Supplementary Materials for

Resident microbes shape the vaginal epithelial glycan landscape

Kavita Agarwal[1,2,3,4,10], Biswa Choudhury[4], Lloyd S. Robinson[1,2,11], Sydney R. Morrill[1,2,3,4], Yasmine Bouchibiti[5,6], Daisy Chilin-Fuentes[7], Sara B. Rosenthal[7], Kathleen M. Fisch[3,7], Jeffrey F. Peipert[8], Carlito B. Lebrilla[5,6], Jenifer E. Allsworth[9], Amanda L. Lewis[1,2,3,4,*], Warren G. Lewis[1,2,3,4,*]

Affiliations:

[1]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110, United States of America.
[2]Center for Women's Infectious Disease Research, Washington University School of Medicine, St. Louis, MO 63110, United States of America.
[3]Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California San Diego (UCSD), La Jolla, CA 92093, United States of America.
[4]Glycobiology Research and Training Center, UCSD, La Jolla, CA 92093, United States of America.
[5]Department of Chemistry, University of California, Davis, Davis, CA 95616, United States of America.
[6]Department of Food Science and Technology, University of California, Davis, Davis, CA 95616, United States of America.
[7]Center for Computational Biology & Bioinformatics, UCSD, La Jolla, CA, USA
[8]Department of Obstetrics and Gynecology, Indiana University School of Medicine, Indianapolis, IN 46202, United States of America.
[9]Department of Biomedical and Health Informatics, University of Missouri, Kansas City School of Medicine, Kansas City, MO 64110, United States of America.

Current Affiliations:
[10]Trivedi School of Biosciences, Ashoka University, Sonipat, Haryana 131029, India
[11]Omniose LLC, St. Louis, MO 63110, United States of America.

***Co-corresponding authors:** wglewis@health.ucsd.edu, a1lewis@health.ucsd.edu

Contents:

Materials and Methods

Unless otherwise specified, all studies here utilized materials derived from specimens collected either in the CHOICE project or the UC Bank project as described in the main manuscript. Methods related to clinical samples (as described below) refer to materials derived from the CHOICE specimens, except the LC-MS analysis of *N*- and *O*-glycans.

.

Culture conditions for cell lines and bacterial strains

VK2/E6E7 cells (ATCC, CRL-2616) were cultured according to supplier instructions at 37 °C, in the presence of 5 % CO_2, using Keratinocyte-Serum Free Medium (GIBCO, 17005-042) supplemented with recombinant epidermal growth factor (0.1 ng / mL), bovine pituitary extract (0.05 mg / mL) and 0.4 mM calcium chloride.

E. coli Top 10 (cloning strain) was grown while shaking in lysogeny broth (LB) at 37 °C or as indicated, with antibiotic selection where required. *E. coli* Top10 (Invitrogen, C404003) was used for cloning putative sialidase-encoding genes into pET28a as described below. *Clear coli* BL21 DE3 (Biosearch Technologies, 60810-1) cells (detoxified LPS) were used for protein preparations used in cell-culture experiments. Clear coli cells were grown while shaking in either LB-Miller (SIGMA) medium supplemented with 0.5% glucose or in Terrific Broth (SIGMA). For plasmid (pET28a) maintenance in Top10 and *Clear coli* BL21 DE3 cells kanamycin was used at 50 µg/ml.

Epithelial cell processing

Epithelial cells obtained from vaginal swabs were processed individually for detection of glycocalyx by transmission electron microscopy (TEM) and sialic acid analysis by MAL-II staining. For lectin analyses which required larger numbers of cells, like quantification of PNA binding by flow cytometry, VECs were pooled from multiple swabs in the same group (Nugent 0-3 or Nugent 7-10) to have sufficient packed cell pellet volume (~50 – 100 µL). For analytical studies focused on *N*- and *O*-glycan characterization, frozen VEC pellets were sent to the University of California San Diego GlycoAnalytics core. An additional selection criterion was applied to these samples: swabs in the BV group with detectable high sialidase activity (relative sialidase activity >3 for most specimens) as estimated by sialidase assay on the first swab eluate, were pooled together. For each study condition, approximately 100-200 µL of packed cell volume was obtained by pooling VECs isolated from 5 or 10 vaginal swab eluates in the same group.

