	2	μ L insulin	2mL DMEM
d100	0.5	1	5	10	μ L insulin	10mL DMEM

Glucagon treatment was done the same as insulin. After induction, biotin labeling can be done.

[00194] Sample Preparation for Proteomics: Cells were cultured in d-100 mm TC-treated Petri dishes. All cells were transiently expressing the desired construct. All cells were labeled with 100 μ M biotin using the methods mentioned above. Labeling was stopped by washing with cold PBS and freezing at -80 °C. The cells were detached from the plate via scraper with lysis

buffer (150 mM NaCl, 0.5 mM tris, 1% NP40, 0.1% SDS) and collected in Eppendorf tubes. The cells were lysed via passage through a needle (at least 10 passes) or sonication and clarified with centrifugation at 10,000 x g for 10 minutes at 4 °C. 100 µL of streptavidin-coated magnetic beads (NEB S1410S) were washed twice with RIPA buffer and then incubated with clarified lysates (400 µg protein) with rotation at 4 °C overnight to enrich biotinylated proteins. The magnetic beads were washed once with 500 µL RIPA buffer, once with 500 µL wash buffer (50 mM Tris, pH 7.4, 2% SDS), and twice with 500 µL RIPA buffer. Magnetic beads were resuspended in 500 µL 10 mM DTT (Dithiothreitol, GoldBio 27565-41-9) in PBS at 37 °C for 30 minutes, then cooled to room temperature. The supernatant was discarded. The magnetic beads were then resuspended in 1 mL 30 mM iodoacetamide (Sigma, 16125) (protected from light) at room temperature for 30 minutes. The supernatant was discarded, and the beads were washed with pure mass-spec grade water. The magnetic beads were resuspended in 300 µL 50% MeCN/50% water (Fisher Chemical, 75-05-8; Fisher Chemical, 7732-18-5) (m.s.-grade). The proteins were then digested with Lys-C protease (Thermo Scientific, 90051) with a 1:100 ratio Lys-C to protein sample (~0.3 µL for 50 µL resin) at 37 °C for 16 hours without shaking. The proteins were further digested with SOLu-Trypsin (Sigma, EMS0004) at a ratio of 1:20 trypsin weight to sample weight (50 µL resin, 3 µL Trypsin) at 47 °C for one hour, then cooled to 37 °C for four hours with rotation. The digestion was quenched by bringing the mixture to a final concentration of 1% formic acid (Thermo Scientific, 85178). The beads were removed from the mixture via magnet (or centrifugation) and washed twice with 200 µL 50% MeCN and once with M.S.-water. Bead fragments were removed with centrifugation (10,000 x g for 10 minutes). Samples were concentrated via a speed vac set to 40 °C, and the residues were stored at -80 °C. Following the manufacturer's protocol, peptide concentrations were determined via Pierce™ Quantitative Fluorometric Peptide Assay kit (ThermoFisher, 23290). The detection of detergents was conducted using an SDS assay, using Stains-all dye. A stock solution of 1.8 mM stains-all was made using 50% propanol: water (protect from light) (e.g., 10 mL solution needs 10 mg of stains-all). A 90 µM working solution was diluted from the stock solution in 5% formamide (OmniPur, 75-12-7) (e.g., for 5 mL; mix 0.25 mL stock, 0.25 mL formamide, 4.5 mL water, 2.5% propanol final). This solution can be stored at room temperature for ~four days in the dark. Pipette 1 µL of sample and 1 µL of a standard curve SDS (Fisher Scientific, BP166-500) sample (0.02-0.1%) into a 96-well plate. The standard curve started at 0.02% with increments of 0.01% (e.g., 0.02, 0.03, 0.04, etc.). 200 µL of the working solution was added to each well with a sample or standard (keep the plate protected from the light). The plate was read using a plate

reader at 445 nm. The samples should have minimal SDS to prevent damage to the mass spectrum column.

[00195] MaxQuant/Perseus proteomics data analysis: After ms/ms, raw files are analyzed using MaxQuant and Perceus, which provides a list of proteins found in samples. That protein list is compared to the known O-GlcNAcome database created by the Olivier-Van Stichelen Lab. (Wulff-Fuentes, E., Berendt, R.R., Massman, L. *et al.* The human O-GlcNAcome database and meta-analysis. *Sci Data* 8, 25 (2021). <https://doi.org/10.1038/s41597-021-00810-4>)

[00196] Expression (Flag) and Labeling activity (Cy5-strep) blots. GAPDH for loading control.

[00197] To test different constructs, the following were generated:

1. mCherry-miniTurbo as a nonspecific labeling control that does not bind sugars
2. hOGA*-miniTurbo is the optimal construct, comprising human O-GlcNAcase residues 1-706 with a D174N mutation
3. The reverse version: miniTurbo-hOGA* did not express well / poor labeling
4. hOGA*-A-miniTurbo is an alternate mutation, D174A – poor expression/labeling
5. GafD-miniTurbo = the bacterial lectin version – no expression/labeling
6. miniTurbo-GafD = the reverse version; no labeling/expression
7. CpOGA-miniTurbo = a previously reported GH84 enzyme; poor expression
8. V2 and V3-miniTurbo = reported/published GlycoID constructs

[00198] Improvement of the labeling specificity for O-GlcNAc by changing the targeting protein.

[00199] Western blotting was performed to demonstrate expression and activity of mCherry-miniTurbo hOGA*-miniTurbo, miniTurbo-hOGA*, hOGA*-A-miniTurbo, GafD-miniTurbo, miniTurbo-GafD, CpOGA-miniTurbo, and V2 and V3-miniTurbo constructs. For this, labeling was conducted at 100 μ M biotin with 6 hours of labeling or 500 μ M biotin with 30 minutes of labeling. mCherry was used as a fluorometric control, which provides strong expression and activity. CpOGA is a bacterial OGA with strong binding affinity for O-GlcNAcylated proteins. Mutated hOGA D147N (hOGA-miniTurbo/miniTurbo-hOGA), alanine mutated hOGA D147A (hOGA-A-miniTurbo), the codon optimized GafD constructs and CpOGA-miniTurbo were compared to one another. The longer labeling period (6 hours with 100 μ M biotin) provided stronger labeling when compared to the shorter time (30 minutes with 500 μ M biotin). The optimal variant was hOGA*-miniTurbo, which is the “GlycoID2” construct according to aspects of the present disclosure.

[00200] Optimization of the amount of time required to biotinylate proteins using the hOGA-miniTurbo was performed by incubation with 0 μ M 25 μ M, 50 μ M, 100 μ M, 250 μ M, or 500 μ M biotin for 10 minutes, 1 hour, or 4 hours. Labeling was found to be possible after 10 minutes, though it requires a high concentration of biotin, while after 4 hours the amount of labeling seems to plateau regardless of biotin concentration. Labeling in further experiments is typically conducted at 1 hour with 100 μ M or 500 μ M biotin.

[00201] After stable transfection of hOGA-miniTurbo-NES in U2OS cells, cells were induced with biotin for 1 hour of labeling. The enriched proteins were identified with proteomics. This list was checked against the reported “O-GlcNAcome Database” to check whether the enriched proteins (labeled by hOGA-miniTurbo constructs) were known O-GlcNAc proteins, see Figure 2. The majority of these GlycoID2 hits are known O-GlcNAc proteins, demonstrating high selectivity of this system.

[00202] Example 2

[00203] Materials and Methods

[00204] The following commercial cell lines, kits, assays, and equipment were used in this example.

[00205] Cell lines: U2OS cells (ATCC #HTB-96); HeLa cells (ATCC #CCL-2).

[00206] Tissue culture: Complete DMEM (Gibco); 0.25% Trypsin (Gibco); Cytiva Fetal Bovine Serum (FBS) (note: contains biotin; not suitable for labeling reactions) Cytiva HyClone Dialyzed FBS, heat-inactivated (biotin-free serum suitable for labeling; Gibco PBS pH 7.4 (1X), Mirus TransIT-LT1 transfection reagent. Cell lysis materials: 20-gauge needles with 1 mL syringes or a cell sonication device, Corning Costar 6-well Clear TC-Treated Well Plates (sterile).

[00207] Western blotting and immunofluorescence: Bolt LDS Sample Buffer (4X), NuPAGE 10% Bis-Tris (Thermo Scientific Mini Gel Tank and Power Supply), Nitrocellulose membrane (Invitrogen iBlot™ 2 Transfer Stacks), Goat anti-mouse IgG (H+L) Secondary Antibody, DYKDDDDK (SEQ ID NO: 25) Tag Mouse Monoclonal Antibodies*, Western blot transfer device (Invitrogen™ iBlot™ 2 Gel Transfer Device), Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488, Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 555*, Nuclear Stain (#R37605), SuperSignal™ West Pico PLUS Chemiluminescent Substrate, , Mild Stripping Buffer, 4% paraformaldehyde, 8-well Chamber Slide (LabTek), Pierce High Sensitivity Streptavidin HRP, Cell Microscope / Imager (Biotek/Agilent Cytation C1), Mitochondria Isolation Kit (#89874).

[00208] Proteomics and functional studies: HPLC-grade acetonitrile, HPLC-grade water, Streptavidin Magnetic Beads (New England Biolabs #S1420S), insulin (10 mg/mL in 25 mM HEPES, sterile), Glucagon (0.05 M acetic acid, 500 µg/mL, filtered sterilized), Dithiothreitol (DTT), Formic acid, Stains-All Solution for SDS Detection, Pierce Fluorometric Kit (Thermo #23290), Petri Dish (100 mm, Sterile), Rotor (End-over-end).

[00209] Construct design and synthesis

[00210] Plasmid constructs encoding the following protein sequences were synthesized by Twist Biosciences. The GlycoID2 plasmid encoded a glycan binding domain (modified human O-GlcNAcase 1-706 with D174N mutation), a miniTurboID labeling domain, and a FLAG epitope tag for detection with short flexible linker repeats (GGGGS) as spacing units. The mCherry-miniTurbo control construct encoded an mCherry red fluorescent protein and the miniTurboID labeling domain.

[00211] GlycoID2 [hOGA(1-706, D174N)-linker-miniTurboID-FLAG tag]

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAGEDNPAGAGGA AVAGAAGGAR
RFLCGVVEGFYGRP WVMEQRKELFRRLQK WELNTYLYAPKDDYKHRMFWREMYSVE
EAEQLMTLISAAREYEIEFIY AISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDI
DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNVSQSPYLRT
VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY
KGRSTELIPRLKGVLTNP NCEFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
IQIKLENEGSDIEDITDVL YSPQMALKLALTEWLQEF GVP HQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATTV VTTVYQEPIMSQGAALS GEPTTLTKEEEKKQPDEEPM DMVVEK
QEETD HKNDNQILSEIVEAKMAEELKPM DTDKESIAESKSP EMSMQEDCISDIAPMQTD
EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVVS VNCKGKDSEKIEE WRSRAAKFEEMCGLVMGMFTRLSN CANRTILYDMYSYVW
DIKSIMSMVKSFVQWL GCRSHSSAQFLIGDQEPWAFRGG LAGEFQRLLPIDGANDLFFQ
PPGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSMIPLLNAKQILGQLDGGSVAVLPVV
DSTNQYLLDRIGELKSGDACIAEYQQAGRGRGRKWFSPFGANLYLSMFWR LKRGPA
AIGLGPVIGIVMAEALRKL GADKVRVKWPNDLYLQDRKLAGILVELAGITGDAAQIVIGA
GINVAMRRVEESV VNQG WITLQEAGINLDRNTLAAMLIRELRAALELFEQEGLAPYLSR
WEKLDNF INRPVKLIIGDKEIFGISRGIDKQGALLLEQDGV IKPWMGGEISLRS AEKGGG
GSKLDYKDDDDK (SEQ ID NO: 28)

[00212] mCherry-miniTurboID-FLAG tag

MVSKGEEDNMAIIEKFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGP
LPFAWDILSPQFMYGSKAYVKHPADIPDY LKLSFPEGFKWERVMNFEDGGVVTVTQDS
SLQDGEFIYKVKLRGTNFP SDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLK
DGGHYDAEVKTTYKAKKPVQLPGA YNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGM
DELYKGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSMIPLLNAKQILGQLDGGSVAVL
PVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRGRKWFSPFGANLYLSMFWR LKRG
PAAIGLGPVIGIVMAEALRKL GADKVRVKWPNDLYLQDRKLAGILVELAGITGDAAQIV
IGAGINVAMRRVEESV VNQG WITLQEAGINLDRNTLAAMLIRELRAALELFEQEGLAPY
LSRWEKLDNF INRPVKLIIGDKEIFGISRGIDKQGALLLEQDGV IKPWMGGEISLRS AEKGG
GGGSKLDYKDDDDK (SEQ ID NO: 29)

[00213] Each GlycoID2 construct was subsequently mutated to insert a localization sequence (highlighted in bold) using the Q5 Site Directed Mutagenesis Kit (New England Biolabs, E0554).

