Supplementary Information

Proximity tagging identifies the glycan-mediated glycoprotein interactors of galectin-1 in muscle stem cells

Zak Vilen,[1] Eugene Joeh,[1] Meg Critcher[1], Christopher G. Parker,[2] Mia L. Huang[1,2,3]

[1] Department of Molecular Medicine, Scripps Research, 120 Scripps Way, Jupiter, FL 33458-5284
[2] Department of Chemistry, Scripps Research, 120 Scripps Way, Jupiter, FL 33458-5284
[3] Corresponding author: miahuang@scripps.edu

TABLE OF CONTENTS

Additional Methods

Generation of APEX2-galectin-1 fusion constructs

Full length DNA insert for NPG1 and CPG1 were synthesized by Genscript. Starting from the human galectin-1 protein sequence, the codon optimized sequence was generated for NPG1 and CPG1 using the proprietary OptimumGene algorithm, with codon adaptation indices of 0.97 and 0.96 respectively. Sequences were ligated into the multiple cloning site of a pETDuet-1 vector and transformed into BL21 chemically competent *E. coli*, and plated on LB agar plates under ampicillin selection (100 µg/mL). Single cell colonies were picked and grown in LB under ampicillin selection overnight. 1 mL of culture was added to 2 L of LB with ampicillin and shaken at 37 °C until OD_{600}~0.6. Protein expression was then induced with 1 mM IPTG (4 hours, 37 °C) before the bacterial pellet was harvested by centrifugation (3,300 xg, 15 minutes, 4 °C). The cell pellet was placed on ice and lysed in 15 mL lysis buffer (20 mM imidazole, 2X PBS) via sonication (5 ms ON, 40ms OFF, 1 s total ON time, 15% amplitude, 4 °C), The crude protein was purified from cell lysate using Ni-NTA beads or an AKTA Start FPLC system. Bound protein was washed with 25 mM imidazole and eluted with 250 mM imidazole before the buffer was exchanged to 1XPBS using a PD-10 column (GE Healthcare). Protein aliquots were stored at 4 °C in 1X PBS for immediate use, or flash frozen in liquid N_2 and kept at –80 °C for long term storage.

Generation of truncated APEX2 only construct

Full length DNA insert for APEX2 was synthesized by Genscript. Starting from the APEX2 engineered enzyme sequence[1], sequences were ligated into the multiple cloning site of a pETDuet-1 vector and transformed into BL21 chemically competent *E. coli* and plated on lB agar plates under ampicillin selection (100 µg/mL). Single cell colonies were picked and grown in LB under ampicillin selection overnight. 1 mL of culture was added to 2 L of LB with ampicillin and shaken (37 °C) until OD_{600}~0.6. Protein expression was then induced with 1 mM IPTG (4 hours, 37 °C) before the bacterial pellet was harvested by centrifugation (3,300 xg, 15 minutes, 4 °C). The cell pellet was placed on ice and lysed in 15 mL lysis buffer (20 mM imidazole, 2X PBS) via sonication (5 ms ON, 40ms OFF, 1 s total ON time, 15% amplitude, 4 °C), The crude protein was purified from cell lysate using Ni-NTA beads or an AKTA Start FPLC system. Bound protein was washed with 25 mM imidazole and eluted with 250 mM imidazole before the buffer was exchanged to 1XPBS using a PD-10 column (GE Healthcare). Protein aliquots were stored at 4 °C in PBS for immediate use, or flash frozen in liquid N_2 and kept at -80 °C for long term storage.

Generation of Integrin-β3 and integrin-αV

Mammalian expression constructs for Human Integrin-αV (HG11269-CH) and Human Integrin-β3 (HG10787-NH) were purchased from Sino biological and amplified in DH10B chemically competent *E. Coli*. Plasmid was isolated using the ZymoPURE II Plasmid Maxiprep Kit (D4203-B) and stored at –20 °C. HEK-293-ES cells were seeded at 1 x 10^8 cells/mL in 18mL of DMEM supplemented with 10% FBS. 20 µg of plasmid DNA was suspended in 1 mL ESF-SFM (Expression Systems, 98-001), while 50 µg of polyethylenimine (PEI) was suspended in 1 mL of ESF-SFM. Plasmid and PEI mixtures were combined and incubated to allow formation of complex (20 min, RT) in a DNA:reagent ratio of 1:2.5. Transfection mixture was added to 18mL culture and incubated (37 °C, 5% CO_2, shaking 140 rpm). Cells were harvested 72 hours post transfection by

centrifugation (600 xg, 5 min). Cell pellets were lysed by sonication and purified as described previously on AKTA Start FPLC system on Ni-NTA column. 300 mL of culture yielded ~5 mg of protein for each construct.

Growth and maintenance of mouse myoblasts

C2C12 mouse myoblast cells (ATCC CRL-1772) were cultured in DMEM supplemented with 15% fetal bovine serum at (37 °C, 5% CO_2). Cells were maintained for up to 20 passages before being discarded. For proximity labeling experiments, plastic dishes were pre-coated with poly-D-lysine and washed with PBS prior to cell seeding to improve cell retention.

Elaborated SEC Methods

Fusion constructs were diluted to desired concentrations in PBS and allowed to reach equilibrium for approximately (18 hours, 4 °C). A 4x400 mm, 5 µm MAbPac SEC-1 column (Thermofisher #074696) was prepared on a DIONEX Ultimate 3000 HPLC system and equilibrated with PBS, pH 7.14 for 15 minutes at a flow rate of 0.25 mL/min. Protein samples were loaded onto a 10 µL loop for injection and run at 0.25 mL/min with PBS, pH 7.14 for 30 minutes. The carbonic anhydrase (29 kDa) ovalbumin (44 kDa), and conalbumin (75 kDa) protein standards from a gel filtration calibration kit (GE Healthcare #28-4038-41) were run in sequence to approximate the sizes of eluting proteins and protein complexes. Protein emission was detected at 280 nm and at 405 nm with the UV detector. Peaks were manually integrated using Chromeleon (Thermo Scientific 7.3) to measure the presence of the monomeric and dimeric species. Values were normalized and DC_{50} values were calculated by non-linear regression in Prism 9.0 (GraphPad).

Imaging

After in situ proximity labeling, cells were fixed with 4% paraformaldehyde in PBS at (RT, 10 mins) and washed with PBS. Biotinylated interactors were probed for with Cy5-streptavidin (Southern Biotech 7105-15, 1 µg/mL) in PBST for 1hr at RT in the dark, and immediately after nuclei were stained with Hoescht 33342 (10 mins, RT) in the dark. Fluorescence microscopy was performed on 24-well plastic dishes with an EVOS M5000 (ThermoFisher) imager.

Gel electrophoresis and blotting

SDS-PAGE analysis of proteins and cell lysates was performed using manually cast 10% or pre-cast 5%-15% gels (Bio-Rad). Samples were loaded in Laemmli buffer supplemented with 10% (v/v) β-mercaptoethanol and boiled (95 °C, 10 mins). Gels were run (95 V, 90 mins) and proteins were transferred to PVDF membranes in running buffer supplemented with 20% (v/v) methanol for (60 mins 300 mA). Membranes were blocked with (5% (w/v) BSA/TBST, 1hr, RT) before incubation with primary antibodies (5% BSA/TBST, overnight, 4 °C) rocking. Following incubation with primary antibodies, blots were washed in TBST and incubated in with secondary antibodies for (1 hour, 5% BSA/TBST, RT). When applicable, blots were stripped with Restore PLUS Western Blot Stripping Buffer (Thermo #46430) for 15 minutes at RT rocking, before being washed with TBST and re-blotted as described above.

Antibodies / bioconjugates for detection and immunoprecipitation

Target	Conjugate	Host	Clone	Cat #	Dilution
Galectin-1	N/A	Goat	N/A	R&D AF1152	1 µg/mL (1:1000)
CD44	N/A	Rabbit	EPR18668	Abcam # ab189524	Western: 0.5 µg/mL(1:1000) IP: (15µg/1.5mg lysate)
Goat IgG	HRP	Donkey	N/A	Abcam 97110	0.1 µg/mL (1:10,000)
Rabbit IgG	HRP	Goat	N/A	Ab6721	0.2 µg/mL (1:10,000)
Biotin	HRP	N/A	N/A	Biolegend 405210	1:500
Biotin	CY5	N/A	N/A	Southern Biotech 7105-15	1 µg/mL (1:1000)

ELISAs

Asialofetuin (5 µg/mL) or appropriate recombinant glycoprotein (40 nM) was immobilized onto a 96 well plate (Nunc Maxisorp) (pH 9.6 carbonate buffer, overnight, 4 °C). The following day, plates were washed 3x with PBST and blocked (2% BSA/PBST, 1 hr, RT). Defined concentrations of PX-Gal1 constructs or recombinant human galectin-1 (Biolegend 553506) were then incubated for 1 hr rocking at RT before unbound protein was washed off with PBST. Bound protein was detected with α-galectin-1 primary antibody (R&D Systems AF1152, 1:1000) in 2% BSA/PBST by rocking (1 hr, RT) followed by HRP-conjugated secondary antibody (Abcam 97110, 1:10,000) in 2% BSA/PBST. After washing, bound antibody was detected with TMB substrate (Surmodics, TMBW-1000-01) and quenched with 2 N H_2SO_4. All conditions were performed in technical triplicate to enable statistical analysis.