Fluorometric assay for sialidase activity

To measure sialidase activity, 50 µL of vaginal swab eluate was diluted 3-fold into 100 mM sodium acetate, pH 5.5, containing 200 µM 4-methylumbelliferyl-Neu5Ac (4-MU-Sialic acid) (Goldbio, M-520-1) in a 96 well round bottom black polypropylene plates and sealed with an optical clear film. Substrate hydrolysis was monitored using the fluorescence of 4-MU (Excitation 365 nm, Emission 440 nm) in a Tecan M200 plate reader every 2 min for 2 h at 37 °C.

Preparation of recombinant *Gardnerella* sialidases NanH2 and NanH3 in *E. coli*

Gardnerella NanH2 and NanH3 (truncated versions lacking the transmembrane segments of both proteins and the signal sequence of NanH2), along with the vector

control, were expressed in *E. coli* and purified by nickel chromatography as previously described (*63*). In experiments with recombinant *Gardnerella* sialidases NanH2/NanH3, and sialidase from *Arthrobacter ureafaciens* (*A.u.* sialidase), enzyme activities were normalized as described earlier (*63*). Similar amounts of enzyme activity were used to investigate their ability to cleave sialic acid from VECs.

Cloning of *Gardnerella* sialidase *nanH2* in pET28a plasmid

For experiments using VK2 cells with *Gardnerella* sialidase, the *nanH2* gene (accession number ATJH01000171) from *G. vaginalis* JCP8151B was used in this study. The truncated *nanH2* gene, lacking the predicted C-terminal transmembrane segment and N-terminal signal peptide, was amplified using pfuUltra (Agilent) polymerase, with the primer pairs 8151B NanH2 F NcoI: 5' TGT CTC CAT GGC TAC TTC AGA AAA ATC C 3' and 8151B NanH2 Trun his R Sal1: 5' ATA ATG TCG ACT GAA CTT CCC GTT GAA ACT TTA C 3'. The previously published pLR43 plasmid (*63*), where NcoI site at 1068 position in the full length *nanH2* gene was mutated for previous study, was used as a template for amplification. The truncated *nanH2* PCR product and the pET28a vector were digested using NcoI and Sal1 (New England Biolabs), followed by ligation using 5 U / µL T4 ligase (Invitrogen). *E. coli* Top10 cells were transformed with the ligation product and transformed colonies containing pET28a with truncated *nanH2* (**fig. S13**) were identified by sequencing.

ClearColi BL21(DE3) electrocompetent cells were transformed with either pET28a with JCP8151B *nanH2-His$_6$,* truncated (lacking transmembrane and signal sequence), or with the empty pET28a vector, by electroporation in 1mm cuvettes (Bio-Rad) at 1800 V (time constant 5.5 ms) following supplier protocol. After electroporation, cells were grown shaking at 250 rpm in the expression recovery medium (Biosearch Technologies) for 1 h at 37 ˚C and transformed colonies were identified by plating on LB with kanamycin.

Expression and purification of recombinant *Gardnerella* sialidase (rNanH2) from Clear coli cells (detoxified LPS)

All glassware used for preparation of media and reagents was baked at 200 °C for at least 2 h to destroy endotoxin. All buffers were prepared using endotoxin-free water (Cytiva). For expression and purification, an overnight culture of the transformed cells in LB (supplemented with 0.5 % glucose and kanamycin) was used to inoculate 450 mL Terrific Broth with kanamycin. The sub-cultured cells were then grown at 37 °C while shaking until they reached an OD of ~0.4, and were then induced with 1 mM IPTG for ~18 h at 18 °C. After induction, cells were pelleted and resuspended in lysis buffer (50 mM Tris-HCl, 300 mM sodium chloride (NaCl), 5 mM imidazole pH 8.0) containing EDTA-free protease inhibitor, 0.2 mg / mL lysozyme, and 1 mM PMSF. Cells were incubated with end-to-end mixing in lysis buffer for 1 h at room temperature, then sonicated on ice using a microtip at 10 % amplitude for 6 min 30 s, with pulse 15 s ON and 30 s OFF. The clarified lysates were incubated with 750 µL of TALON® Metal Affinity Resin cobalt beads (Takara) with end-to-end mixing at 4 °C for 1 h to allow for protein binding. Thereafter, beads were washed using 20 mL of wash buffer (50 mM Tris HCl and 300 mM NaCl) with increasing concentrations of imidazole from 20 mM to 40 mM (pH 8.0). Beads were incubated with end-to-end mixing at 4 °C for 10 min between each wash. Protein eluates were collected