[00214] **Cyt-GlycoID2**

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAPGEDNPAGAGGA AVAGAAGGAR
RFLCGVVEGFYGRP WVMEQRKELFRRLQK WELNTYLYAPKDDYKHRMF WREMY SVE
EAEQLMTLISAAREYEIEFIYAISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDI
DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPT EYCGTFCYPNVSQSPYLRT
VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY
KGRSTELIPRLKGVLTNP NCEFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
IQIKLENEGSDIEDIETDVL YSPQMALKLALTEWLQEF GVP HQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATTVVT TVYQEPIMSQGAALS GEPTTLTKEEEKKQPDEEPM DMVVEK
QEETHDKNDNQILSEIVEAKMAEELKPM DTDKESIAESKSP EMSMQEDCISDIAPMQTD
EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVSVNCKGKDSEKIEE WRSRAAKFEEMCGLVMGMFTRL SNCANRTILYDMYSYVW
DIKSIMSMVKSFVQWL GCRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ
PPGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGMIPLLNAKQILGQLDGGSVAVLPVV
DSTNQYLLDRIGELKSGDACIAEYQQAGRGRGRKWFSPFGANLYLSMFWR LKRGPA
IGLGPVIGIVMAEALRKL GADKVRVKWPNDLYLQDRKLAGILVELAGITGDAAQIVIGA
GINVAMRRVEESV VNQGWITLQEAGINLDRNTLAAMLIRELRAALELFEQEGLAPYLSR
WEKLDNFINRPVKLIIGDKEIFGISRGIDKQGALLLEQDGVIPWPMGGEISLRS AEKGGG
GSKLDYKDDDDKGGGGSLPPLERLTL (SEQ ID NO: 30)

[00215] **Mem-GlycoID2**

MGCINSKRKDDGGGGSGSMVQKESQATLEERESELSSNPAASAGASLEPPAAPAPGEDN
PAGAGGA AVAGAAGGAR RFLCGVVEGFYGRP WVMEQRKELFRRLQK WELNTYLYAP
KDDYKHRMF WREMY SVEEAEQLMTLISAAREYEIEFIYAISPGLDITFSNPKEVSTLKRK
LDQVSQFGCRSFALLFNDIDHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPT
EYCGTFCYPNVSQSPYLRTVGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIW
DNIHANDYDQKRLFLGPYKGRSTELIPRLKGVLTNP NCEFEANYVAIHTLATWYKSNM
NGVRKDVVMTDSEDSTVSIQIKLENEGSDIEDIETDVL YSPQMALKLALTEWLQEF GVP HQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATTVVT TVYQEPIMSQGAALS GEPTTLTKEEEKKQPDEEPM DMVVEK
QEETHDKNDNQILSEIVEAKMAEELKPM DTDKESIAESKSP EMSMQEDCISDIAPMQTD
EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVSVNCKGKDSEKIEE WRSRAAKFEEMCGLVMGMFTRL SNCANRTILYDMYSYVWDI
KSIMSMVKSFVQWL GCRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQPPKL
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGMIPLLNAKQILGQLDGGSVAVLP
VVDSTNQYLLDRIGELKSGDACIAEYQQAGRGRGRKWFSPFGANLYLSMFWR LKRGPA
AIGLGPVIGIVMAEALRKL GADKVRVKWPNDLYLQDRKLAGILVELAGITGDAAQIVIGA
GINVAMRRVEESV VNQGWITLQEAGINLDRNTLAAMLIRELRAALELFEQEGLAPYLSR
WEKLDNFINRPVKLIIGDKEIFGISRGIDKQGALLLEQDGVIPWPMGGEISLRS AEKEFG
GGGGSDYKDDDDK (SEQ ID NO: 31)

[00216] **Mit-GlycoID2**

MLGFVGRVAAAPASGALRRRLTPSASLPPAQLLLRAAPTAVHPVRDYAGGGGSGSM
VQKESQATLEERESELSSNPAASAGASLEPPAAPAPGEDNPAGAGGA AVAGAAGGAR

FLCGVVEGFYGRPWVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMYSVVE
 AEQLMTLISAAREYEIEFIYAISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDID
 HNMCAADKEVFSSFAHAQVSITNEIYQYLGEPEFTLFCPTEYCGTFCYPNVSQSPYLRTV
 GEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKIIRAPVIWDNIHANDYDQKRLFLGPYK
 GRSTELIPRLKGVLTNPNCFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVSI
 QIKLENEGSDIEDTDLVLYSPQMALKLALTEWLQEFVPHQYSSRQVAHSGAKASVVD
 GTPLVAAPSLNATTVVTTVYQEPIMSQGAALSGETTTLTKEEEKKQPDEEPMDMVVEK
 QEETDHKNDNQILSEIVEAKMAEELKPMDDTKESIAESKSPMSMQEDCISDIAPMQTD
 EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
 SSVVSVNCKGKDSEKIEEWSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
 DIKSIMSMVKSFFVQWLGCRRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ
 PPKLGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGMIPLLNKQILGQLDGGSVAVL
 PVVDSTNQYLLDRIGELKSGDACIAEYQQAGRGSRGRKWFSFGANLYLSMFWRKRG
 PAAIGLGPVIGIVMAEALRKLKADKVRVKWPNDLYLQDRKLAGILVELAGITGDAAQIV
 IGAGINVAMRRVEESVVNQGWITLQEAGINLDRNTLAAMLIRELRAALELFEQGLAPY
 LSRWEKLDNFNRPVKLIIGDKEIFGISRGIDKQGALLLEQDGVIPKPMGGEISLRS AEKE
 FGGGGS DYKDDDDK (SEQ ID NO: 32)

[00217] Nuc-GlycoID2

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAGEDNPAGAGGA AVAGAAGGAR
 RFLCGVVEGFYGRPWVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMYSVVE
 EAEQLMTLISAAREYEIEFIYAISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDI
 DHNMCAADKEVFSSFAHAQVSITNEIYQYLGEPEFTLFCPTEYCGTFCYPNVSQSPYLRT
 VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKIIRAPVIWDNIHANDYDQKRLFLGPY
 KGRSTELIPRLKGVLTNPNCFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
 IQIKLENEGSDIEDTDLVLYSPQMALKLALTEWLQEFVPHQYSSRQVAHSGAKASVVD
 GTPLVAAPSLNATTVVTTVYQEPIMSQGAALSGETTTLTKEEEKKQPDEEPMDMVVEK
 QEETDHKNDNQILSEIVEAKMAEELKPMDDTKESIAESKSPMSMQEDCISDIAPMQTD
 EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
 SSVVSVNCKGKDSEKIEEWSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
 DIKSIMSMVKSFFVQWLGCRRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ
 PPGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGMIPLLNKQILGQLDGGSVAVLPV
 DSTNQYLLDRIGELKSGDACIAEYQQAGRGSRGRKWFSFGANLYLSMFWRKRGPA
 IGLGPVIGIVMAEALRKLKADKVRVKWPNDLYLQDRKLAGILVELAGITGDAAQIVIGA
 GINVAMRRVEESVVNQGWITLQEAGINLDRNTLAAMLIRELRAALELFEQGLAPYLSR
 WEKLDNFNRPVKLIIGDKEIFGISRGIDKQGALLLEQDGVIPKPMGGEISLRS AEKGGG
 GSKLDYKDDDDKGGGGSDPKKKRKVDPKKRKVDPKKRKV (SEQ ID NO: 33)

[00218] All constructs were verified by DNA sequencing.

[00219] Cell Culture

[00220] Human cell lines were obtained from ATCC. All consumables (pipette tips, glass Pasteur pipettes, Eppendorf tubes) were sterilized via autoclave. U2OS cells were cultured in DMEM (Sigma Aldrich, D6429) supplemented with 10% (v/v) HyClone Fetal Bovine Serum (Cytiva, SH30396.03) and 1% HyClone Penicillin-Streptomycin solution (Cytiva, SNUC-TUBOID0010) at 37 °C under 5% CO₂. All mammalian cell manipulations were done inside a laminar flow hood sterilized with 70% ethanol.

[00221] Transfection & stable cell selection

[00222] Cells were seeded in the desired plate, flask, or dish (see below) before transfecting with GlycoID2 plasmid DNA. Before seeding, cells were carefully washed with 7 mL of sterile PBS pH 7.4 (ThermoFisher, 10010-023). The 1.5 mL of Trypsin 0.25% solution (Cytiva, SNUC-TUBOID0031.01) was added to the flask and incubated for 5 minutes at 37 °C under 5% CO₂. The trypsin was then neutralized with serum-containing growth media, where the trypsin can be further removed via centrifugation (300 x g for 3 minutes) in a sterile 15 mL centrifuge tube (ThermoFisher, 14-955-237). The cell pellet was washed with PBS and recentrifuged. The cells can then be seeded into the desired flask (2.1x10⁶ cells for T-75 flask (USA Scientific, CC7682-4875), 2.2x10⁶ cells for 100 mm dish (ThermoFisher, FB0875713), 0.3x10⁶ cells for 6-well dish (ThermoFisher, 07-200-83), or 0.1 x10⁶ cells for 12-well dish [Corning, 3512). Cells were transfected using TransIT-LT1 transfection reagent (Mirus Bio, MIR 2304) for transient transfection, following the manufacturer's protocol. Transfected cells were incubated for 48-72 hours before use.

[00223] Immunoblotting/streptavidin labeling

[00224] To analyze cell lysates via immunoblot, cells were collected with a cell scraper from plates in RIPA buffer containing protease inhibitors (150 mM NaCl, 1% Nonidet P-40, 0.5% Na-deoxycholate, 0.1% sodium dodecyl sulfate [SDS (Sodium Dodecyl Sulfate)], and 50 mM Tris-pH 7.4). The cell lysates were briefly sonicated and centrifuged (17,000 x g for 20 minutes at 4 °C) to collect the soluble protein fraction. Mitochondrial protein isolation was done via the ThermoFischer Mitochondria Isolation Kit (#89874). Protein concentration was determined via Pierce Rapid Gold BCA Protein Assay Kit (ThermoFisher, A53225). Samples were boiled at 75 °C in an SDS gel loading buffer for 10 minutes. Proteins were separated on a 10% gradient gel (NuPAGE™ 4 to 12%, Bis- 197 Tris, 1.0–1.5 mm, Mini Protein Gels; ThermoFisher, NP0321BOX) and transferred to a nitrocellulose membrane (iBlot™ 2 Transfer Stacks, nitrocellulose, mini; ThermoFisher, IB23002) using an iBlot 2 dry blotting system (ThermoFisher, IB21001). After blocking with 5% w/v bovine serum albumin (Research Products International, A30075) in 1×TBST buffer (10 mM Tris-pH8, 150 mM NaCl, 0.05% Tween 20) for 1 hour, the membrane was incubated overnight with the appropriate antibody following the manufacturer's protocol. The signals from the antibodies were detected via the iBright™ FL1500 instrument (ThermoFisher, A44115). The membranes were incubated with Cy5- or HRP-conjugated streptavidin to detect biotinylated proteins.

[00225] Biotin Labeling with Fusion Constructs

[00226] For biotin labeling experiments of stably transfected cells, biotin was added after induction of insulin/glucagon or the addition of inhibitors. Biotin (Carbosynth, 58-85-5) was diluted to 100 μ M (or desired concentration) in complete growth media and added directly to cells. The cells were incubated at 37 °C for the desired amount of time. For western blots and proteomics, labeling was stopped by washing with cold PBS and freezing at -80 °C.

[00227] Insulin/Glucagon Induction

[00228] For insulin induction of desired cells, the cells are serum-starved overnight before induction using DMEM media without fetal bovine serum (or reduced FBS as needed). To stabilize the cells, the cells should be cultured in media with dialyzed FBS before serum starvation. Once ready, insulin is directly diluted into the cell plate's well to the desired concentration and mixed gently while swirling.

[00229] Immunohistochemistry and Immunofluorescence

[00230] Immunofluorescence of localized constructs U2OS cells were seeded onto 8-well glass chamber slides (ThermoFisher #154534PK) before transfection with GlycoID2 constructs. After transfection and expression for 48 hours, the media was removed, and the cells were fixed with 4% paraformaldehyde and then washed three times with PBS (except when using the MitoView dye). Nuclear staining was accomplished using NucBlue (ThermoFisher), 2 drops of the dye per mL of media incubated for 20 minutes at 37 °C. Fluorescent images were captured using a fluorescence imager.

[00231] Activity assays with biotin / siRNA / inhibitor controls

[00232] For the OGT knockdown, cells were transfected with Dharmacon ON-TARGET plus SMART pool human OGT siRNA (# L-019111-00-0005), SMART pool human OGA siRNA (# L-012805-00-0005), or ON-TARGET control pool non-targeting pool siRNA (#D001810-10-05) as control using DharmaFECT, described by the manufacturer. The SMART pool consists of a combination of 4 different siRNA oligomers optimized for a knockdown in human cell lines. These were used instead of two distinct siRNA sequences for OGT knockdown and its corresponding SMART pool control knockdown. The dried siRNA pellets were recentrifuged and resuspended in RNase-free 1x siRNA buffer (60 mM KCl, 6 mM HEPES-pH 7.5, and 0.2 mM MgCl₂) to a final concentration of 20 μ M and aliquoted into 20 μ L samples. Plates were seeded to the desired confluency. For transfection, the siRNA aliquot was diluted to 5 μ M and transfected using the following table:

		Tube1:		Tube2:			
--	--	--------	--	--------	--	--	--

		Diluted siRNA		Diluted DharmaFEC T			
Plate format	Surface Area (cm ² /well)	Vol. of 5 μ M siRNA (μ L)	Serum- free media (μ L)	DharmaFEC T reagent (μ L)	Serum -free media (μ L)	Complete media (μ L/well)	Total transfection vol. (μ L/well)
96	0.3	0.5	9.5	0.05-0.5	9.95- 9.5	80	100
24	2	2.5	47.5	0.25-2.5	49.75- 47.5	400	500
12	4	5	95	0.5-5.0	99.5- 95.0	800	1000
6	10	10	190	1.0-10.0	199.0- 190.0	1600	2000

[00233] For 24-well plates, 2 μ L of DharmaFECT reagent was sufficient for this siRNA transfection. The reagents were gently mixed via pipetting and incubated for 5 minutes at room temperature. The tubes were then combined and incubated for an additional 20 minutes. The media was removed from the plate and replaced with the appropriate amount of transfection reagent. For protein analysis, the transfected plates were incubated at 37 °C under 5% CO₂ for 48-96 hours.