Direct blotting. Membranes were blocked (5% BSA/TBST, 1 hr) before incubation with NPG1 or APEX2 (5 µM, 5 %BSA/TBST, overnight, 4 °C). Blots were washed (TBST, 15 mins, 3x) before incubation in 500 µM biotin-phenol (5% BSA/TBST, 30 mins, RT, rocking). Bound protein was activated by the addition of H_2O_2 (1 mM) and the labeling reaction was allowed to proceed (2 mins, RT, rocking). Biotinylated protein was detected by incubation with HRP conjugated streptavidin in 5% BSA/TBST(1 hr. RT), which was visualized with SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo #34577). Blots were imaged using a ChemiDoc MP imaging system (Bio-Rad).

Elaborated Proteomics Sample Preparation

1.4 x 10^7 C2C12 cells per plate were seeded on 15 cm^{-1} dishes pre-coated with poly-D-lysine and allowed to grow (overnight, 37 °C). The next day, cells were washed with PBS and incubated with various concentrations of NPG1 in DMEM (30 mins, 37 °C). Cells were washed 2x with PBS, before being incubated with biotin-phenol (500 µM, 30 mins, 37 °C) in normal growth media. Taggin was initiated with H_2O_2 (1 mM, 1 min, RT). Media was quickly aspirated and the reaction was quenched by washing 3 times with freshly prepared quencher solution (10 mM sodium ascorbate, 10 mM sodium azide, 5 mM Trolox in PBS). Cells were washed with PBS, scraped, and pelleted (600 xg, 5 min) for storage overnight at –80°C. Cell pellets were resuspended in 400 µL of PBS and lysed by sonication. Protein concentrations were normalized to 1.5 – 2 mg/mL at a final volume of 500 µL of PBS using the Lowry assay. 2 mL of cold LCMS grade MeOH was added to each lysate and protein was allowed to precipitate overnight at –20 °C. The resultant precipitate was pelleted by centrifugation (3260 xg, 15 minutes, 4 °C), and washed once with LCMS grade MeOH. The resultant pellet was allowed to dry (RT, 10 min) before being dissolved in freshly prepared proteomics grade 6 M Urea in PBS. 10 µL of 10% (v/v) SDS was added to each sample, and the mixture was sonicated in a benchtop water bath sonicator until fully dissolved. Dissolved proteins were reduced by adding 50 µL of a freshly prepared 1:1 solution of 20 mM TCEP (PBS) and 600 mM K_2CO_3 (PBS) and incubating samples shaking (30 mins, 37 °C). Free thiols were alkylated by the addition of 70 µL of freshly prepared 400 mM iodoacetamide and incubation in the dark (RT, 30 mins). 130 µL of 10% SDS was added to each sample, and samples were diluted with 5.5 mL of PBS before the addition of 100 µL washed streptavidin agarose beads. Samples were incubated while rotating to allow binding (1.5 hr, RT). Beads were pelleted by centrifugation at 500xg (2 min, 4 °C) and washed with 5 mL of 0.2% SDS in PBS, 1x PBS, and 100 mM TEAB, pH 8.5. Beads were pelleted, resuspended in 100 mM TEAB, and transferred to LoBind Microcentrifuge tubes (Fisher #13-698-794). Trypsin solution was prepared by adding 2 mL of 100 mM TEAB and 20 µL of $CaCl_2$ to one vial of sequencing grade porcine trypsin (Promega, 20 µg). 200 µL of trypsin solution was added to each sample, and samples were digested overnight shaking at 37 $^{\circ}$C. Beads were pelleted by centrifugation at 500xg (2 min, 4^{o}C) and the supernatant was transferred to a new LoBind Microcentrifuge tube. Each sample was labeled with a tandem mass tag (Thermo Scientific cat# A34808). For each sample, 8 µL of a 20 µg/µL stock of TMT reagent was added along with MS-grade acetonitrile to a final acetonitrile concentration of 30% (v/v) followed by incubation (RT, 1 hour) with occasional vortex. The reaction was quenched by addition of 6 µL hydroxylamine (6 µL, 15 mins). Samples were then acidified with formic acid (5 µL), before being dried by vacuum centrifugation. Dried samples were combined by redissolving each sample in 5% MeCN in H_2O, 0.1% formic acid. The combined sample was desalted with two C-18 columns (Thermo Fisher #89870) according to the manufacturers protocol. The desalted sample was dried by vacuum centrifugation and stored at -80 °C until ready for injection.

Extended LC/MS Gradient for Proteomics

Peptides were on-line eluted into a Fusion Tribrid mass spectrometer (Thermo Scientific) from an EASY PepMapTM RSLC C18 column (2 mm, 100 Å, 75 mm x 50 cm, Thermo Scientific), using a hold of 3% solvent B (80/20 acetonitrile/water, 0.1% formic acid) for 10 min, followed by a gradient

of 3–38% solvent B in 192 min, then a 38–75% solvent B increase in 5 min, a 75–100% solvent B in 1 min, a hold of 100% solvent B for 5 min, a decrease down to 3% solvent B in 1 min, and finally a 6 min hold of 3% solvent B. All flow rates were 250 nL/min delivered using a nEasy-LC1000 nano liquid chromatography system (Thermo Scientific). Solvent A consisted of water and 0.1% formic acid.

Extended (SPS)-MS3 method for Proteomics

scanning between 345-1500 m/z at a resolution of 120,000 for MS1 in the Orbitrap mass analyzer at an AGC target of 1E5 and a maximum injection time of 50 msec, and performing CID in the ion trap of peptide monoisotopic ions with charge 2–8 above an intensity threshold of 5E3, using a quadrupole isolation of 1.6 m/z, and at a CID energy of 30%. The ion trap AGC target was set to 1.8E4 with a maximum injection time of 120 msec. Dynamic exclusion duration was set at 20 sec and ions were excluded after 1 time within +/- 10 ppm mass tolerance window. The top 10 MS2 ions in the ion trap between 400–1200 m/z were then chosen for HCD at 65% energy and detection occurred in the Orbitrap at a resolution of 50,000, an AGC target of 1E5, and an injection time of 120 msec (MS3). All scan events occurred within a 3 sec specified cycle time.

Proteomics Data Analysis in Proteome Discoverer

Peptide sequences were determined by matching protein databases with the acquired fragmentation pattern by the SEQEST HT algorithm. The precursor mass tolerance was set to 10 ppm and fragment ion mass tolerance to 0.6 Da. One missed cleavage site of trypsin was allowed. Carbamidomethyl and TMT-10plex were used as a static modification. Oxidation was used as variable modification. All spectra were searched against the proteome database of *mus musculus* (17,068 sequences) using a target false discovery rate of 1%. Proteins identified were filtered by at least 3 unique peptides prior to statistical analysis using Prism 9.0 (GraphPad Software Inc.). TMT values obtained from proteome discoverer were transformed with log2-(x) and p-values were obtained from t-test function over two biological replicates with 3x technical replicates each.

Elaborated CD44 Immunoprecipitation

Confluent 15 cm^2 dishes of C2C12 cells were scraped and pelleted by centrifugation (500 xg, 3 min, 4 °C). Cells were washed with PBS 2x and then lysed with RIPA buffer supplemented with protease inhibitor cocktail and sonication. Lysate was pre-clarified by centrifugation at (16,000 xg, 15 min, 4 °C) and the supernatant was added to 100 µL of washed Protein G Sepharose Fast Flow beads (GE Healthcare) and incubated (10 mins at 4 °C) to remove non-specific bead binding. Beads were pelleted by centrifugation at 1,450 xg for (10 minutes, 4 °C) and the supernatant was collected. 10 µg of primary antibody (Abcam # ab189524, Clone: EPR18668) was added to pre-cleared cell lysate and incubated (overnight, 4 °C) to form the immunocomplex. The immunocomplex was captured by incubating samples with 100 µL of Protein G Sepharose Fast Flow beads (overnight, 4 °C). Beads were pelleted by centrifugation at (1,450 xg, 10mins, 4 °C) and the supernatant was set aside. Beads were washed with PBS, and the bound protein was eluted with 100 µL of glycine (0.2 M, pH 2.6) and incubating (RT, 10 min). Beads were pelleted and the supernatant was harvested before being quenched with an equal volume of cold Tris buffer (1 M, pH 9.0). Elution and quenching were repeated two more times for a total of three elutions. 4x Laemmli buffer was added to the beads and boiled (95 °C, 10 mins) to remove any remaining immunocomplex from the beads.