in 500 μL fractions using 1.5 mL of elution buffer 1 (50 mM Tris-HCl, 300 mM NaCl and 300 mM imidazole pH 8.0) followed by 2.5 mL of elution buffer 2 (50 mM Tris-HCl, 300 mM NaCl and 650 mM imidazole pH 8.0).

Eluate fractions were evaluated for purity by SDS-PAGE, followed by staining with InstantBlue Coomassie Protein Stain (Abcam). A prominent band at the expected molecular mass of 100 kDa was observed in the eluate, that also expressed sialidase activity as confirmed using 4MU-Sia assay. Clear Coli cells transformed with empty pET28a vector were processed in parallel and eluates from this preparation were used as vector control in experiments with VK2 cells. Before use in cell culture, protein and vector control preparations were buffer exchanged using tissue culture grade (endotoxin-free) DPBS (no calcium, no magnesium) with PD Mini-Trap G-25 columns (Cytiva). Sialidase activity in the buffer exchanged fractions of rNanH2 preparation was estimated using 4MU-Sia assay, along with AUS. The rNanH2 concentration used in cell culture experiments had sialidase activity equivalent to 10 mU of AUS.

Visualization of glycocalyx on VECs by transmission electron microscopy (TEM)

VECs eluted from individual vaginal specimens of women with or without BV were washed with phosphate buffered saline (PBS) and processed for glycocalyx staining. For treatment of VECs with recombinant *Gardnerella* sialidase NanH2, cells obtained from individual specimens of women without BV were divided equally into two microcentrifuge tubes such that one half was treated with NanH2 in 300 µL swab elution buffer (SEB), and the other was mock treated with the same amount of buffer alone. Cells were kept at 37 °C on a rotor. After 1 h, cells were pelleted by centrifugation at 2000 g for 5 min and washed with 1 mL PBS twice. VECs were processed for glycocalyx detection using ruthenium red and osmium tetroxide (OsO_4) as described by Fassel and Edmiston (*75*). Briefly, VECs isolated from swab eluates were fixed and stained using 100 mM sodium cacodylate buffer (Polysciences Inc., Warrington, PA), pH 7.2, containing 0.15% ruthenium red, 2% paraformaldehyde and 2.5% glutaraldehyde for 1 h at RT. VECs were then washed with 100 mM cacodylate buffer containing 0.15% ruthenium red. Secondary fixation was performed with 1% OsO_4 and 0.15% ruthenium red in 100 mM cacodylate buffer for 1 h (Polysciences Inc.). Samples were then rinsed extensively in deionized water (dH2O) prior to en bloc staining with 1% aqueous uranyl acetate for 1 h (Ted Pella Inc., Redding, CA). Following several rinses in dH20, samples were dehydrated in a graded series of ethanol and embedded in Eponate 12 resin (Ted Pella Inc.). Sections of 95 nm were cut with a Leica Ultracut UCT ultramicrotome (Leica Microsystems Inc., Bannockburn, IL), stained with uranyl acetate and lead citrate, and viewed on a JEOL 1200 EX transmission electron microscope (JEOL USA Inc., Peabody, MA) equipped with an AMT 8 megapixel digital camera and AMT Image Capture Engine V602 software (Advanced Microscopy Techniques, Woburn, MA). For glycocalyx scoring, the microscopist and scorers were blinded to the experimental treatment, as well as the BV status of the sample. 10 images of the epithelial cell membrane were acquired from each specimen at 25,000 X magnification. Images were labelled randomly to facilitate the blinded glycocalyx scoring. Images were scored between zero to five based on the abundance of glycocalyx, with zero being least abundant or nearly absent and five being

highly abundant. Data from three independent scorers was combined to compute the average glycocalyx abundance score for each image.