[00234] Proteomics sample prep and analysis

[00235] Cells were cultured in d-100 mm TC-treated Petri dishes. All cells were transiently expressing the desired construct. All the cells were labeled with 100 μ M biotin using the abovementioned methods. Labeling was stopped by washing with cold PBS and freezing at -80 °C. The cells were detached from the plate via scraper with lysis buffer (150 mM NaCl, 0.5 mM tris, 1% NP40, 0.1% SDS) and collected in passes) or sonication and clarified with centrifugation at 10,000 x g for 10 minutes at 4 °C.

[00236] After gentle shaking to suspend beads in solution, 100 μ L of streptavidin-coated magnetic beads (NEB S1410S) were washed twice with RIPA buffer and then incubated with clarified lysates (400 μ g protein) with rotation at 4 °C overnight to enrich biotinylated proteins.

The magnetic beads were washed once with 500 μ L RIPA buffer, once with 500 μ L wash buffer (50 mM Tris, pH 7.4, 2% SDS), and twice with 500 μ L RIPA buffer. Magnetic beads were resuspended in 500 μ L 10 mM DTT (Dithiothreitol, GoldBio 27565-41-9) in PBS at 37 °C for 30 minutes, then cooled to room temperature. The supernatant was discarded. The magnetic beads were then resuspended in 1 mL 30 mM iodoacetamide (Sigma, 16125) (protected from light) at room temperature for 30 minutes. The supernatant was discarded, and the beads were washed with pure mass-spec grade water. The magnetic beads were resuspended in 300 μ L 50% MeCN/50% water (Fisher Chemical, 75-05-8; Fisher Chemical, 7732-18-5) (m.s.-grade).

[00237] The proteins were then digested with Lys-C protease (Thermo Scientific, 90051) with a 1:100 ratio Lys-C to protein sample (~0.3 μ L for 50 μ L resin) at 37 °C for 16 hours without shaking. The proteins were further digested with SOLu-Trypsin (Sigma, EMS0004) at a ratio of 1:20 trypsin weight to sample weight (50 μ L resin, 3 μ L trypsin) at 47 °C for one hour, then cooled to 37 °C for four hours with rotation. The digestion was quenched by bringing the mixture to a final concentration of 1% formic acid (Thermo Scientific, 85178). The beads were removed from the mixture via a magnet (or centrifugation) and washed twice with 200 μ L 50% MeCN and once with M.S.-water. Bead fragments were removed with centrifugation (10,000 x g for 10 minutes). Samples were concentrated via a speed vac set to 40 °C, and the residues were stored at -80 °C.

[00238] Following the manufacturer's protocol, peptide concentrations were determined via Pierce™ Quantitative Fluorometric Peptide Assay kit (ThermoFisher, 23290). The detection of detergents was conducted using an SDS assay, using Stains-all dye. A stock solution of 1.8 mM stains-all was made using 50% propanol: water (protect from light) (e.g., 10 mL solution requires 10 mg of stains-all). A 90 μ M working solution was diluted from the stock solution in 5% formamide (OmniPur, 75-12-7) (e.g., for 5 mL, mix 0.25 mL stock, 0.25 mL formamide, 4.5 mL water, 2.5% propanol final). This solution can be stored at room temperature for ~four days in the dark. Pipette 1 μ L of sample and 1 μ L of a standard curve SDS (Fisher Scientific, BP166-500) sample (0.02-0.1%) into a 96-well plate. The standard curve started at 0.02% with increments of 0.01% (e.g., 0.02, 0.03, 0.04, etc.). 200 μ L of the working solution was added to each well with a sample or standard (keep the plate protected from the light). The plate was read using a plate reader at 445 nm. The samples should have minimal SDS to prevent damage to the mass spectrum column.

[00239] Data analysis & statistics

[00240] Statistical analysis was performed using Bioconductor R 3.3.2 and GraphPad Prism. Graphs were generated with GraphPad Prism. P values ≤ 0.05 are reported as significant and *P ≤ 0.05 , **P < 0.01 , ***P < 0.001 indicates level of significance. Linear regression or Student's t-test (two-sided) were applied when two conditions were compared, including qRT-PCR analysis. For analysis of 3 or more experiment conditions, a mixed-model approach was applied. For protein interactome studies, the STRING-db dataset was downloaded and used within Cytoscape via the stringApp plugin.

[00241] A system for highly selective O-GlcNAc labeling in the cytosol, mitochondria, nucleus, and intracellular plasma membrane are described herein according to aspects of the present disclosure, also called "GlycoID2" herein, is shown diagrammatically in **Figure 1C**. In this example, a catalytically inactive mutant of the human O-GlcNAcase (hOGA) was used as an O-GlcNAc binding agent. These GlycoID2 systems which had over 90% selectivity for directly labeling known O-GlcNAc proteins in live cells. Additionally, the extended subcellular location allowed complementary tracking of O-GlcNAc protein dynamics in distinct compartments. As a practical consideration, a puromycin selection marker was added and stable cell lines expressing GlycoID2 constructs were successfully generated. O-GlcNAc changes in response to two complementary hormones, insulin and glucagon were analyzed using a fusion protein which provides highly selective O-GlcNAc labeling in the cytosol, mitochondria, nucleus, and intracellular plasma membrane according to aspects of the present disclosure. Comparative proteomics obtained using compositions and method of the present disclosure revealed differential insulin- and glucagon-induced O-GlcNAcylation patterns between short 1 hour, medium 2-4 hour, and extended 16-hour time points. O-GlcNAc modifications regulate proteins in many different signaling pathways, nutrient sensing events, and stress. GlycoID2 tools have numerous uses, such as, spatiotemporal studies of O-GlcNAcylated cellular proteins in a variety of cell processes.

[00242] **Results**

[00243] **Constructs and Activity**

[00244] Monomeric O-GlcNAc binding proteins were used in this example to simplify their assembly and utility in live cells, including GafD lectin, and bacterial O-GlcNAcase from *Clostridium perfringens* (CpOGA).³¹ Because glycosylhydrolase enzymes require catalytic acid residues, CpOGA can be mutated with a catalytically inactivating D298N mutation to act as a potent O-GlcNAc binding protein, referred to as CpOGA*.³² Variants of CpOGA* are useful as

an O-GlcNAc binding protein in protein lysates, and are used in applications including far western blot imaging,³² affinity enrichment,³³ and proximity ligation assays.³⁴

[00245] The human OGA (hOGA) with a mutation D174N lacks the enzyme activity.³⁵ hOGA contains three domains, an N-terminal catalytic domain, a stalk domain that plays a role in enzyme dimerization, and a C-terminal pseudo-histone acetyl transferase (HAT-like) domain of unknown function. In this example (hOGA) with enzyme inactivation mutation(s) were also shortened by deletion of the C-terminal 213 amino acid residues. The “hOGA*” construct therefore contained a D174N mutation and spanned residues 1-706 of hOGA, which is normally a 916 residue protein.³⁸

[00246] In this example a “miniTurboID proximity labeling domain”, i.e. a mutant *E. coli* biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein, was attached to the mutant hOGA protein, generating a fusion protein of the present disclosure.

[00247] A 25-residue flexible linker (GGGGS₅) was disposed between the O-GlcNAc binding domain (i.e. hOGA* or cpOGA*) of the fusion protein and the miniTurboID proximity labeling domain. Figure 3 diagrammatically shows selected fusion protein constructs including linkers. As a non-glycan targeting control, the red fluorescent protein mCherry-linker-miniTurboID construct was generated. For hOGA*, the miniTurboID proximity labeling domain was attached to either the N- or C-terminus. Both mutation D174A and D174N variants were tested to remove the key catalytic residue of hOGA. CpOGA* was used with miniTurboID attached to the a C-terminal. GafD GlycoID constructs (V2G and V3G) as well as non-targeted cytosolic and nuclear constructs (V2 and V3) were used.

[00248] The expression and labeling activity of each construct was tested to determine which would be the most active for proximity labeling of O-GlcNAc. Expression blots were produced using antibodies for Flag, V5, or HA tags for new constructs, cyt-GlycoID, and nuc-GlycoID, respectively. It was observed that hOGA*-miniTurboID was expressed at higher levels than its reversed miniTurbo-hOGA*. Also, hOGA(D174N) expressed at higher levels than hOGA(D174A) mutation variants. The cpOGA* variant did not express at high levels. GlycoID constructs V3G (nuc-GlycoID) and V2G (cyt-GlycoID) had lower expression levels than hOGA(D174N)-miniTurbo. The control construct mCherry-miniTurbo, expressed well and showed high activity. In particular, the expression levels of mCherry-miniTurbo and hOGA*-miniTurbo were similar, such that mCherry-miniTurbo can be used as a control for targeted vs. untargeted O-GlcNAc labeling studies.

[00249] As an initial activity test, 100 μ M biotin was administered and the labelling reaction was allowed to proceed for 6 hours. The biotinylation activity between all of these constructs was compared to the GafD constructs. It was observed that the hOGA* had improved protein labeling with biotin over the former GlycoID constructs V2G and V3G. The strongest labeling was observed with the non-sugar targeted control mCherry-miniTurbo. Because mCherry-miniTurboID was not designed to bind or target specific subsets of protein, this is indicative of nonspecific binding. Strong labeling was observed using hOGA(D174N)-miniTurbo. Strep-HRP: streptavidin-horse radish peroxidase. V2 = cyt-miniTurboID, V2G = cyt-GafD-miniTurboID, V3 = nuc-miniTurboID, and V3G = cyt-GafD-miniTurboID. The optimized hOGA(D174N)-linker-miniTurboID construct was called GlycoID2 for the purposes of this study.

[00250] To characterize the labeling activity of GlycoID2 fusion proteins, a short biotin concentration/time optimization study was used. Following 48 hours expression with a plasmid encoding GlycoID2 (hOGA*-miniTurboID), incubation was performed with 0 μ M, 25 μ M, 50 μ M, 100 μ M, 250 μ M, or 500 μ M biotin for 10 minutes, 1 hour, 4 hours, 6 hours, or overnight before quenching via media removal and freezing at -80°C . Flag-HRP was used as expression/loading control for GlycoID2 fusion protein, and biotin labeling was tracked using streptavidin-HRP. It was found that biotinylation activity can be observed as soon as 10 minutes, following induction with high concentrations of biotin above 250 μ M. Concentration-dependent labeling was observed at short 1 hour time points or longer 4 hour time points, up to 500 μ M biotin. At longer time points, increased biotin concentration did not make a difference in 6 hour or overnight labeling even at 25 μ M biotin.

[00251] Various cellular growth media and controls for O-GlcNAc levels were tested. Glucose levels are known to correlate with O-GlcNAc levels in several tissues. For this experiment, HeLa cells were cultured in high glucose (4.5 g/L) or low glucose (1.0 g/L) media during the expression of GlycoID2 constructs. After 48 hours, labeling was initiated with 100 μ M biotin for 6 hours. Bars indicate standard error of the mean (SEM), N = 3 per condition. mT = miniTurboID. Performing GlycoID2 vs. mCherry-miniTurbo labeling experiments in high and low glucose conditions did not make an O-GlcNAc-dependent difference in labeling, see **Figure 4**.

[00252] As a further control experiment, siRNA was used to knock down OGT and OGA to reduce or increase O-GlcNAc levels, respectively. GlycoID2 and siRNA for OGT, OGA, or a scrambled negative control (Scr) were transfected simultaneously. After 48 hours, biotin

labeling was induced. Lower panels are Coomassie brilliant blue gel stains, as an additional loading control. After 48 hours, OGT knockdown was apparent, and biotinylation levels were slightly reduced. On the other hand, 48 hours after OGA knockdown did not show strong knockdown by OGA western blot. The siRNAs targeting endogenous OGA are not expected to knock-down GlycoID2 sequences because the DNA/RNA sequences are different between the endogenous and the engineered OGA sequences.

[00253] Location-targeted GlycoID2 variants

[00254] The distribution of OGT between the cytosol, mitochondria, nucleus, and the plasma membrane, such as during insulin signaling, is thought to be a major determinant of O-GlcNAc regulation. To study the four primary locations where OGT is active within cells, the cytosol, mitochondria, nucleus, and the plasma membrane, protein localization signals were used to target GlycoID2 constructs to these subcellular compartments.