Elaborated Sample Preparation for N-Glycomics

C2C12 cells were cultured to ~ 90% confluency in 10 cm^2 plates, yielding approximately 8 x 10^6 cells per plate. Cells were harvested into 15 mL falcon tubes and washed four to five times with cold PBS followed by centrifugation (600 rpm, 10 min, 4 °C). The cell surface (glyco)proteins were digested by incubating the cell pellet with trypsin from bovine pancreas (Sigma) (500 µL per sample, 2.5 mg/mL stock solution) (15 min, 37 °C). Digested samples were brought to 4 °C on ice followed by centrifugation (14,000 xg, 15 mins, 4 °C). Trypsin was heat-inactivated by boiling the supernatant containing the glycopeptides (100 °C, 5 min). The sample was cooled on ice for 5-10 mins. Glycans were released from the glycopeptides by incubating the sample with 2.0 µL of PNGase F overnight at 37 °C (7.5 µg/µL in 25 mM Tris pH 7.2 with 50% glycerol). Once released, the glycans were purified from the remaining peptides through a C18 column (Thermo Scientific HyperSep, #03251257). The C18 columns were first equilibrated by running methanol, 5% aqueous acetic acid, *n*-propanol, and 5% aqueous acetic acid, in order. Once the column was ready, samples were first acidified to 5% acetic acid, spun down to remove any precipitate at 14,000 xg for 5 min at 4 °C), then loaded to the column. Once the sample had fully entered the column, the freed glycans were eluted with a total of 5 mL of 5% (v/v) acetic acid in water. The eluates were combined, frozen and lyophilized. Experiments were repeated across two independent biological replicates. Permethylated glycan samples were dissolved in 20% MeCN (aq.) prior to LC-MS/MS analysis and stored at -20 °C.

Extended LC gradient for N-Glycomics

Column temperature was maintained at 50-55 °C. Mobile phase A consisted of water (18.2 MΩ) with 0.1% formic acid (Thermo Scientific #28905), and mobile phase B was MeCN (Fisher #A994) with 0.1% formic acid. Flow rate was maintained at 100 or 120 µL/min. The elution gradient was from 20% buffer B to 60% buffer B over span of at least 90 min, increased up to 95% B over the final 15 min.

Extended MS Method for N-Glyomics

The mass spectrometer was operated in positive ion mode in the mass range of 700 m/z – 2,000 m/z with a spray voltage of 3.5 kV. MS^2 data was collected in data-dependent acquisition mode, with the top four most abundant ions with signal intensity over 10,000 counts selected from the full MS^1 scan for collision induced dissociation (CID). Calibration was performed regularly (Thermo LTQ ESI Positive Ion Calibration Solution, #88322) to ensure accuracy of 10 ppm or less.

Amino acid sequence of NPG1 construct

APEX2 in blue | spacer sequence | Human Galectin-1 (Uniprot P09382)

N Terminal HIS-APEX2-GAL1 (NPG1)
Theoretical pI: 5.73
Theoretical Mw: 43149.60
EFMHHHHHHGKSYPTVSADYQDAVEKAKKKLRGFIAEKRCAPLMLRLAFHSAGTFDKGTKTGGPF
GTIKHPAELAHSANNGLDIAVRLLEPLKAEFPILSYADFYQLAGVVAVEVTGGPKVPFHPGREDKPEP
PPEGRLPDPTKGSDHLRDVFGKAMGLTDQDIVALSGGHTIGAAHKERSGFEGPWTSNPLIFDNSYF
TELLSGEKEGLLQLPSDKALLSDPVFRPLVDKYAADEDAFFADYAEAHQKLSELGFADAGSGGGGS
ACGLVASNLNLKPGECLRVRGEVAPDAKSFVLNLGKDSNNLCLHFNPRFNAHGDANTIVCNSKDGG
AWGTEQREAVFPFQPGSVAEVCITFDQANLTVKLPDGYEFKFPNRLNLEAINYMAADGDFKIKCVAF
D

Sequence comparison between human and mouse galectin-1

Human Galectin-1 (Uniprot P09382)
Theoretical Mw: 14,716
MACGLVASNLNLKPGECLRVRGEVAPDAKSFVLNLGKDSNNLCLHFNPRFNAHGDANTIVCNSKD
GGAWGTEQREAVFPFQPGSVAEVCITFDQANLTVKLPDGYEFKFPNRLNLEAINYMAADGDFKIKC
VAFD

Mouse Galectin-1 (Uniprot P16045)
Theoretical Mw: 14,866
MACGLVASNLNLKPGECL VRGEVASDAKSFVLNLGKDSNNLCLHFNPRFNAHGDANTIVCNTKED
GTWGTEHREPAFPFQPGSITEVCITFDQADLTIKLPDGHEFKFPNRLNMEAINYMAADGDFKIKCVA
FE

-> Sequence identity by NCBI Protein BLAST = 119/135 (88%)

Amino acid sequence of APEX2 truncated construct

Theoretical Mw: 28,804
MHHHHHHGKSYPTVSADYQDAVEKAKKKLRGFIAEKRCAPLMLRLAFHSAGTFDKGTKTGGPF
GTIKHPAELAHSANNGLDIAVRLLEPLKAEFPILSYADFYQLAGVVAVEVTGGPKVPFHPGREDKP
EPPPEGRLPDPTKGSDHLRDVFGKAMGLTDQDIVALSGGHTIGAAHKERSGFEGPWTSNPLIFD
NSYFTELLSGEKEGLLQLPSDKALLSDPVFRPLVDKYAADEDAFFADYAEAHQKLSELGFADAGS
GGGGSLPETG

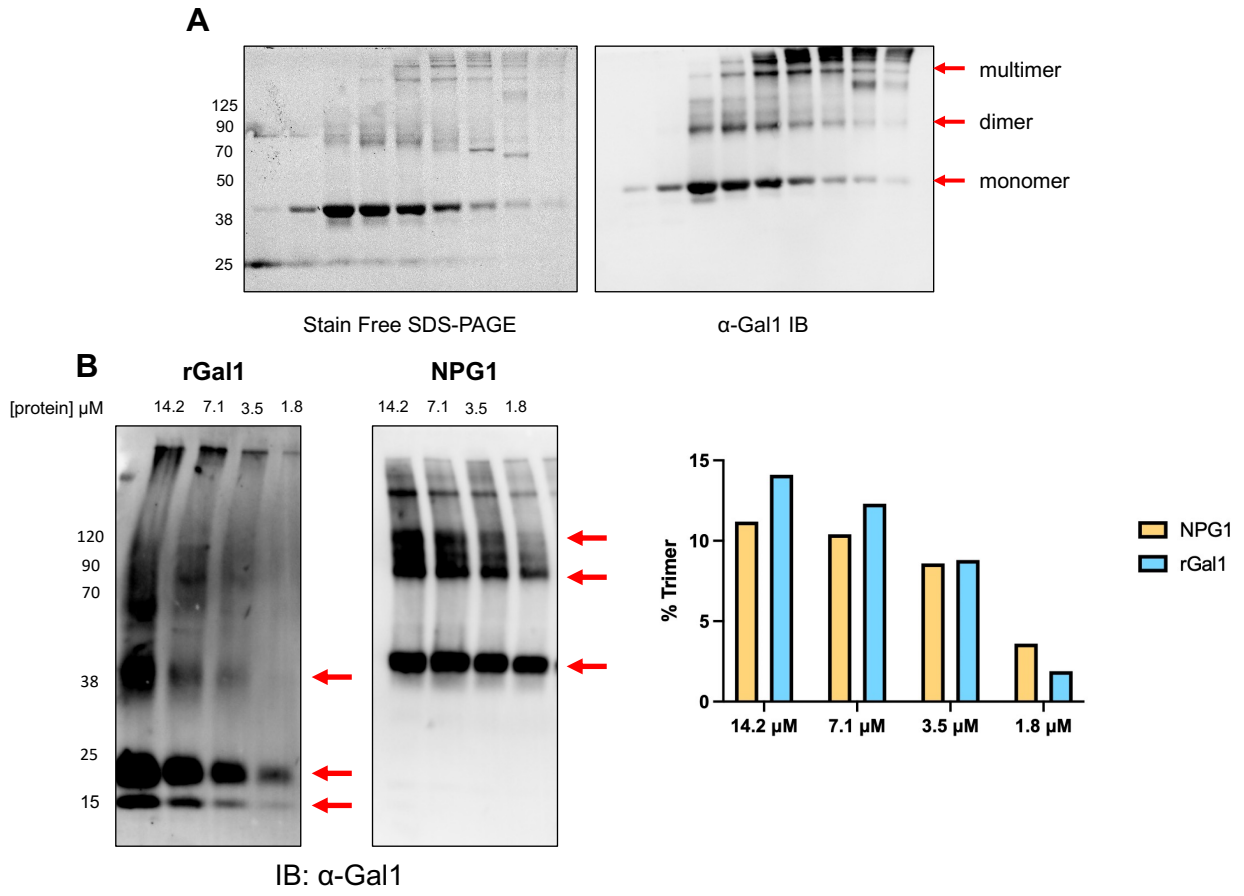


Figure S1. Characterization of NPG1. a) SDS-PAGE (left) and western blot analysis (right) of NPG1 fractions following expression and purification. The presence of higher molecular weight bands (> 43 kDa corresponding to the monomer) suggest the spontaneous formation of the dimer and higher molecular weight species in high concentrations of NPG1. Samples were run on a 10% SDS-PAGE gel (90 V, 95 min) and transferred to a PVDF membrane. The membrane was blocked and galectin-1 was detected with α-galectin-1 antibody. b) α-galectin-1 Western blots of decreasing concentrations of recombinant human galectin-1 and NPG1. Galectin-1 detected by α-galectin-1 antibody (R&D AF1152, 1 μg/mL; 1:1000) and α-goat HRP (Abcam 97110, 0.1 μg/mL; 1:10,000). Concentrations decrease from left to right (14.2 μM, 7.1 μM, 3.5 μM, 1.8 μM) and monomers, dimers, and trimer bands are indicated with red arrows. Multimer species are only observable at high concentrations, and disappear rapidly upon dilution, which is quantified by densitometric analysis in ImageLab (BioRad) as (adjusted band volume/adjusted lane volume). The rate at which trimers of rGal1 and NPG1 disappear upon dilution is approximately equal, indicating that multimer formation is conserved between recombinant galectin-1 and the NPG1 construct.