Lectin staining of vaginal epithelial cells

Cells were isolated from vaginal swab eluates of women with and without BV as described above. Cells were derived from individual women whenever possible; when cell numbers were insufficient for flow cytometry, cells from 2-5 individuals were pooled for analysis. Cells were washed three times with PBS containing 0.08% sodium azide (wash buffer). Prior to lectin staining, cell pellets for each sample (with ~50 µL packed cell volume) was re-suspended in 1 mL wash buffer, divided equally into 2 tubes and centrifuged at 2000 x g for 2 min. Supernatants were removed, pellets were resuspended in 300 µL of SEB (untreated controls) or SEB containing 15 mU of exogenous sialidase from *Arthrobacter ureafaciens* (*A.u.* sialidase) (EY Labs, EC-32118-5). Cells were incubated at 37 °C on a rotor to avoid clumping. After 1 h, cells were washed with 1 mL of lectin staining buffer (LSB) (10 mM HEPES, pH 7.5, 0.15 M sodium chloride and 0.1 mM calcium chloride). Cells were stained with peanut agglutinin (PNA) by incubating with 10 µg/mL of PNA-rhodamine (PNA-Rh) (Vector Lab, RL-1072) in 300 µL LSB in an ice bucket with intermittent mixing (every 10-15 min). After approximately 1 h, PNA stained cells were washed with LSB, re-suspended in 150 µL LSB and analyzed by flow cytometry. About 10,000 events were acquired for the same gated population for each sample (see **fig. S7A**). Untreated cells were used to assess auto fluorescence; PNA-Rh binding was analyzed in the FL-2 channel on a BD FACSCalibur flow cytometer. Exposed galactose was calculated as a percentage of PNA-binding mean fluorescence intensity (MFI) observed to untreated cells compared to the PNA-binding MFI to VECs from the same sample pretreated with *A.u.* sialidase.

For MAL-II staining, cells from individual specimens were incubated with 5 µg/mL of MAL-II biotin (Vector Laboratories, B-1265) in 300 µL LSB in an ice bucket with intermittent mixing (every 10-15 min). After ~1 h, cells were washed with an avidin staining buffer (ASB) (1% BSA, 0.05% NaN_3 in PBS). To detect biotin, cells were then stained with 2 µg/mL Neutravidin-OregonGreen 488 (Invitrogen, A6374) in 300 µL ASB for ~45 min at 4 °C degrees in an ice bucket. Cells stained with secondary only were included as control. Finally, the MAL-II stained cells and controls were washed with PBS and processed further for imaging.

For confocal microscopy, lectin-stained cells were re-suspended in PBS containing 600 nM DAPI (Invitrogen) and incubated in an ice bucket for 15 min. 10 µL of the cell suspension was then transferred to a microscope slide, covered with a cover slip and sealed with a transparent nail polish. Images were captured from at least three independent fields for each sample using LSM 880 confocal laser scanning microscope (Carl Zeiss) using 40 X objective lens with immersion oil.

Detection of Lewis X (CD15) on VECs

Cells were isolated from vaginal swab eluates of women without BV as described above. Cells were washed three times with PBS containing 0.08% sodium azide (wash buffer). Prior to staining with anti-CD15, that recognizes Lewis X antigen, the cell pellet for each