[00255] Site-directed mutagenesis was used to assay targeting signal peptides and orientation (N- vs. C-terminus) for GlycoID2 targeting sequences. These mutagenesis methods were successful for the cytosol, plasma membrane, and nucleus, and the resulting constructs used are called cyt-GlycoID2, mem-GlycoID2, and nuc-GlycoID2, respectively herein. Briefly, a C-terminal Rev-associated nuclear export signal (NES) was used for the cyt-GlycoID2, a C-terminal triple SV40 nuclear localization signal (NLS) was used for nuc-GlycoID2, and an N-terminal Lyn kinase plasma membrane signal (PLS) was used for mem-GlycoID2. Mitochondria targeting sequences (MTS) are typically long signal peptides that are cleaved following mitochondrial import, and indeed mutagenesis was unable to install a MTS sequence. Mitochondria targeting sequences ATP-synthase, cytochrome c oxidase subunit 8 (COX8), superoxide dismutase 2 (SOD2), and a mitochondrial aldolase, were used. The human ATP-synthase MTS was used for the mit-GlycoID2 construct. Immunofluorescence was used to confirm the intended subcellular localization for each construct and revealed the expected location for all of cyt-GlycoID2, mem-GlycoID2, mit-GlycoID2, and nuc-GlycoID2,

[00256] Following confirmation of GlycoID2 construct localization, a puromycin selection marker was included in expression construct plasmids to obtain stable pools of cells expressing cyt-GlycoID2, nuc-GlycoID2, mem-GlycoID2, or mit-GlycoID2.

[00257] U2OS cells were used for establishment of one set of cell lines since U2OS cells natively respond to signaling hormones like insulin and glucagon. First, cyt-GlycoID2 (hOGA*-based O-GlcNAc binding) was compared with cyt-GlycoID (GafD-based O-GlcNAc binding).

[00258] Labeling was performed for 1 hour or 4 hours with 0 μ M, 25 μ M, 50 μ M, 100 μ M, 250 μ M, or 500 μ M biotin concentration. GlycoID2 expression was determined with anti-Flag blot and cyt-GlycoID expression was determined with anti-V5 tag blot. Coomassie gel was used as loading control. Immunoblotting was then performed to compare the activity of stabilized cyt-GlycoID2 with the cyt-GlycoID constructs in U2OS cells.

[00259] The labeling pattern looked nearly identical between the two constructs, with exceptions being prominent bands at the molecular weight of each construct (50 vs. 110 kDa, respectively), which showed autobiotinylation of the GlycoID2 or GlycoID construct itself. Notably, in this study with stabilized GlycoID and GlycoID2 proteins, relatively strong background labeling was observed even in lanes not induced with biotin. Though the K_m for miniTurboID is in the micromolar range and efficient labeling conditions typically require at least 100 μ M biotin addition, it was reasoned that in the stabilized cell lines low levels of constant biotinylation might build up on proteins. Biotin is not a component in Dulbecco's Modified Eagle's Medium (DMEM) that was used for U2OS cell culture but is present in micromolar levels in the fetal bovine serum (FBS) that was used to supplement cell culture media. Switching to dialyzed FBS, ameliorated the background labeling issue after just 24 hours, and 48 hours of culture in biotin-free media eliminated nearly all background bands.

[00260] Proteomic O-GlcNAc analysis with targeted GlycoID2 constructs

[00261] The selectivity of GlycoID2 labeling for O-GlcNAcylated proteins was tested in cells by mass spectrometry-based proteomics analysis. Briefly, cells expressing a GlycoID2 expression construct were treated in the condition of interest such as insulin stimulation (or baseline culture conditions as control) and protein proximity labeling was induced with biotin. After the desired labeling time, cells were harvested, biotinylated proteins were enriched by streptavidin binding. An on-bead digestion protocol was used with LysC and trypsin proteases before peptides were quantified and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). For the initial test, U2OS cells stably expressing cyt-GlycoID2 under puromycin selection were used. Cells were labeled in unstimulated, baseline cell culture conditions with 500 μ M biotin and labeling proceeded for 1 h at 37 °C in normal tissue culture conditions (5% CO₂ atmosphere, standard high glucose DMEM media). To assess variability between sampling, two biological replicate studies were conducted with 4 technical replicates each. After processing, the protein IDs present in at least 2/4 technical replicates were mapped between the two biological replicates, see **Figure 5A**. Samples were acquired under identical conditions approximately 4 months apart. 80% and 67% overlap between the two samples,

respectively, was observed indicating moderate to good reproducibility for cyt-GlycoID2 labeling in stable expression cell lines. N = 4 technical replicates per biological replicate.

[00262] The two replicate data sets identified 279 and 335 proteins with strong MS/MS intensities in at least 2 of 4 technical replicates per experiment, respectively. The labeled protein IDs with the human O-GlcNAcome were determined via comparison with the database reported by Wulff-Fuentes *et al*³⁹ and the O-GlcNAcATLAS reported by Ma *et al*⁴⁰. The overlap with known O-GlcNAcylated proteins was 97% and 94% between the two biological replicates, indicating excellent selectivity for human O-GlcNAc protein labeling. **Figure 5B** shows overlap with the O-GlcNAcome database (accessed February, 2023). Unreported indicates protein IDs were not present in the O-GlcNAcome database as of February, 2023. The primary subcellular locations of the hits from the cyt-GlycoID2 biological replicates was analyzed using the workflow and database reported by Yan *et al*⁴¹, see **Figure 5C**. The majority were located in the cytosol, which matches the immunofluorescence location of this construct. The major overlap with cytosolic O-GlcNAc proteins indicates that cyt-GlycoID2 labeled the intended targets between two independent trials.

[00263] Spatiotemporal analysis of O-GlcNAc dynamics during insulin and glucagon signaling

[00264] Homeostasis is regulated in multicellular organisms through a complex network of hormones and other cellular regulators, which act via receptors to initiate signaling. Cell context, including receptor expression levels, concentration-dependent distance from hormone production sites, and epigenetics add additional layers of regulation. Two reciprocal hormones for glucose metabolism and uptake vs. secretion are insulin and glucagon. The role of insulin-driven O-GlcNAc regulation is well-studied, but glucagon signaling activities on O-GlcNAcylation patterns is less conclusive. Without wishing to be bound by theory, it is believed that insulin and glucagon stimulation would lead to differential protein O-GlcNAcylation. Targeted GlycoID2 constructs according to aspects of the present disclosure would reveal such differences.

[00265] Having established that GlycoID2 constructs could selectively label O-GlcNAcylated proteins in as short a time as 1 hour, a 1 hour time point was used for a comparative labeling study. Serum-starved U2OS cells stably expressing cyt-GlycoID2 were used, and signaling pathways were induced in these cells by treating with 1 µg/mL insulin or 1 µg/mL glucagon. O-GlcNAcylated proteins in the cytosol were tracked over time following 1 hour, 2 hours, 4 hours, or 16 hours of stimulation and different proteins were analyzed. Biotin

labeling time was held constant at 1 hour for each insulin/glucagon time point, that is, biotin was added 1 hour before each sample was harvested. Proteins were analyzed for significant difference from control conditions that were not stimulated with hormone but were induced with biotin 1 hour before harvesting. An analysis of all time points showed less than 20% overlap between the O-GlcNAc proteins induced by insulin compared with the O-GlcNAc proteins induced by glucagon, see **Figure 6A**). Moreover, overlap of proteins between the 4 time points within the insulin or the glucagon datasets also revealed highly varied responses at the 1 hour, 2 hours, 4 hours, or 16 hours times, see **Figure 6B, 6C**). Akt-Serine473 phosphorylation was used as a marker for insulin induction of O-GlcNAc proteins, see **Figure 6D**) and CREB phosphorylation as a marker of glucagon induction of O-GlcNAc proteins, see **Figure 6E**. Volcano plots for 3 insulin time points were generated demonstrating that O-GlcNAc patterns dynamically change between two complementary signaling systems in human cell lines.

[00266] These results show that fusion proteins of the present disclosure which include a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein were useful to detect O-GlcNAcylation patterns in targeted locations, including the cytosol, mitochondria, plasma membrane, and nucleus, and differential alteration of O-GlcNAcylation patterns by insulin or glucagon stimulation in mammalian cells expressing the fusion proteins. Overall insulin-induced O-GlcNAcylation was distinct from glucagon-induced O-GlcNAcylation, and both patterns changed dynamically at the 1 hour, 2 hour, 4 hour, and 16 hour time points. These data are consistent with the rapid sequential release of insulin during periods of nutrient excess, which can be secreted in pulses as rapid as 15-30 minutes.

[00267] References

- 1 Bar-Peled, L. & Kory, N. Principles and functions of metabolic compartmentalization. *Nat Metab* 4, 1232-1244, doi:10.1038/s42255-022-00645-2 (2022).
- 2 Lee, W. D., Mukha, D., Aizenshtein, E. & Shlomi, T. Spatial-fluxomics provides a subcellular-compartmentalized view of reductive glutamine metabolism in cancer cells. *Nature communications* 10, 1351, doi:10.1038/s41467-019-09352-1 (2019).
- 3 Campanella, M. E., Chu, H. & Low, P. S. Assembly and regulation of a glycolytic enzyme complex on the human erythrocyte membrane. *Proc Natl Acad Sci U S A* 102, 2402-2407, doi:10.1073/pnas.0409741102 (2005).

- 4 Wolfe, K. et al. Dynamic compartmentalization of purine nucleotide metabolic enzymes at leading edge in highly motile renal cell carcinoma. *Biochem Biophys Res Commun* 516, 50-56, doi:10.1016/j.bbrc.2019.05.190 (2019).
- 5 Orre, L. M. et al. SubCellBarCode: Proteome-wide Mapping of Protein Localization and Relocalization. *Mol Cell* 73, 166-182.e167, doi:10.1016/j.molcel.2018.11.035 (2019).
- 6 Ellisdon, Andrew M. & Halls, Michelle L. Compartmentalization of GPCR signalling controls unique cellular responses. *Biochemical Society Transactions* 44, 562-567, doi:10.1042/bst20150236 (2016).
- 7 Zaccolo, M. & Pozzan, T. Discrete microdomains with high concentration of cAMP in stimulated rat neonatal cardiac myocytes. *Science* 295, 1711-1715, doi:10.1126/science.1069982 (2002).
- 8 Baillie, G. S. Compartmentalized signalling: spatial regulation of cAMP by the action of compartmentalized phosphodiesterases. *The FEBS journal* 276, 1790-1799, doi:https://doi.org/10.1111/j.1742-4658.2009.06926.x (2009).
- 9 Doigneaux, C. et al. Hypoxia drives the assembly of the multienzyme purinosome complex. *J Biol Chem* 295, 9551-9566, doi:10.1074/jbc.RA119.012175 (2020).
- 10 Nelson, Z. M., Leonard, G. D. & Fehl, C. Tools for investigating O-GlcNAc in signaling and other fundamental biological pathways. *Journal of Biological Chemistry* 300, doi:10.1016/j.jbc.2023.105615 (2024).
- 11 Gonzalez-Rellan, M. J., Fondevila, M. F., Dieguez, C. & Nogueiras, R. O-GlcNAcylation: A Sweet Hub in the Regulation of Glucose Metabolism in Health and Disease. *Frontiers in Endocrinology* 13, doi:10.3389/fendo.2022.873513 (2022).
- 12 Fahie, K. M. M., Papanicolaou, K. N. & Zachara, N. E. Integration of O-GlcNAc into Stress Response Pathways. *Cells* 11, doi:10.3390/cells11213509 (2022).
- 13 Hart, G. W. Nutrient regulation of signaling and transcription. *J Biol Chem* 294, 2211-2231, doi:10.1074/jbc.AW119.003226 (2019).
- 14 Levine, Z. G. & Walker, S. The Biochemistry of O-GlcNAc Transferase: Which Functions Make It Essential in Mammalian Cells? *Annual review of biochemistry* 85, 631-657, doi:10.1146/annurev-biochem-060713-035344 (2016).
- 15 Levine, Z. G. et al. Mammalian cell proliferation requires noncatalytic functions of O-GlcNAc transferase. *Proceedings of the National Academy of Sciences* 118, e2016778118, doi:10.1073/pnas.2016778118 (2021).