Table S1. Analysis parameters for ELISA data (**Fig. 1C**). Data analyzed using Graphpad Prism 9.0.

log(agonist) vs. response -- Variable slope (four parameters)	NPG1	NPG1 + Lac	CPG1	CPG1 + Lac
Best-fit values				
Bottom	0.7236	-0.09384	-1.869	-0.2402
Top	100.1	49.15	97.18	68.43
LogEC50	1.509	2.740	1.751	2.809
HillSlope	1.463	1.787	1.318	1.498
EC50	32.29	550.1	56.31	644.3
Span	99.37	49.25	99.05	68.67
95% CI (profile likelihood)				
Bottom	-3.202 to 4.371	-3.652 to 3.228	-4.749 to 0.8501	-1.486 to 0.9744
Top	96.80 to 103.6	40.48 to 98.75	93.76 to 100.9	63.22 to 77.04
LogEC50	1.446 to 1.570	2.555 to 3.421	1.693 to 1.810	2.743 to 2.913
HillSlope	1.207 to 1.811	0.7591 to ???	1.124 to 1.550	1.203 to 1.870
EC50	27.93 to 37.19	359.1 to 2638	49.34 to 64.63	553.7 to 819.1
Goodness of Fit				
Degrees of Freedom	17	17	17	17
R squared	0.9937	0.9272	0.9954	0.9942
Sum of Squares	236.4	475.6	160.7	58.31
Sy.x	3.729	5.289	3.075	1.852
Number of points				
# of X values	21	21	21	21
# Y values analyzed	21	21	21	21

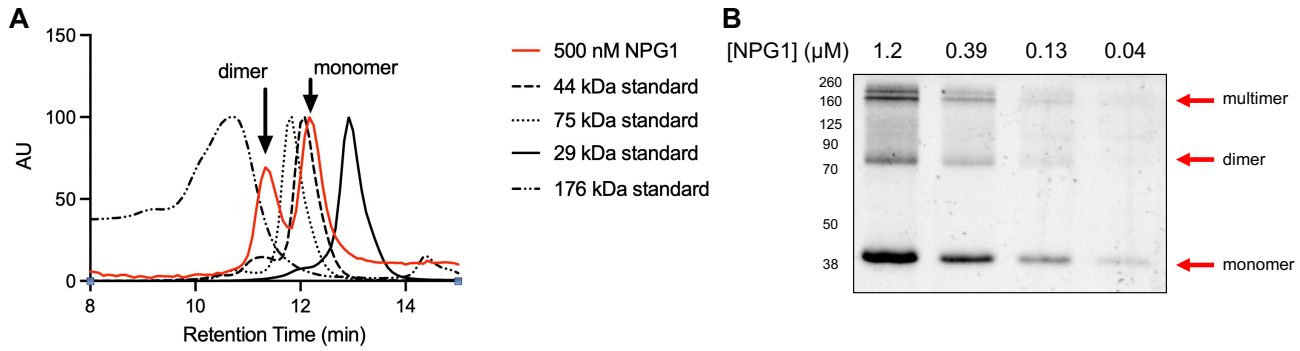


Figure S2. Evaluation of NPG1 dimerization. a) SEC spectra of protein standards and representative NPG1 containing both monomer and dimer peaks. Protein standards included carbonic anhydrase (29 kDa) ovalbumin (44 kDa), and conalbumin (75 kDa) obtained from a gel filtration calibration kit (GE Healthcare #28-4038-41) as well as α-mouse HRP conjugate IgG (~176 kDa). Samples and standards were equilibrated for 18 hrs at 4 °C before being run in PBS pH 7.4 for 30 minutes. The retention time of the major peaks (monomer and dimer) correspond to the expected molecular weights of the NPG1 construct (44 kDa and 88 kDa respectively). b) SDS-PAGE gel run under non-reducing conditions of a serial dilution of NPG1 construct. NPG1 was diluted and allowed to equilibrate overnight at 4°C in PBS before being run on a 10% SDS-PAGE gel (95 V, 90 min) and stained with Coomassie blue. Gel indicates the presence of the monomer (~44 kDa), dimer (~88 kDa), and multimeric species, all indicated with red arrows. At lower concentrations, the monomer is the dominant species.

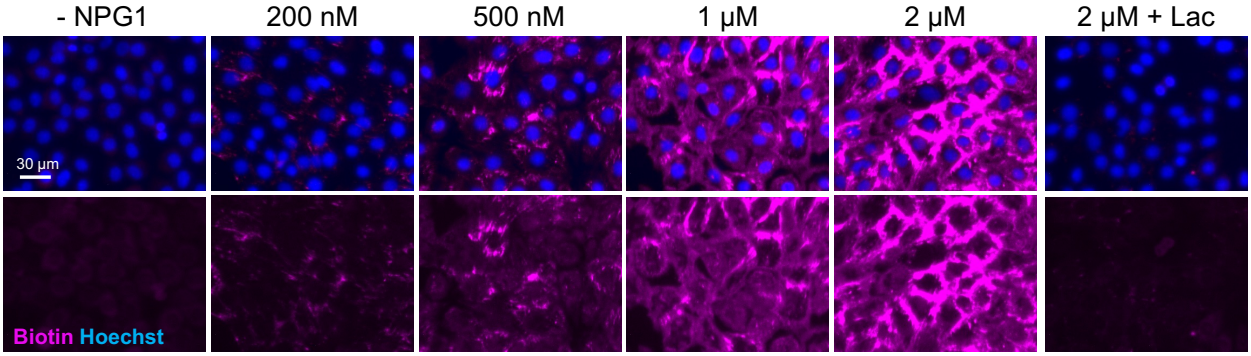


Figure S3. Extended dose dependent labeling of live myoblasts by NPG1. Live C2C12 cells were pre-incubated with varying concentrations of NPG1 (200 nM – 2 μM) before labeling was initiated by the addition of biotin-phenol (500 μM) and H_2O_2 (1 mM). After quenching, cells were fixed in 4% PFA and stained using Hoescht and a streptavidin-CY5 conjugate. The biotin signal (purple) was observed at concentrations as low as 200 nM, and increased in a dose dependent fashion. Stains appeared to localized primarily at the cell surface and ECM, surrounding but not overlapping the nucleus (blue). The addition of lactose (100 mM) drastically decreased the apparent biotin signal indicating that a majority of the detected interactions are glycan-mediated.

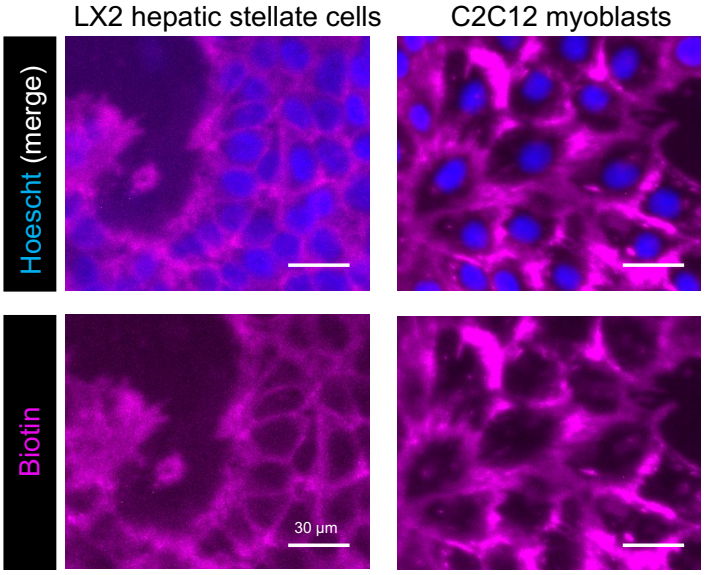


Figure S4. Fluorescence microscopy of galectin fusion constructs and ECM components APEX labeling performed in two different cell lines, C2C12 myoblasts with high ECM volume and LX2 hepatic stellate cells with low ECM stained for biotin (purple) post APEX tagging with APEX-galectin fusion constructs. LX2 cells labeled with galectin-3 fusion protein demonstrate strong signal localized to cell surface, seen as a thin line surrounding nuclei (blue). C2C12 cells labeled with galectin-1 fusion protein (NPG1) demonstrate strong signal localized to cell surface and ECM between cells, visible as thick diffuse signal surrounding nuclei. The pattern of staining generated by NPG1 in C2C12s closely resembles the staining of ECM components, such as perlecan and collagen, in undifferentiated C2C12 myoblasts[1].