- 16 Yang, X. et al. Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. *Nature* 451, 964-969, doi:10.1038/nature06668 (2008).
- 17 Perez-Cervera, Y. et al. Insulin signaling controls the expression of O-GlcNAc transferase and its interaction with lipid microdomains. *Faseb j* 27, 3478-3486, doi:10.1096/fj.12-217984 (2013).
- 18 Fehl, C. & Hanover, J. A. Tools, tactics and objectives to interrogate cellular roles of O-GlcNAc in disease. *Nature chemical biology* 18, 8-17, doi:10.1038/s41589-021-00903-6 (2022).
- 19 Samavarchi-Tehrani, P., Samson, R. & Gingras, A. C. Proximity Dependent Biotinylation: Key Enzymes and Adaptation to Proteomics Approaches. *Mol Cell Proteomics* 19, 757-773, doi:10.1074/mcp.R120.001941 (2020).
- 20 Kang, M.-G. & Rhee, H.-W. Molecular Spatiomics by Proximity Labeling. *Accounts of Chemical Research* 55, 1411-1422, doi:10.1021/acs.accounts.2c00061 (2022).
- 21 Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* 196, 801-810, doi:10.1083/jcb.201112098 (2012).
- 22 Geri, J. B. et al. Microenvironment mapping via Dexter energy transfer on immune cells. *Science* 367, 1091-1097, doi:10.1126/science.aay4106 (2020).
- 23 Liu, Y., Nelson, Z. M., Reda, A. & Fehl, C. Spatiotemporal proximity labeling tools to track GlcNAc sugar-modified functional protein hubs during cellular signaling. *ACS Chemical Biology*, doi:10.1021/acscchembio.2c00282 (2022).
- 24 Nelson, Z. M., Kadiri, O. & Fehl, C. GlycoID Proximity Labeling to Identify O-GlcNAcylated Protein Interactomes in Live Cells. *Curr Protoc* 4, e1052, doi:10.1002/cpz1.1052 (2024).
- 25 Hsu, K.-L., Gildersleeve, J. C. & Mahal, L. K. A simple strategy for the creation of a recombinant lectin microarray. *Molecular bioSystems* 4, 654-662, doi:10.1039/b800725j (2008).
- 26 Carrillo, L. D., Krishnamoorthy, L. & Mahal, L. K. A Cellular FRET-Based Sensor for β -O-GlcNAc, A Dynamic Carbohydrate Modification Involved in Signaling. *Journal of the American Chemical Society* 128, 14768-14769, doi:10.1021/ja065835+ (2006).
- 27 Carrillo, L. D., Froemming, J. A. & Mahal, L. K. Targeted in vivo O-GlcNAc sensors reveal discrete compartment-specific dynamics during signal transduction. *J Biol Chem* 286, 6650-6658, doi:10.1074/jbc.M110.191627 (2011).
- 28 Branon, T. C. et al. Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* 36, 880-887, doi:10.1038/nbt.4201 (2018).

- 29 Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* 47, D607-d613, doi:10.1093/nar/gky1131 (2019).
- 30 Ambrosi, M., Cameron, N. R. & Davis, B. G. Lectins: tools for the molecular understanding of the glycode. *Org Biomol Chem* 3, 1593-1608, doi:10.1039/b414350g (2005).
- 31 Macauley, M. S., Whitworth, G. E., Debowski, A. W., Chin, D. & Vocadlo, D. J. *O*-GlcNAcase Uses Substrate-assisted Catalysis: KINETIC ANALYSIS AND DEVELOPMENT OF HIGHLY SELECTIVE MECHANISM-INSPIRED INHIBITORS. *Journal of Biological Chemistry* 280, 25313-25322, doi:10.1074/jbc.M413819200 (2005).
- 32 Mariappa, D. et al. A mutant *O*-GlcNAcase as a probe to reveal global dynamics of protein *O*-GlcNAcylation during *Drosophila* embryonic development. *The Biochemical journal* 470, 255-262, doi:10.1042/bj20150610 (2015).
- 33 Selvan, N. et al. A mutant *O*-GlcNAcase enriches *Drosophila* developmental regulators. *Nature chemical biology* 13, 882-887, doi:10.1038/nchembio.2404 (2017).
- 34 Song, J. et al. *O*-GlcNAcylation Quantification of Certain Protein by the Proximity Ligation Assay and *Clostridium perfringens* OGAD298N (CpOGAD298N). *ACS Chemical Biology* 16, 1040-1049, doi:10.1021/acscchembio.1c00185 (2021).
- 35 Wang, P. et al. *O*-GlcNAc cycling mutants modulate proteotoxicity in *Caenorhabditis elegans* models of human neurodegenerative diseases. *Proc Natl Acad Sci U S A* 109, 17669-17674, doi:10.1073/pnas.1205748109 (2012).
- 36 Groves, J. A., Maduka, A. O., O'Meally, R. N., Cole, R. N. & Zachara, N. E. Fatty acid synthase inhibits the *O*-GlcNAcase during oxidative stress. *J Biol Chem* 292, 6493-6511, doi:10.1074/jbc.M116.760785 (2017).
- 37 Ge, Y. et al. Target protein deglycosylation in living cells by a nanobody-fused split *O*-GlcNAcase. *Nature chemical biology* 17, 593-600, doi:10.1038/s41589-021-00757-y (2021).
- 38 Elbatrawy, A. A., Kim, E. J. & Nam, G. *O*-GlcNAcase: Emerging Mechanism, Substrate Recognition and Small-Molecule Inhibitors. *ChemMedChem* 15, 1244-1257, doi:10.1002/cmdc.202000077 (2020).
- 39 Wulff-Fuentes, E. et al. The human *O*-GlcNAcome database and meta-analysis. *Scientific Data* 8, 25, doi:10.1038/s41597-021-00810-4 (2021).

- 40 Ma, J., Li, Y., Hou, C. & Wu, C. O-GlcNAcAtlas: A database of experimentally identified O-GlcNAc sites and proteins. *Glycobiology*, doi:10.1093/glycob/cwab003 (2021).
- 41 Yan, T. et al. Proximity-labeling chemoproteomics defines the subcellular cysteinome and inflammation-responsive mitochondrial redoxome. *Cell Chem Biol* 30, 811-827.e817, doi:10.1016/j.chembiol.2023.06.008 (2023).
- 42 Liu, S. et al. An organism-wide atlas of hormonal signaling based on the mouse lemur single-cell transcriptome. *Nature communications* 15, 2188, doi:10.1038/s41467-024-46070-9 (2024).
- 43 Lei, L. et al. Multilevel Differential Control of Hormone Gene Expression Programs by hnRNP L and LL in Pituitary Cells. *Mol Cell Biol* 38, doi:10.1128/mcb.00651-17 (2018).
- 44 Bansal, P. & Wang, Q. Insulin as a physiological modulator of glucagon secretion. *American Journal of Physiology-Endocrinology and Metabolism* 295, E751-E761, doi:10.1152/ajpendo.90295.2008 (2008).
- 45 Potter, S. C. et al. Dissecting OGT's TPR domain to identify determinants of cellular function. *Proc Natl Acad Sci U S A* 121, e2401729121, doi:10.1073/pnas.2401729121 (2024).
- 46 He, J. et al. Spatiotemporal Activation of Protein O-GlcNAcylation in Living Cells. *Journal of the American Chemical Society* 144, 4289-4293, doi:10.1021/jacs.1c11041 (2022).
- 47 Sacoman, J. L., Dagda, R. Y., Burnham-Marusich, A. R., Dagda, R. K. & Berninsone, P. M. Mitochondrial O-GlcNAc Transferase (mOGT) Regulates Mitochondrial Structure, Function, and Survival in HeLa Cells. *J Biol Chem* 292, 4499-4518, doi:10.1074/jbc.M116.726752 (2017).
- 48 Lang, D. A., Matthews, D. R., Peto, J. & Turner, R. C. Cyclic oscillations of basal plasma glucose and insulin concentrations in human beings. *N Engl J Med* 301, 1023-1027, doi:10.1056/nejm197911083011903 (1979).
- 49 Wareham, N. J., Phillips, D. I., Byrne, C. D. & Hales, C. N. The 30 minute insulin incremental response in an oral glucose tolerance test as a measure of insulin secretion. *Diabet Med* 12, 931, doi:10.1111/j.1464-5491.1995.tb00399.x (1995).
- 50 Rayaprolu, S. et al. Cell type-specific biotin labeling in vivo resolves regional neuronal and astrocyte proteomic differences in mouse brain. *Nature communications* 13, 2927, doi:10.1038/s41467-022-30623-x (2022).
- 51 Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angewandte Chemie International Edition* 44, 7342-7372, doi:10.1002/anie.200501023 (2005).

- 52 Fan, J., Krautkramer, K. A., Feldman, J. L. & Denu, J. M. Metabolic regulation of histone post-translational modifications. *ACS Chem Biol* 10, 95-108, doi:10.1021/cb500846u (2015).
- 53 Smith, K., Shen, F., Lee, H. J. & Chandrasekaran, S. Metabolic signatures of regulation by phosphorylation and acetylation. *iScience* 25, 103730, doi:https://doi.org/10.1016/j.isci.2021.103730 (2022).
- 54 Ceddia, R. B. Direct metabolic regulation in skeletal muscle and fat tissue by leptin: implications for glucose and fatty acids homeostasis. *International Journal of Obesity* 29, 1175-1183, doi:10.1038/sj.ijo.0803025 (2005).
- 55 Paneque, A., Fortus, H., Zheng, J., Werlen, G. & Jacinto, E. The Hexosamine Biosynthesis Pathway: Regulation and Function. *Genes (Basel)* 14, doi:10.3390/genes14040933 (2023).
- 56 Yang, X. & Qian, K. Protein O-GlcNAcylation: emerging mechanisms and functions. *Nature Reviews Molecular Cell Biology* 18, 452 (2017).
- 57 Ong, Q., Han, W. & Yang, X. O-GlcNAc as an Integrator of Signaling Pathways. *Front Endocrinol (Lausanne)* 9, 599, doi:10.3389/fendo.2018.00599 (2018).
- 58 Lin, W., Gao, L. & Chen, X. Protein-Specific Imaging of O-GlcNAcylation in Single Cells. *ChemBioChem* 16, 2571-2575, doi:https://doi.org/10.1002/cbic.201500544 (2015).
- 59 Kasprowicz, A. et al. Exploring the Potential of β -Catenin O-GlcNAcylation by Using Fluorescence-Based Engineering and Imaging. *Molecules* 25, 4501 (2020).
- 60 Hwang, B. B., Engel, L., Goueli, S. A. & Zegzouti, H. A homogeneous bioluminescent immunoassay to probe cellular signaling pathway regulation. *Commun Biol* 3, 8, doi:10.1038/s42003-019-0723-9 (2020).
- 61 Hall, M. P. et al. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem Biol* 7, 1848-1857, doi:10.1021/cb3002478 (2012).
- 62 Shaner, N. C. et al. A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nat Methods* 10, 407-409, doi:10.1038/nmeth.2413 (2013).
- 63 Lee, O.-H. et al. Genome-wide YFP Fluorescence Complementation Screen Identifies New Regulators for Telomere Signaling in Human Cells*. *Molecular & Cellular Proteomics* 10, S1-S11, doi:https://doi.org/10.1074/mcp.M110.001628 (2011).
- 64 Zhou, J., Lin, J., Zhou, C., Deng, X. & Xia, B. An improved bimolecular fluorescence complementation tool based on superfolder green fluorescent protein. *Acta Biochimica et Biophysica Sinica* 43, 239-244, doi:https://doi.org/10.1093/abbs/gmq128 (2011).

65 Lang, Y., Li, Z. & Li, H. Analysis of Protein-Protein Interactions by Split Luciferase Complementation Assay. *Curr Protoc Toxicol* 82, e90, doi:10.1002/cptx.90 (2019).

[00268] Item List

[00269] Item 1. A fusion protein comprising: a glycan binding component linked to a mutant *E. coli* biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant *E. coli* biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein.

[00270] Item 2. The fusion protein of item 1, wherein the glycan binding component is selected from the group consisting of: an enzyme, a lectin with the proviso that the lectin is not a GalD lectin, a collectin, a ficolin, a C-reactive protein, and a carbohydrate-binding domain of any thereof.

[00271] Item 3. The fusion protein of item 1, wherein the glycan binding component is selected from the group consisting of: an aptamer, an antibody, and an antigen-binding fragment of an antibody.

[00272] Item 4. The fusion protein of item 1, item 2 or item 3, wherein the glycan binding component is a mutant O-GlcNAcase enzyme derived from a member of the GH84 family of glycosylhydrolases and lacking enzymatic glycosylhydrolase activity.

[00273] Item 5. The fusion protein of any one of items 1 to 4, wherein the mutant O-GlcNAcase enzyme derived from a member of the GH84 family of glycosylhydrolases and lacking enzymatic glycosylhydrolase activity is a mutant human O-GlcNAcase (hOGA) enzyme, a glycan specific binding fragment thereof, or a variant of either thereof.

[00274] Item 6. The fusion protein of item 5, wherein the mutant human O-GlcNAcase (hOGA) enzyme is a D174N mutant human O-GlcNAcase (hOGA) enzyme which includes the amino acid sequence:

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAPGEDNPAGAGGA AVAGAAGGAR
RFLCGVVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMYSVE
EAEQLMTLISAAREYEIEFIYAISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDI
DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNVSQSPYLRT
VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY
KGRSTELIPRLKGVLTNP NCFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
IQIKLENEGSDEDIETDVL YSPQMALKLALTEWLQEF GVP HQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATT VTTVYQEPIMSQGAALS GEPTTLTKEEEKKQPDEEPM DMVVEK
QEETD HKNDNQILSEIVEAKMAEELKPM DTDKESIAESKSP EMSMQEDCISDIAPMQTD

EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVSVNCKGKDSEKIEEWSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
DIKSIMSMVKSFVQWLGCRRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ
PP (SEQ ID NO:4)

or a variant thereof.

[00275] Item 7. The fusion protein of any one of items 1 to 6, wherein the glycan binding component has a C-terminus and an N-terminus, the mutant E. coli biotin ligase BirA has a C-terminus and an N-terminus, and the C-terminus of the glycan binding component is linked to the N-terminus of the mutant E. coli biotin ligase BirA or the N-terminus of the glycan binding component is linked to the C-terminus of the mutant E. coli biotin ligase BirA.

[00276] Item 8. The fusion protein of any one of items 1 to 7, wherein the glycan binding component is linked to the mutant E. coli biotin ligase BirA by a linker disposed between the glycan binding component and the mutant E. coli biotin ligase BirA.