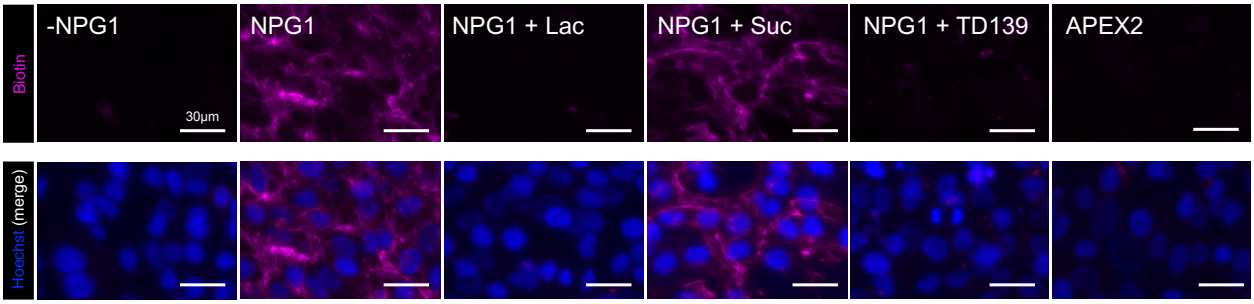


Figure S5. Extended controls for proximity tagging. Prior to labelling, 5μM NPG1 was pre-incubated with lactose (100 mM), sucrose (100 mM), or TD139 (62 μM). Live C2C12 cells were then incubated with 5μM NPG1 (+ or – inhibitors) or a truncated APEX construct not containing the galectin-1 sequence. Labeling was initiated by further incubated with biotin-phenol (500 μM) and H_2O_2 (1 mM). After quenching, cells were fixed in 4% PFA and stained with Hoescht and a streptavidin-CY5 conjugate. A strong biotin signal could be observed localized to the ECM in the NPG1 condition and the NPG1 + sucrose condition. Lactose and TD139 competed out the biotin signal, indicating successful inhibition of carbohydrate binding activity. Incubation with the truncated APEX construct did not yield any significant signal when compared to the negative control, indicating that the localization and labeling is the result of galectin-1 rather than the APEX2 enzyme.

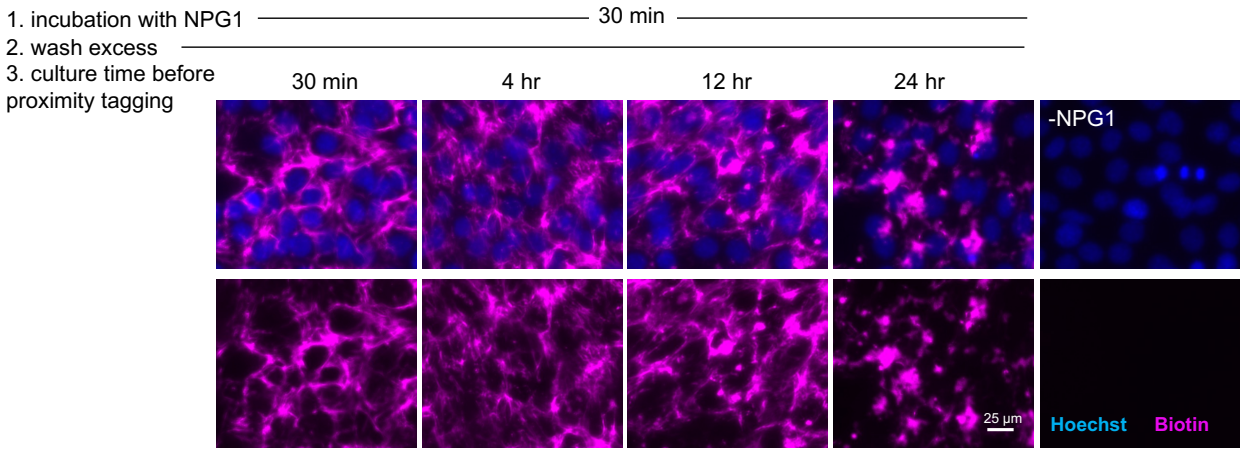


Figure S6. Fluorescence microscopy images of C2C12 myoblasts following extended pre-incubation with NPG1 and proximity tagging. NPG1 (5 μM) was incubated with live cells in serum free growth media for 30 mins., and excess unbound NPG1 was removed by washing 2x with DPBS. The cells were then continuously cultured in DMEM for the indicated time points (30 min - 24 hr) prior to proximity tagging, fixation, and fluorescence imaging. All timepoints generated significant fluorescence signal compared to the negative control (no NPG1). Whereas shorter continuous incubation periods (30 min - 4hr) generated signal primarily localized to the cell surface or the extracellular matrix, with pro-longed continuous incubation, an increased localization of the biotin signal into the cytosol is observed. This is in agreement with previous literature, where galectin-1 has been observed to localize to extracellular vesicles during myoblast differentiation[2, 3].

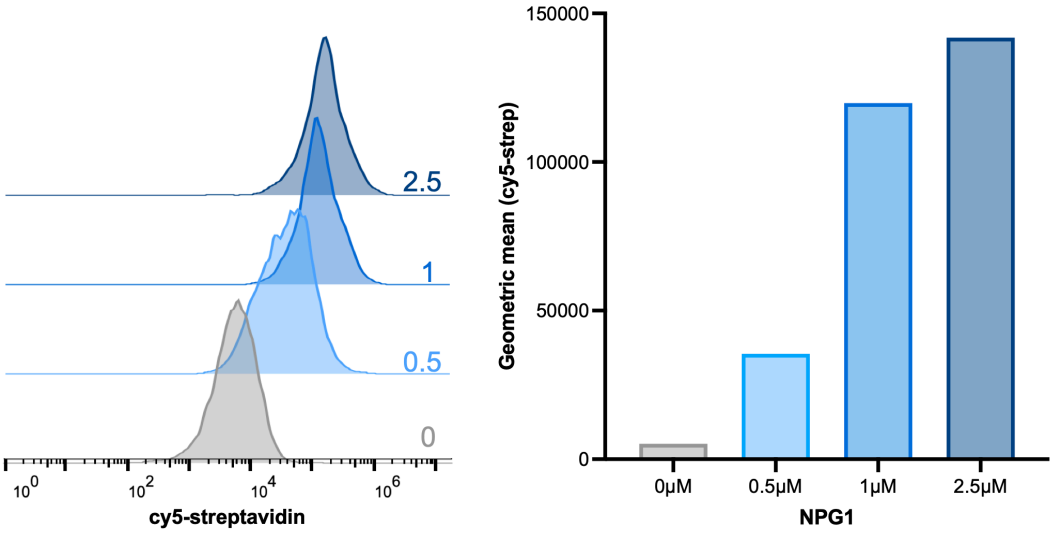


Figure S7. Detection of biotinylated interactors by flow cytometry. Incubation of proximity-tagged myoblasts following staining with CY5-streptavidin (Southern Biotech 7105-15, 1μg/mL; 1:1000) exhibits a dose dependent increase in fluorescence upon NPG1 treatment compared to non-treated cells (grey). The histogram (left) indicates a dose dependent shift in detected fluorescence, which is quantified by the shift in geometric mean of the peaks (right).

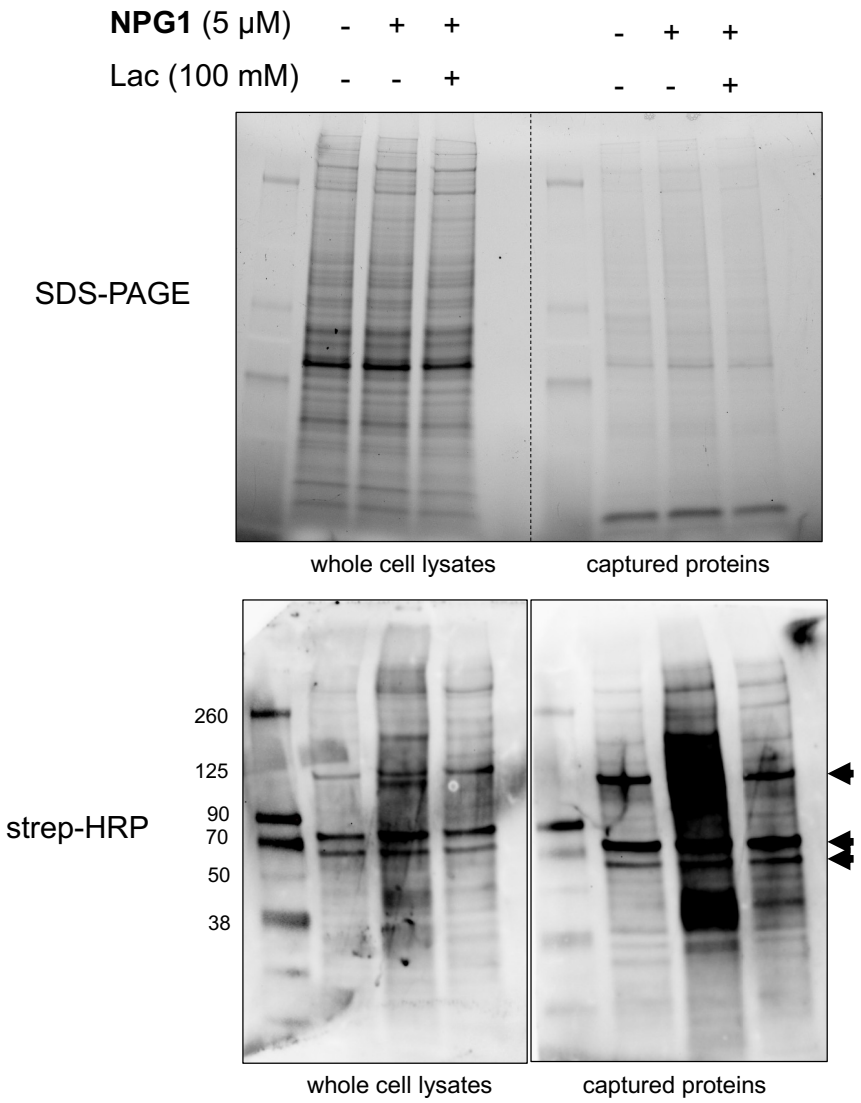


Figure S8. Enrichment of NPG1 interactors by streptavidin solid support. Proximity tagging was performed on live C2C12 myoblasts with 5 μM NPG1 construct before cells were lysed. 15 μg of lysate was set aside, and 200 μg of lysate was enriched on streptavidin beads for 1.5 hr at RT. After washing, the enriched lysate was eluted from beads with 250 mM imidazole, and enriched proteins were run alongside the unenriched, whole cell lysate. Lysates were blotted with streptavidin conjugated to HRP (Biolegend 405210, 1:500), revealing that despite significantly greater amount of protein visualized in the SDS-PAGE, the enriched sample contained a relatively larger number of biotinylated interactors, indicating successful enrichment. Endogenously biotinylated proteins detected were detected in both the negative control and competed conditions and are indicated with arrowheads. These bands correspond approximately to the known endogenously biotinylated mammalian proteins pyruvate carboxylase (129 kDa), propionyl-CoA carboxylase (80 kDa), and methylcrotonoyl-CoA carboxylase (79 kDa).

Table S2. Highly significant and enriched protein interactors of NPG1. Glyco-site annotation and GO localization assigned by Uniprot. For GO localization, CM = Cell Membrane, ECM = Extracellular Matrix

Gene Name	Protein Name	Uniprot ID	Glyco-Sites	GO Localization
Cadm1	Cell adhesion molecule 1	Q8R5M8	6	CM
Cd44	CD44 antigen	P15379	10	CM
Prnp	Major prion protein	P04925	2	CM
Cd80	T-lymphocyte activation antigen CD80	Q00609	6	CM
Fbln1	Fibulin-1	Q08879	3	Secreted
Atp1b3	Sodium/potassium-transporting ATPase subunit beta-3	P97370	2	CM
F5	Coagulation factor V	O88783	8	Secreted
Cd200	OX-2 membrane glycoprotein	O54901	6	CM
Ptgfrn	Prostaglandin F2 receptor negative regulator	Q9WV91	8	ER Membrane
Smpdl3b	Acid sphingomyelinase-like phosphodiesterase 3b	P58242	5	Secreted
Lphn2; Adgrl2	Adhesion G protein-coupled receptor L2	Q8JZZ7	3	CM
Vcam1	Vascular cell adhesion protein 1	P29533	5	CM
Efnb1	Ephrin-B1	P52795	1	CM
Epha2	Ephrin type-A receptor 2	Q03145	2	CM
Itga5	Integrin alpha-5	P11688	14	CM
Cdh2	Cadherin-2	P15116	7	CM
Atp1b1	Sodium/potassium-transporting ATPase subunit beta-1	P14094	3	CM
Cd97; Adgre5	Adhesion G protein-coupled receptor E5	Q9Z0M6	7	CM
Dag1	Dystroglycan	Q62165	7	CM, Secreted
Tpbg	Trophoblast glycoprotein	Q9Z0L0	2	CM
Anpep	Aminopeptidase N	P97449	13	CM
Gas1	Growth arrest-specific protein 1	Q01721	2	CM
Enpp1	Ectonucleotide pyrophosphatase/phosphodiesterase family member 1	P06802	6	CM
Plaur	Urokinase plasminogen activator surface receptor	P35456	7	CM, Secreted
Vasn	Vasorin	Q9CZT5	6	CM, ECM
Nradd	Death domain-containing membrane protein NRADD	Q8CJ26	2	CM, Nucleus
Itgb5	Integrin beta-5	O70309	7	CM
Emb	Embigin	P21995	9	CM
Plxnb2	Plexin-B2	B2RXS4	8	CM, ECM
Cdh13	Cadherin-13	Q9WTR5	7	CM
Fbln2	Fibulin-2	P37889	4	Secreted, ECM
Antxr1	Anthrax toxin receptor 1	Q9CZ52	3	CM
Cr1l	Complement component receptor 1-like protein	Q64735	2	CM
Kirrel	Kin of IRRE-like protein 1	Q80W68	4	CM
Antxr2	Anthrax toxin receptor 2	Q6DFX2	1	CM
Slc3a2	4F2 cell-surface antigen heavy chain	P10852	8	CM
Egfr	Epidermal growth factor receptor	Q01279	10	CM
Pdgfrb	Platelet-derived growth factor receptor beta	P05622	11	CM
Islr	Immunoglobulin superfamily containing leucine-rich repeat protein	Q6GU68	2	Secreted
Ptk7	Inactive tyrosine-protein kinase 7	Q8BKG3	10	CM
Itgb3	Integrin beta-3	O54890	5	CM

Tfrc	Transferrin receptor protein 1	Q62351	5	CM
Itgb1	Integrin beta-1	P09055	12	CM
Cp	Ceruloplasmin	Q61147	6	Secreted
Cd14	Monocyte differentiation antigen CD14	P10810	5	CM
Notc2	Neurogenic locus notch homolog protein 2	O35516	6	CM, ECM
Igsf8	Immunoglobulin superfamily member 8	Q8R366	3	CM
Cd109	CD109 antigen	Q8R422	7	CM
Cdh15	Cadherin-15	P33146	5	CM
Mfge8	Lactadherin	P21956	4	CM, ECM
Ece1	Endothelin-converting enzyme 1	Q4PZA2	10	CM
Cdon	Cell adhesion molecule-related/down-regulated by oncogenes	Q32MD9	8	CM
Itga3	Integrin alpha-3	Q62470	13	CM
Itgav	Integrin alpha-V	P43406	11	CM
Igf1r	Insulin-like growth factor 1 receptor	Q60751	16	CM
Hspg2	Basement membrane-specific heparan sulfate proteoglycan core protein	Q05793	13	Secreted, ECM
Sgcd	Delta-sarcoglycan	P82347	3	CM
Nptn	Neuroplastin	P97300	6	CM
Env	MLV-related proviral Env polyprotein	P10404	5	CM
Mrc2	C-type mannose receptor 2	Q64449	5	CM
Adam10	Disintegrin and metalloproteinase domain-containing protein 10	O35598	4	CM
Flrt2	Leucine-rich repeat transmembrane protein FLRT2	Q8BLU0	1	CM
Osmr	Oncostatin-M-specific receptor subunit beta	O70458	14	CM
Gpc1	Glypican-1	Q9QZF2	5	CM, ECM
H2-K1	H-2 class I histocompatability antigen	P04223-2	0	CM
Cspg4	Chondroitin sulfate proteoglycan 4	Q8VHY0	16	CM, ECM
Vcan	Versican core protein	Q62059	17	Secreted, ECM
Slc44a2	Choline transporter-like protein 2	Q8BY89	3	CM
Efemp2	EGF-containing fibulin-like extracellular matrix protein 2	Q9WVJ9	2	Secreted, ECM
Mmp14	Matrix metalloproteinase-14	P53690	0	CM
Myadm	Myeloid-associated differentiation marker	O35682	0	CM
Atp1a1	Sodium/potassium-transporting ATPase subunit alpha-1	Q8VDN2	0	CM
Ppic	Peptidyl-prolyl cis-trans isomerase C	P30412	0	Cytoplasm
Lgals1	Galectin-1	P16045	0	Secreted

Table S3. Glycoproteins sourced for ELISA validation. Proteins produced in house are marked*