[00277] Item 9. The fusion protein of any one of items 1 to 8, further comprising a localization signal peptide.

[00278] Item 10. The fusion protein of any one of items 1 to 9, further comprising a localization signal peptide capable of promoting localization of the fusion protein to a subcellular compartment selected from the group consisting of: nucleus, cytosol, mitochondria, endoplasmic reticulum, and plasma membrane.

[00279] Item 11. The fusion protein of any one of items 1 to 10, further comprising an exogenous detectable tag.

[00280] Item 12. A method of detecting proteins proximal to a target protein, comprising:
contacting a living cell with the fusion protein according to any one of items 1 to 11 under compatible biological conditions, whereby the fusion protein specifically binds to a glycosylation post-translational modification of a target protein of the cell;
providing biotin to the living cell, whereby the mutant E. coli biotin ligase BirA ligates biotin to proteins proximal to the target protein; and
detecting the biotinylated proteins, thereby detecting proteins proximal to the target protein.

[00281] Item 13. The method of item 12, wherein detecting the biotinylated proteins comprises purifying the biotinylated proteins and detecting the purified biotinylated proteins.

[00282] Item 14. The method of item 13, wherein detecting the purified biotinylated proteins comprises mass spectrometry.

[00283] Item 15. The method of item 13 or item 14, wherein detecting the purified biotinylated proteins comprises chromatography.

[00284] Item 16. The method of item 15, wherein the chromatography comprises gel electrophoresis.

[00285] Item 17. The method of item 16, wherein the chromatography comprises gel electrophoresis and transfer of the electrophoresed purified biotinylated proteins to a membrane.

[00286] Item 18. The method according to any one of items 12 to 17, wherein contacting the living cell with the fusion protein comprises introducing an expression construct encoding the fusion protein into the cell.

[00287] Item 19. An expression construct comprising a nucleic acid encoding the fusion protein according to any one of items 1 to 11.

[00288] Item 20. A cell comprising the expression construct of item 19.

[00289] Item 21. A composition substantially as shown or described herein.

[00290] Item 22. A method of detecting proteins proximal to a target protein substantially as shown or described herein.

[00291] Item 23. The fusion protein of any one of items 1 to 11 wherein the fusion protein includes: 1) SEQ ID NO: 3 or a variant thereof and SEQ ID NO: 2 or a variant thereof; 2) SEQ ID NO: 3 or a variant thereof and SEQ ID NO: 11 or a variant thereof; 3) SEQ ID NO: 34 or a variant thereof and SEQ ID NO: 2 or a variant thereof; or SEQ ID NO: 4 or a variant thereof and SEQ ID NO: 11 or a variant thereof.

[00292] Any patents or publications mentioned in this specification are incorporated herein by reference to the same extent as if each individual publication is specifically and individually indicated to be incorporated by reference.

[00293] The compositions and methods described herein are presently representative of preferred embodiments, exemplary, and not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art. Such changes and other uses can be made without departing from the scope of the invention as set forth in the claims.

CLAIMS

1. A fusion protein comprising: a glycan binding component linked to a mutant E. coli biotin ligase BirA, the glycan binding component capable of specific binding to a glycosylation post-translational modification of a target protein and the mutant E. coli biotin ligase BirA having enzymatic activity to ligate biotin to proteins proximal to the target protein.

2. The fusion protein of claim 1, wherein the glycan binding component is selected from the group consisting of: an enzyme, a lectin with the proviso that the lectin is not a GafD lectin, a collectin, a ficolin, a C-reactive protein, and a carbohydrate-binding domain of any thereof.

3. The fusion protein of claim 1, wherein the glycan binding component is selected from the group consisting of: an aptamer, an antibody, and an antigen-binding fragment of an antibody.

4. The fusion protein of claim 1, 2 or 3, wherein the glycan binding component is a mutant O-GlcNAcase enzyme derived from a member of the GH84 family of glycosylhydrolases and lacking enzymatic glycosylhydrolase activity.

5. The fusion protein of any one of claims 1 to 4, wherein the mutant O-GlcNAcase enzyme derived from a member of the GH84 family of glycosylhydrolases and lacking enzymatic glycosylhydrolase activity is a mutant human O-GlcNAcase (hOGA) enzyme, a glycan specific binding fragment thereof, or a variant of either thereof.

6. The fusion protein of claim 5, wherein the mutant human O-GlcNAcase (hOGA) enzyme is a D174N mutant human O-GlcNAcase (hOGA) enzyme which includes the amino acid sequence:

MVQKESQATLEERESELSSNPAASAGASLEPPAAPAGEDNPAGAGGA AVAGAAGGAR
RFLCGVVEGFYGRP WVMEQRKELFRRLQKWELNTYLYAPKDDYKHRMFWREMYSVE
EAEQLMTLISAAREYEIEFIY AISPGLDITFSNPKEVSTLKRKLDQVSQFGCRSFALLFNDI
DHNMCAADKEVFSSFAHAQVSITNEIYQYLGE PETFLFCPTEYCGTFCYPNV SQSPYLRT
VGEKLLPGIEVLWTGPKVVSKEIPVESIEEVSKI IKRAPVIWDNIHANDYDQKRLFLGPY

KGRSTELIPRLKGVLTNPNCFEANYVAIHTLATWYKSNMNGVRKDVVMTDSEDSTVS
IQIKLENEGSDEDIETDVLVYSPQMALKLALTEWLQEFVPHQYSSRQVAHSGAKASVVD
GTPLVAAPSLNATTVVTTVYQEPIMSQGAALSGEPTTLTKEEEKKQPDEEPMDMVVEK
QEETHDKNDNQILSEIVEAKMAEELKPMDDTKESIAESKSPMSMQEDCISDIAPMQTD
EQTNKEQFVPGPNEKPLYTAEPVTLEDLQLLADLFYLPYEHGPKGAQMLREFQWLRAN
SSVSVNCKGKDSEKIEEWSRAAKFEEMCGLVMGMFTRLSNCANRTILYDMYSYVW
DIKSIMSMVKSFVQWLGCRRSHSSAQFLIGDQEPWAFRGGLAGEFQRLLPIDGANDLFFQ
PP (SEQ ID NO:4)

or a variant thereof.

7. The fusion protein of any one of claims 1 to 6, wherein the glycan binding component has a C-terminus and an N-terminus, the mutant E. coli biotin ligase BirA has a C-terminus and an N-terminus, and the C-terminus of the glycan binding component is linked to the N-terminus of the mutant E. coli biotin ligase BirA or the N-terminus of the glycan binding component is linked to the C-terminus of the mutant E. coli biotin ligase BirA.

8. The fusion protein of any one of claims 1 to 7, wherein the glycan binding component is linked to the mutant E. coli biotin ligase BirA by a linker disposed between the glycan binding component and the mutant E. coli biotin ligase BirA.

9. The fusion protein of any one of claims 1 to 8, further comprising a localization signal peptide.

10. The fusion protein of any one of claims 1 to 9, further comprising a localization signal peptide capable of promoting localization of the fusion protein to a subcellular compartment selected from the group consisting of: nucleus, cytosol, mitochondria, endoplasmic reticulum, and plasma membrane.

11. The fusion protein of any one of claims 1 to 10, further comprising an exogenous detectable tag.

12. A method of detecting proteins proximal to a target protein, comprising:

contacting a living cell with the fusion protein according to any one of claims 1 to 11 under compatible biological conditions, whereby the fusion protein specifically binds to a glycosylation post-translational modification of a target protein of the cell;

providing biotin to the living cell, whereby the mutant E. coli biotin ligase BirA ligates biotin to proteins proximal to the target protein; and

detecting the biotinylated proteins, thereby detecting proteins proximal to the target protein.

13. The method of claim 12, wherein detecting the biotinylated proteins comprises purifying the biotinylated proteins and detecting the purified biotinylated proteins.

14. The method of claim 13, wherein detecting the purified biotinylated proteins comprises mass spectrometry.

15. The method of claim 13 or claim 14, wherein detecting the purified biotinylated proteins comprises chromatography.

16. The method of claim 15, wherein the chromatography comprises gel electrophoresis.

17. The method of claim 16, wherein the chromatography comprises gel electrophoresis and transfer of the electrophoresed purified biotinylated proteins to a membrane.

18. The method according to any one of claims 12 to 17, wherein contacting the living cell with the fusion protein comprises introducing an expression construct encoding the fusion protein into the cell.

19. An expression construct comprising a nucleic acid encoding the fusion protein according to any one of claims 1 to 11.