Glycoprotein	Uniprot ID	Species	Produced in	Catalog # / manufacturer
Notch-2	O35516.2	Mouse	CHO cell line	5196-NT-050 /R&D systems
Gas-1	Q01721	Mouse	mouse myeloma cell line	2644-GS-050 /R&D systems
Ece-1	Q4PZA2.1	Mouse	CHO cell line	5796-ZN-010 /R&D systems
Epha-2	Q03145	Mouse	mouse myeloma cell line	639-A2-200 /R&D systems
Integrin alpha-V*	P06756	Human	HEK293-T cell line	N/A
Integrin beta-3*	P05106	Human	HEK293-T cell line	N/A

Table S4. Analysis parameters for ELISA data Fig. 4A. Data analyzed using GraphPad Prism version 9.0

log(agonist) vs. normalized response -- Variable slope	Itgβ-3	Itgβ-3 + Lac	Ece-1	Ece-1 + Lac	Epha-2	Epha-2 + Lac
Best-fit values						
LogEC50	1.621	2.025	2.193	2.402	1.137	1.656
HillSlope	0.8288	0.8492	0.9055	0.8672	0.8110	0.9125
EC50	41.80	106.0	155.9	252.2	13.72	45.28
95% CI (profile likelihood)						
LogEC50	1.570 to 1.672	1.977 to 2.074	2.063 to 2.323	2.293 to 2.510	1.109 to 1.166	1.609 to 1.703
HillSlope	0.7642 to 0.9018	0.7848 to 0.9215	0.7307 to 1.150	0.7253 to 1.051	0.7751 to 0.8493	0.8434 to 0.9905
EC50	37.18 to 46.98	94.84 to 118.5	115.5 to 210.3	196.3 to 323.6	12.85 to 14.66	40.67 to 50.42
Goodness of Fit						
Degrees of Freedom	22	22	22	22	22	22
R squared	0.9953	0.9954	0.9650	0.9734	0.9985	0.9959
Sum of Squares	178.9	164.8	1270	845.9	55.80	165.1
Sy.x	2.852	2.737	7.599	6.201	1.593	2.739
Number of points						
# of X values	24	24	24	24	24	24
# Y values analyzed	24	24	24	24	24	24

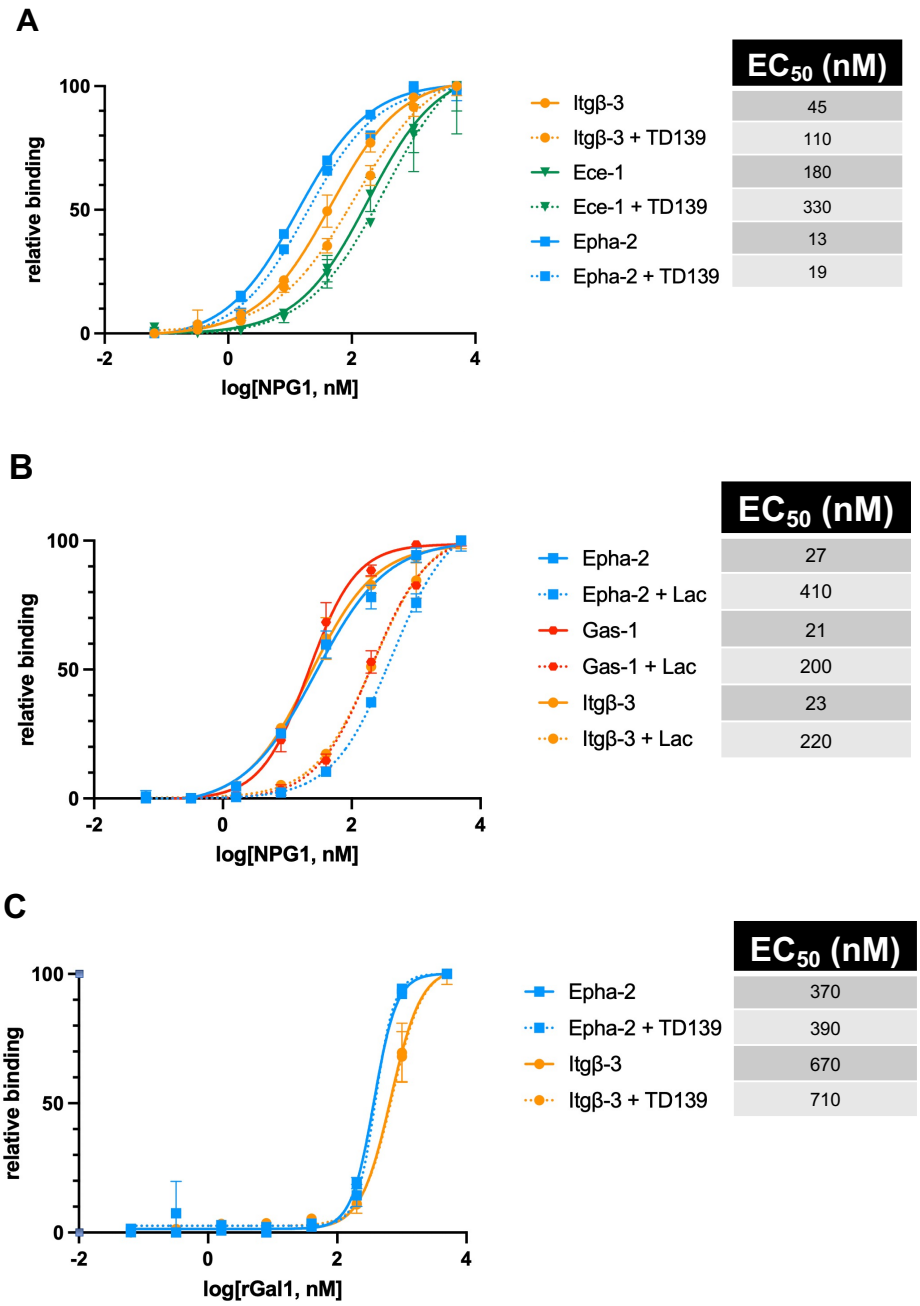


Figure S9. Additional ELISAs evaluating binding of NPG1 and recombinant galectin-1. a) Glycoproteins identified within the proteomics dataset were produced in-house in HEK293T cells or purchased commercially. In the competition conditions, proteins were pre-incubated for 15 min in TD139 (62 μM). Data was analyzed in GraphPad Prism version 9.0. All constructs saturated the plate, indicating significant binding to NPG1. Itgβ-3, Epha-2, and Ece-1 demonstrate partial competition upon incubated with TD139, suggesting that glycans may partially mediate this interaction. b) Competition conditions were performed by pre-incubating glycoproteins of interest in lactose (200 mM). Epha-2, Gas-1, and Itgβ-3, demonstrated significant binding to NPG1, and were significantly competed by the addition of 200 mM lactose, suggesting glycans are partially responsible for the NPG1-glycoprotein interaction. **C)** Serial dilutions of recombinant human galectin-1 purchased commercially were applied to the coated plates rather than the NPG1 construct. Significant binding was detected between Epha-2 and Itgβ-3 and rGal1, however neither interaction was successfully competed by the addition of TD139 (62 μM).

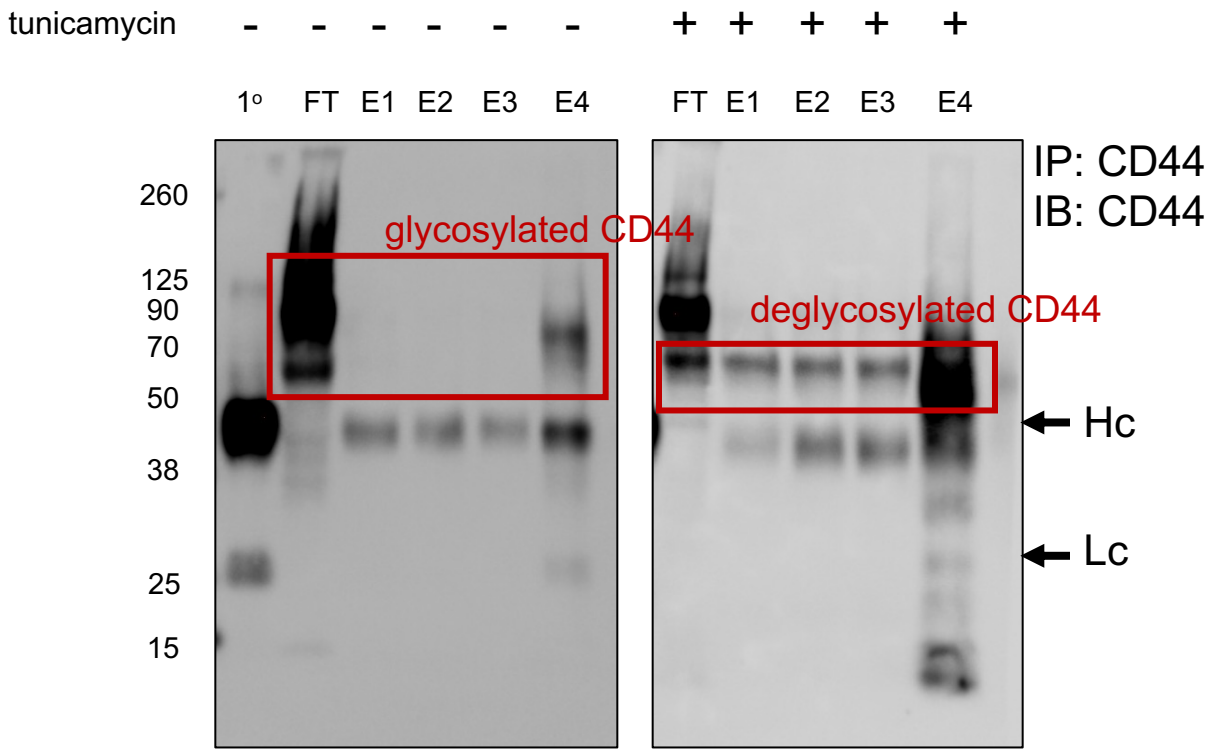


Figure S10. Immunoprecipitation of CD44 from C2C12 myoblasts. Tunicamycin treated cells were pre-incubated with 10 µg/mL Bound protein was eluted 3x with 0.2 M glycine buffer pH 2.6 (E1-E3) and once by boiling in 4x Laemmli buffer (E4). Elutions were run in reducing conditions on SDS-PAGE gel and the presence of CD44 was evaluated by western blotting. We detected the presence of primary antibody in all of the elution fractions at ~50 kDa (Hc) and ~25 kDa (Lc) indicated by arrows. CD44 was detected in flowthrough lanes (FT) as a smear between 60 and 200 kDa. Treatment with tunicamycin reduced detection of CD44, and reduced the detected molecular weight due to a decrease in N-glycosylation. In the non-treated sample, CD44 is eluted in E4 at approximately 86 kDa (the anticipated weight) while in the tunicamycin treated samples CD44 is detected in E1-E3 as a single band around 70 kDa and as a smear in E4. The lack of glycosylation in the tunicamycin treated samples is shown by an overall decrease in higher molecular weight species. Tunicamycin treated samples appear to elute bound CD44 during glycine elutions, while non-treated CD44 is only immediately detectable in the boiled elution. This may indicate that properly glycosylated CD44 protein is the preferred ligand for the α-CD44 antibody.

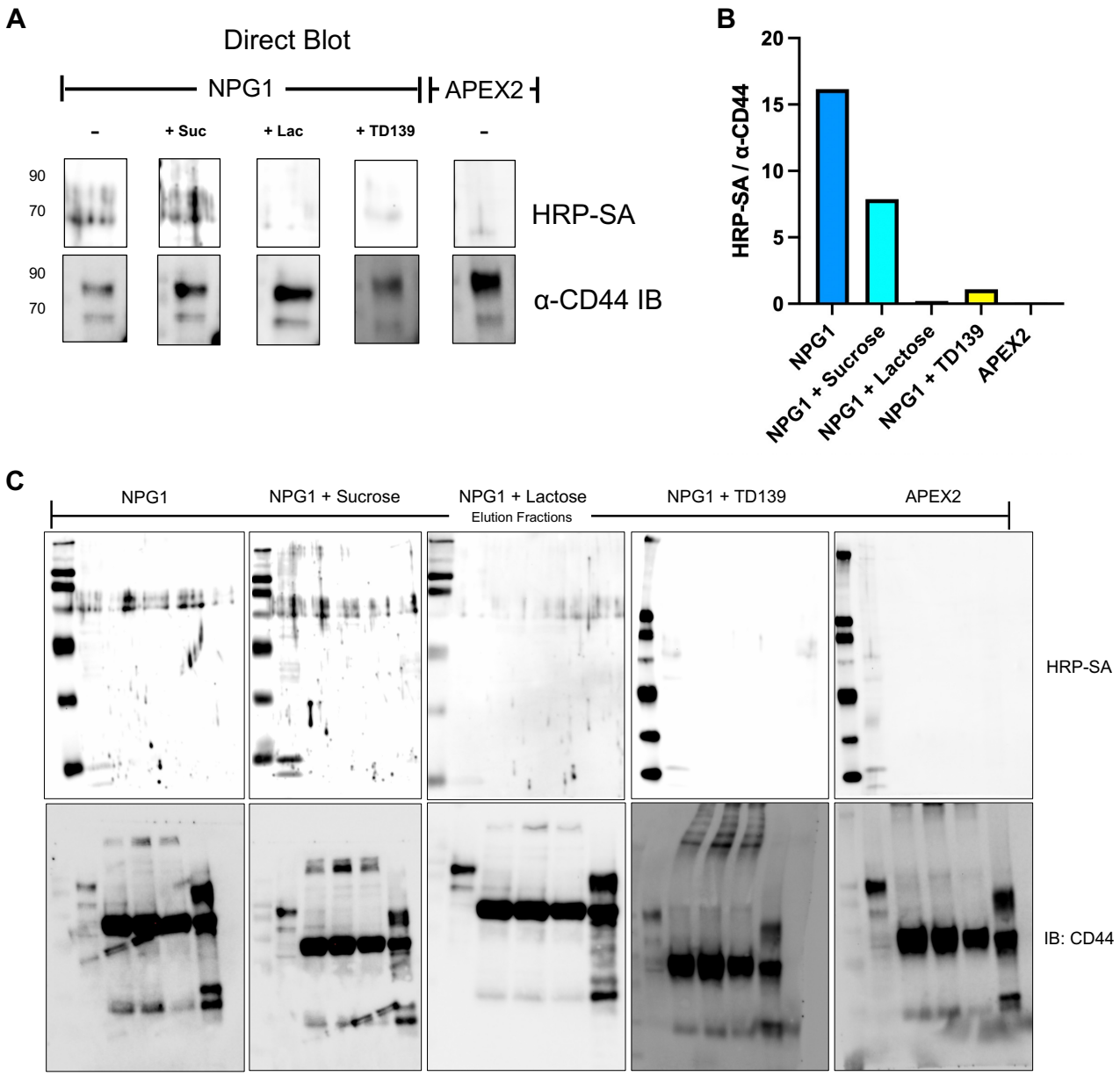
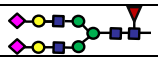
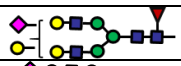
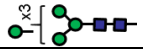
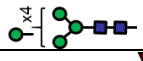
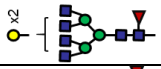
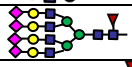
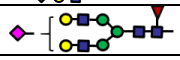
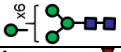
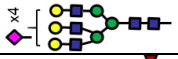
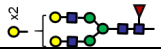
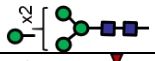
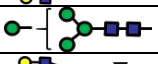
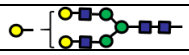


Figure S11. Glycosylation of CD44 mediates binding to NPG1. a) In a direct blotting assay of immunoprecipitated CD44 with either NPG1 or APEX2-LPETG, we observed that only NPG1 generates an appreciable signal. This signal is abrogated by excess lactose (100 mM) or TD139 (62 mM), but not by a non-competing glycan (sucrose, 100 mM). b) Densitometric quantification of direct blot signals showing that NPG1 but not APEX2 alone binds to immunoprecipitated CD44 in a glycan-mediated manner. c) Full Western blots of CD44 IP elutions as shown in Fig. S11a

Table S5. Identified glycans on C2C12 cell surface.

Ranking	Naming	Relative abundance (%)	Structure
1	FA2G2S2	22.76227	

2	FA2G3S1	8.370008	
3	A3G3S3	7.620745	
4	FA2	5.849314	
5	M6	4.089862	
6	M7	3.830458	
7	FA2G1X1	3.75568	
8	A3G3S1	3.283508	
9	A2G2S2	3.114756	
10	FA3G1	2.938131	
11	FA3G2	2.721369	
12	FA2G1	2.56369	
13	M8	2.54436	
14	FA4G1	2.429844	
15	FA4G2	2.295244	
16	FA3	2.267537	
17	FA4G4S4	2.094499	
18	FA2G2S1	1.883016	
19	M9	1.803146	
20	FA3G3X3	1.626349	
21	A3G3S4	1.614652	
22	FA2G3	1.26211	
23	FA2G2X2	1.110434	
24	FA1	1.082008	
25	FA2G4	1.051302	
26	M5	0.970853	
27	FA3G3	0.927792	
28	M4	0.611072	
29	FA4G4	0.42066	
30	FA1G1	0.316912	
31	A2	0.316284	
32	A3G1	0.269792	

33	A2G3	0.207874	

REFERENCES

1. Lam, S.S., et al., *Directed evolution of APEX2 for electron microscopy and proximity labeling.* Nature Methods, 2015(12): p. 51-54.
2. Cooper, D. and S.H. Barondes, *Evidence for Export of a Muscle Lectin from Cytosol to Extracellular Matrix and for a Novel Secretory Mechanism.* The Journal of Cell Biology, 1990. **110**: p. 1681-1691.
3. Harrison, L. and T.J. Wilson, *The 14kDa β-galactoside binding lectin in myoblast and myotube cultures: localization by confocal microscopy.* Journal of Cell Science, 1992. **101**: p. 635-646.