20. A cell comprising the expression construct of claim 19.

21. A composition substantially as shown or described herein.

22. A method of detecting proteins proximal to a target protein substantially as shown or described herein.

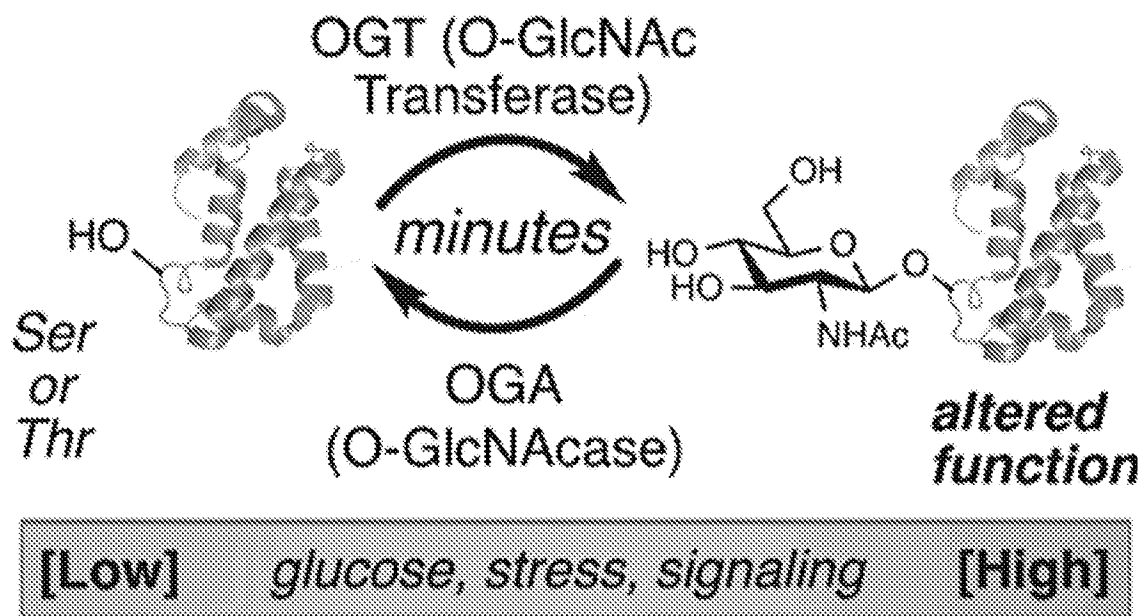


FIG. 1A

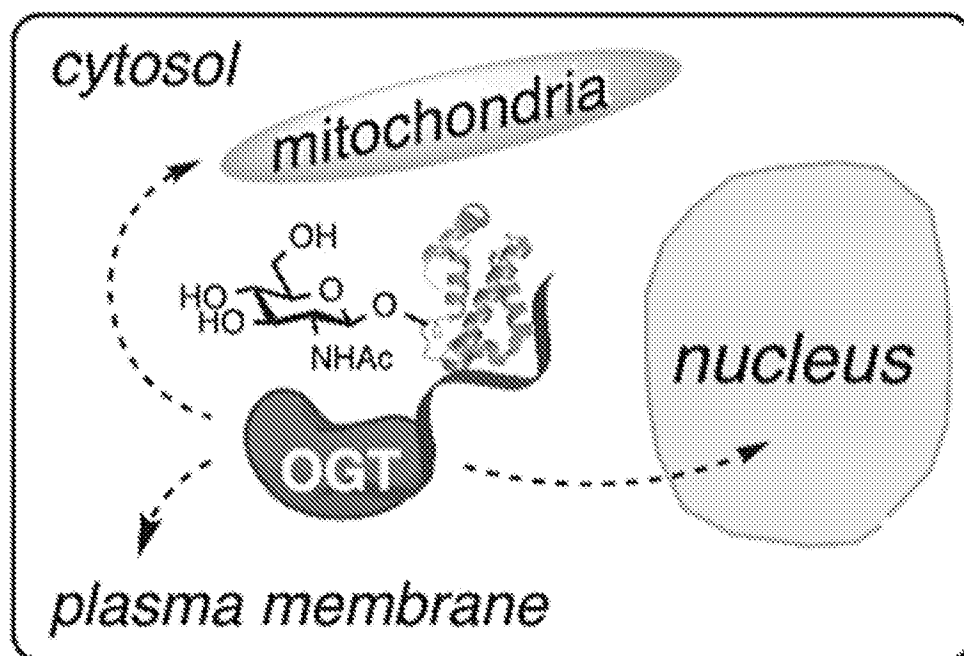


FIG. 1B

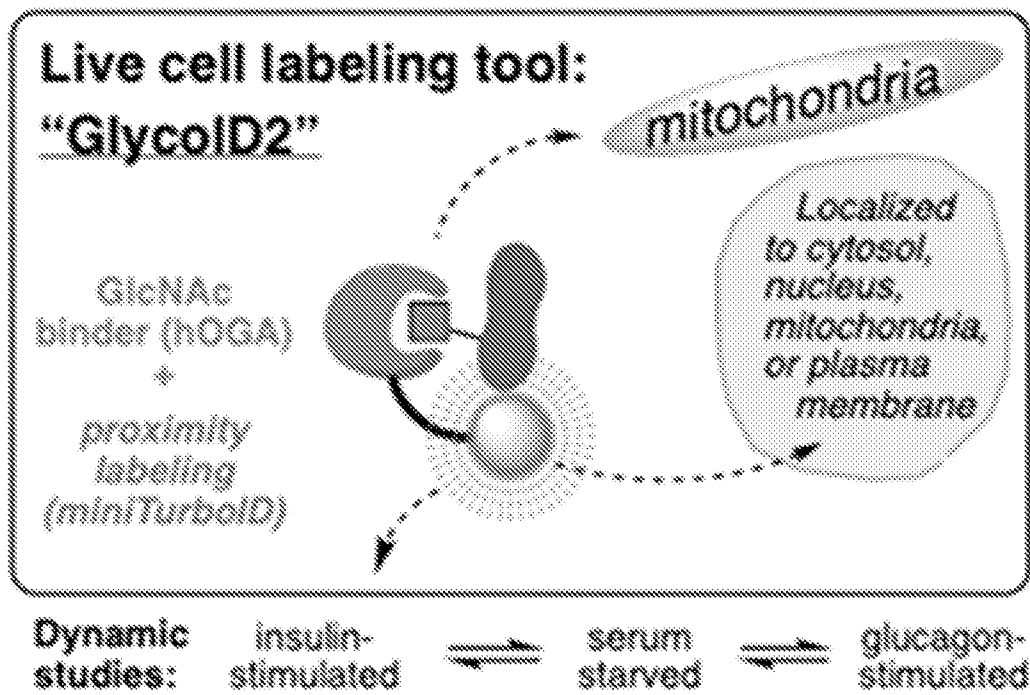


FIG. 1C

Stable Expression of hOGA-miniTurbo-NES -Percent of Enriched Proteins Known to be O-GlcNAcylated Full List

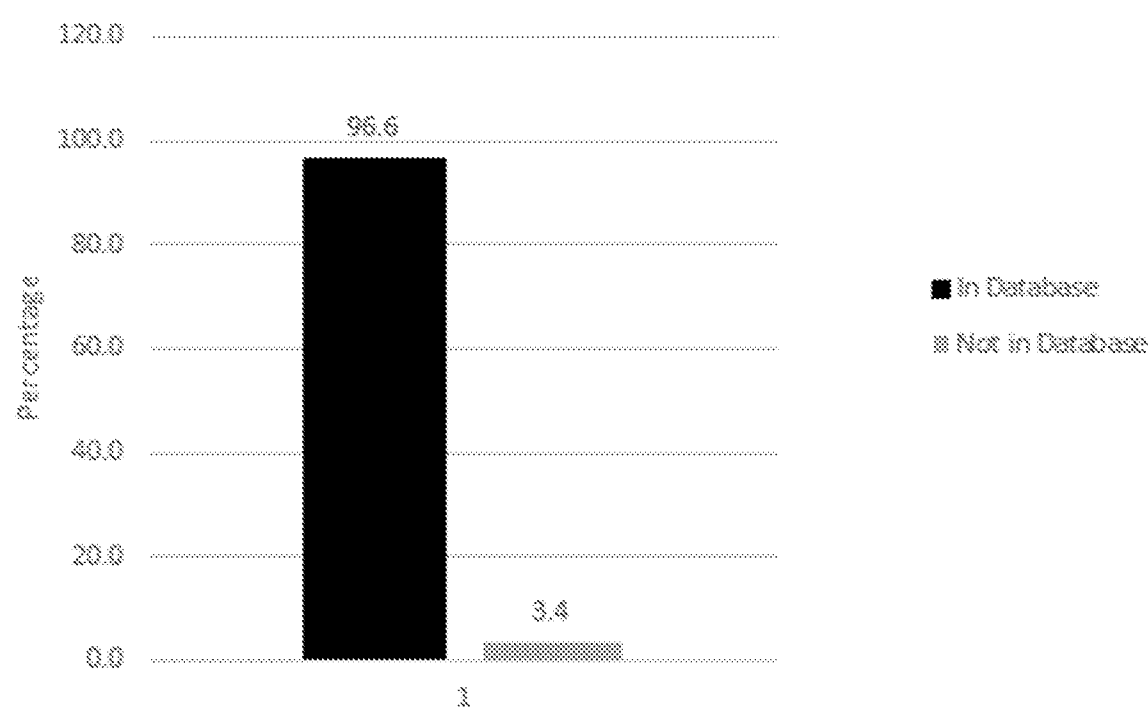


FIG. 2

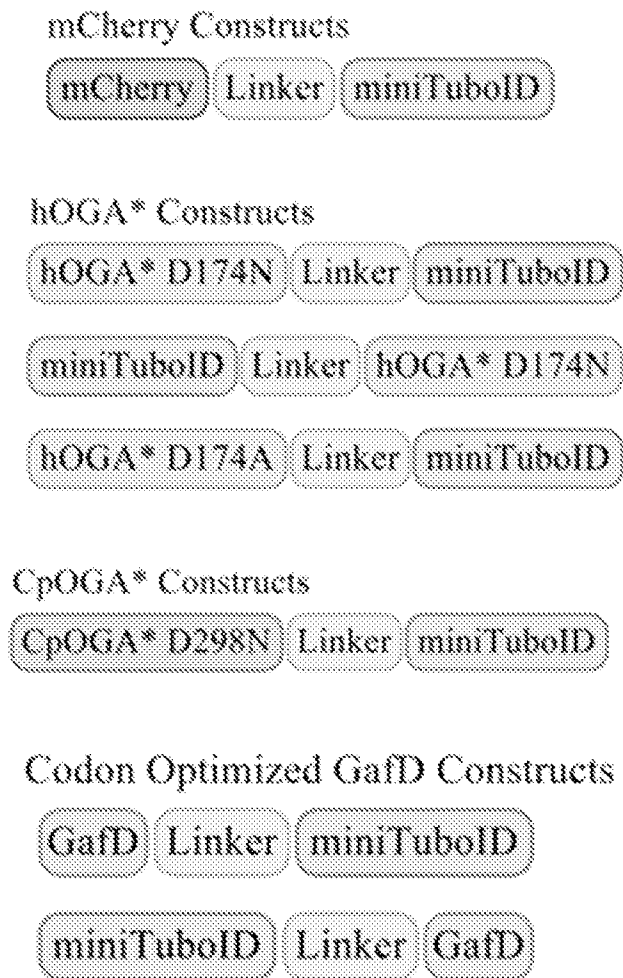


FIG. 3

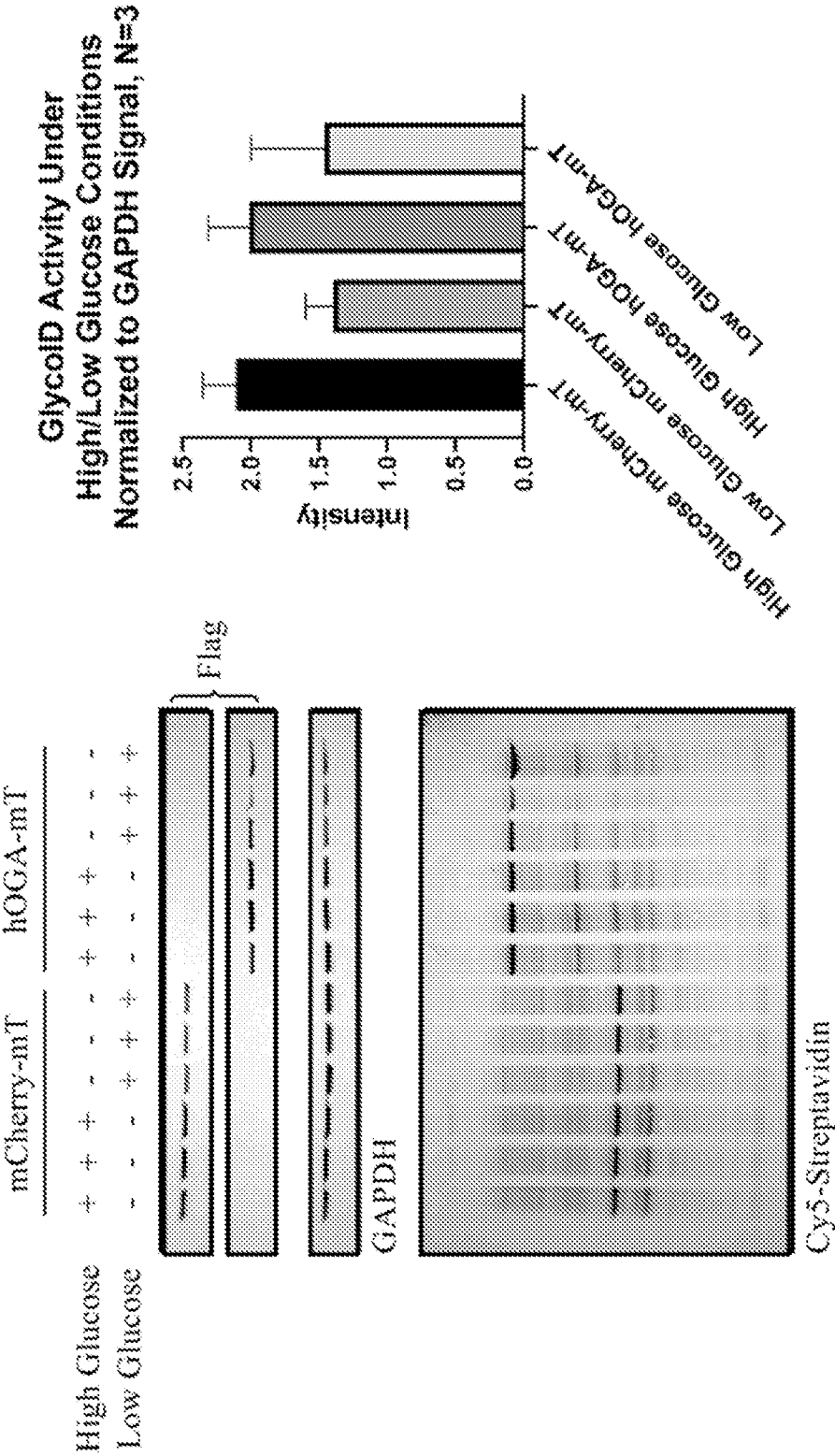


FIG. 4

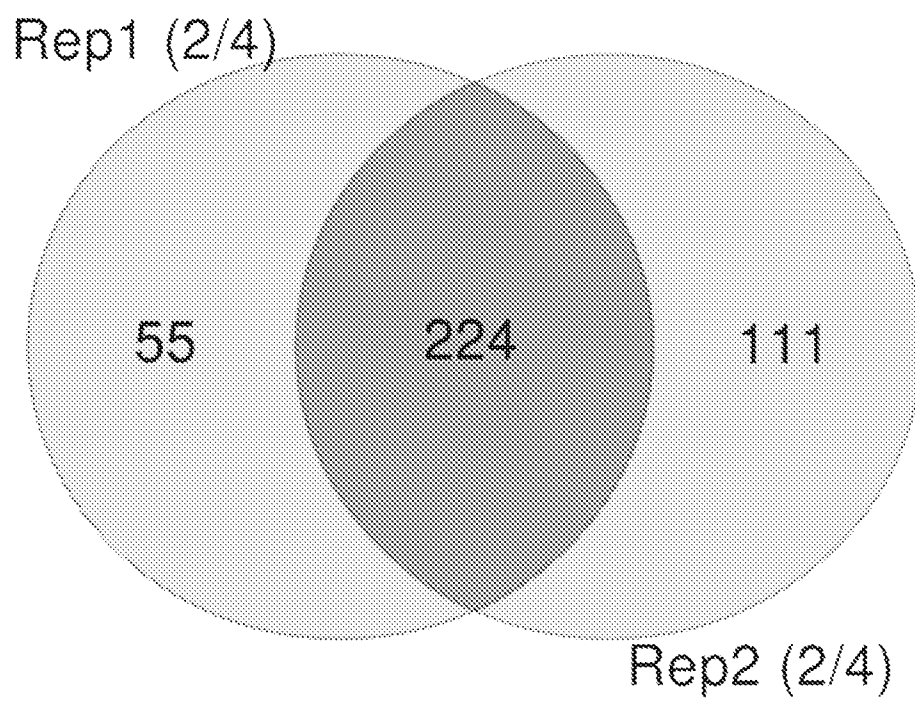


FIG. 5A

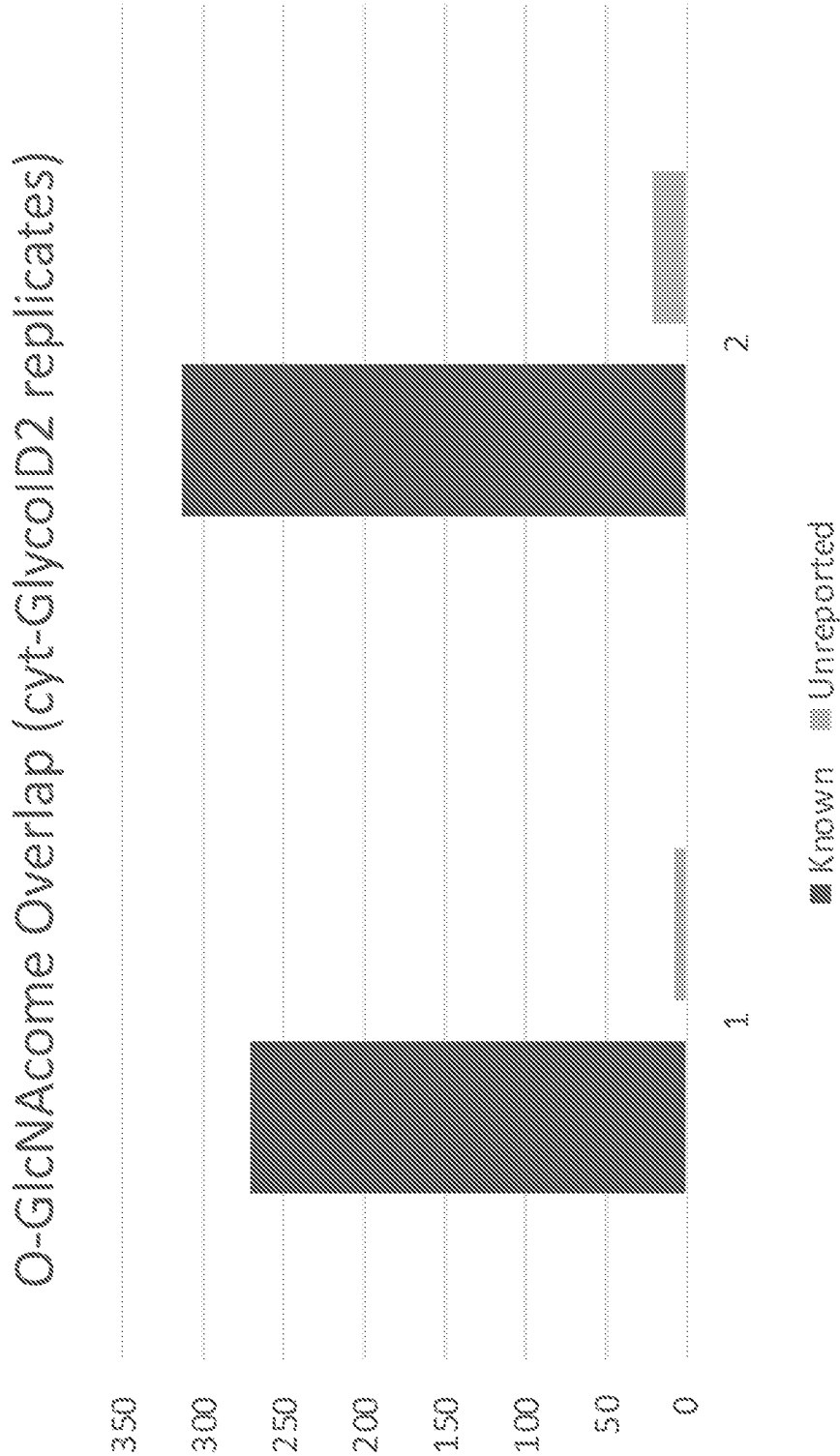


FIG. 5B

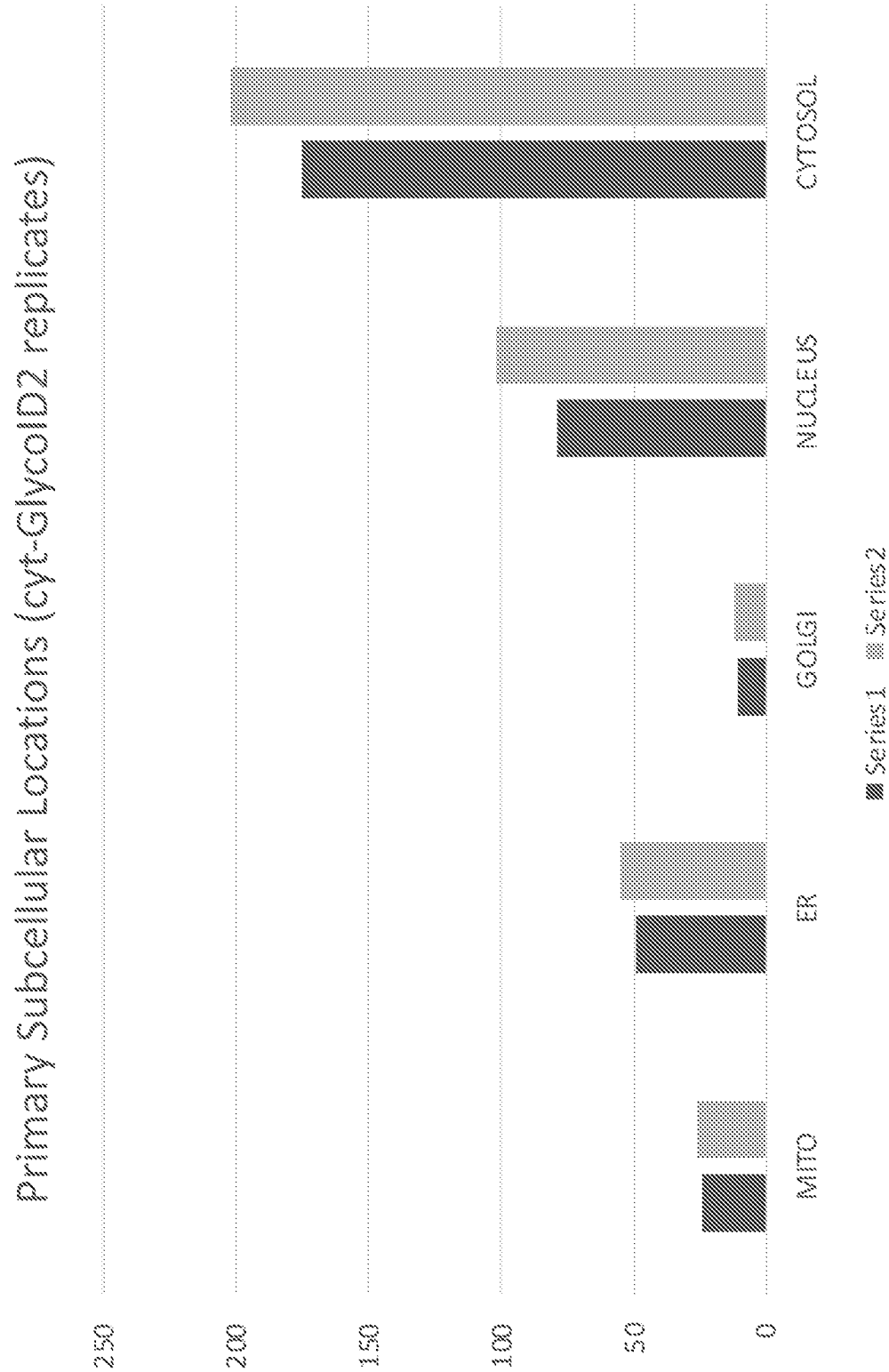


FIG. 5C

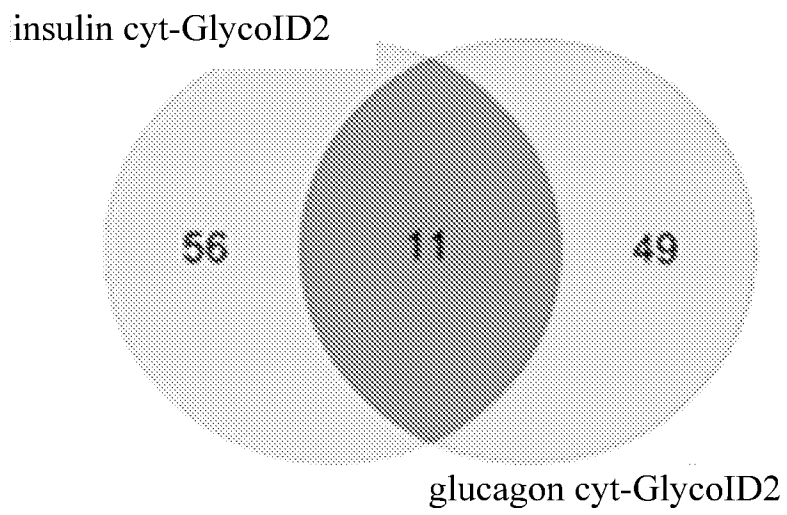


FIG. 6A

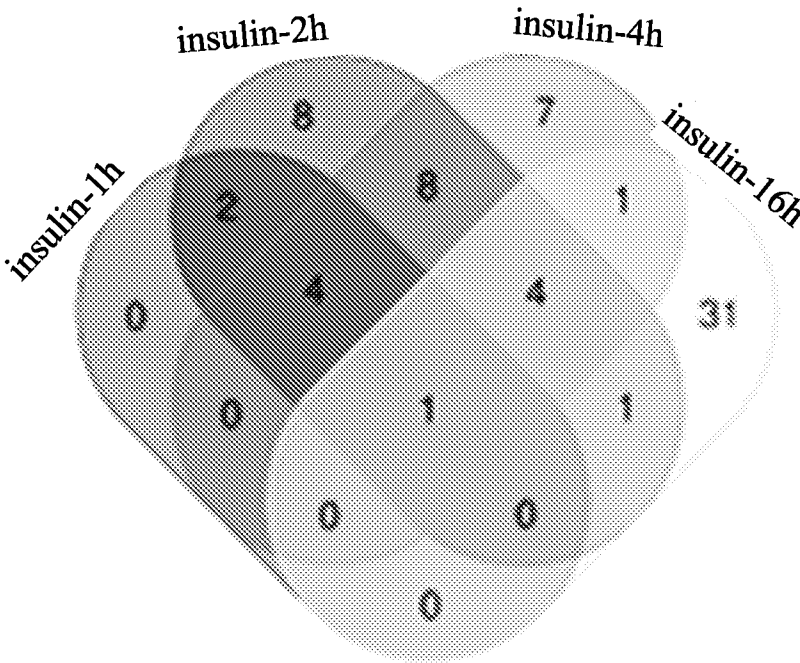


FIG. 6B

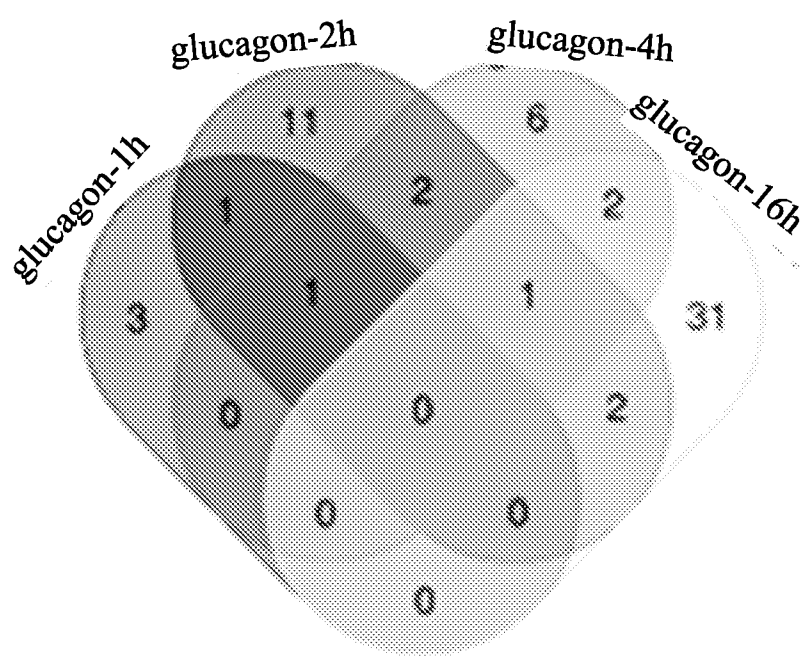


FIG. 6C

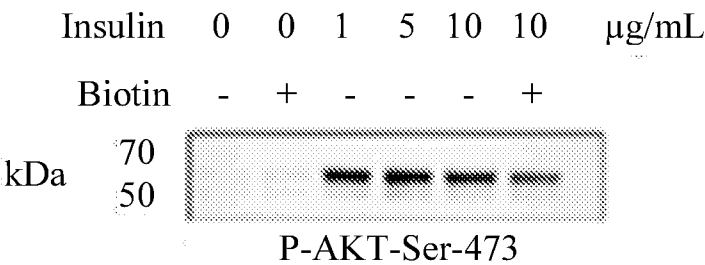
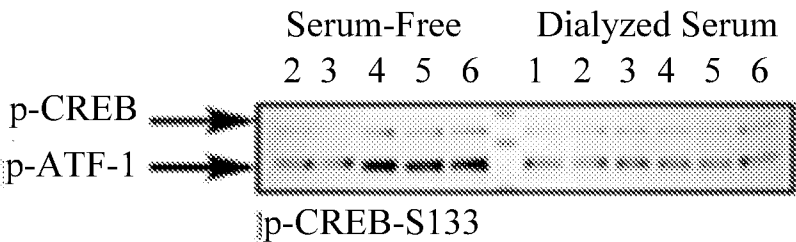


FIG. 6D



Lane 1 = control
Lanes 2-6 = glucagon stimulation

FIG. 6E