sample was resuspended in 1 mL wash buffer, divided equally into 2 tubes and centrifuged at 2000 x g for 2 min. Supernatants were removed, and the pellet was resuspended in 300 µL of SEB or SEB containing 15 mU of *A.u.* sialidase (EY Labs). Cells were incubated in this buffer at 37 °C on a rotor to avoid clumping. After 1 hour, cells were washed with 1 mL of LSB. To block non-specific binding, cells were incubated with 5 µg/mL of Fc Block (BD, 564220) in 600 µL of LSB for 15 min on a rotor at room temperature. Next, each specimen was divided equally into three tubes, 200 µL each, for staining with either 5.6 µg/mL secondary antibody alone (anti-mouse IgM-AlexaFluor488, Jackson ImmunoResearch) or 5 µg/mL of isotype control (mouse IgM, BD 555581) or 5 µg/mL of primary antibody (anti-Lewis X (Human CD15), BD 555400, clone HI98 (also known as HIM1)) followed by the secondary antibody. Cells stained with anti-Lewis X or isotype control were incubated in an ice bucket for 1 h with intermittent mixing (every 10-15 min). After washing the cells with 1 mL of LSB, a second staining step was performed in a similar manner by adding the secondary antibody to all the samples. After 40 min cells were washed two times with LSB, re-suspended in 150 µL LSB and analyzed by flow cytometry. Unstained cells were processed similarly in parallel without any antibody staining. About 10,000 events were acquired for each sample; anti-Lewis X binding was analyzed in the FL-1 channel on a BD FACSCalibur flow cytometer. For confocal microscopy, samples were processed and imaged as described above.

Preparation of crude cell lysate for glycan extraction

All analyses of VEC *N*- and *O*-glycans were performed at GlycoAnalytics Core at the University of California San Diego, California, USA. VECs isolated from vaginal swab (100 µL of packed cell pellet) were re-suspended in 300 µL of ultrapure water (Invitrogen) with 5 µL EDTA free protease inhibitor (EMD-Millipore Corp.) and sonicated in an ultrasonic water bath (Fisher Scientific Sonicator Model FS-9) until it formed a uniform suspension in water. Protein was quantified using BCA assay and further extraction of glycans was performed using a defined amount of protein as described in the methods below.

Isolation and purification of *N*-glycans

N-linked glycans were isolated from glycoproteins in the crude cell lysate by overnight incubation with glycerol free recombinant PNGase-F (NEB, P0709S) at 37 °C. Briefly, cell extract for 500 µg of protein was mixed with a denaturing buffer (NEB kit) and boiled at 100 °C for 15 min. Proteins were solubilized by incubation with NP-40 for 1h at room temp, with intermittent vortexing every 15 min. *N*-glycans were extracted using PNGase-F and purified using a tandemly connected Sep-Pak C18 and porous graphitized carbon (PGC) cartridge equilibrated in water. PGC bound glycans were eluted with 30% acetonitrile containing 0.1% trifluoroacetic acid (TFA) in water and then lyophilized. The purified pool of glycans were processed further as required for each analysis.

Isolation and purification of *O*-glycans

To release *O*-linked glycans by reductive β-elimination, crude cell lysate for 500 µg of protein was incubated with 0.05 M NaOH / 1 M $NaBH_4$ at 45 °C for 16 h with constant stirring. Samples were then neutralized using ice-cold 30% aqueous acetic acid, followed by removal of sodium salts using cation exchange resin (Dowex 50Wx8, H^+ Form, Bio-Rad 142-1451). The samples were co-evaporated using a 9:1 (v/v) methanol/acetic acid

mixture and with absolute methanol to remove borate. The extracted *O*-glycans were purified using Sep-Pak C18 column, where the flow through from Sep-Pak C18 was lyophilized and used for further analysis (*118*). Additionally, for linkage analysis by gas chromatography-mass spectrometry (GC-MS), charged *O*-linked glycans were separated (described below) from neutral *O*-glycans to reduce interference from contaminating neutral saccharides in the vaginal specimens.

Sialic acid analysis on *N*- and *O*-glycans by reverse phase-liquid chromatography

For sialic acid analysis, glycans derived from 25 µg of protein extract were heated to 80 °C in 2 M acetic acid for 3 h. The released sialic acids were collected by ultrafiltration through a 3,000 MWCO filter and derivatized with 1,2-diamino-4,5-methylenedioxybenzene (DMB) (Sigma Aldrich), as described earlier (*119*). The DMB-derivatized sialic acids were eluted isocratic with acetonitrile (8%) and methanol (7%) in Milli-Q® water and analyzed either by HPLC using Acclaim 120 C18, 5 µM column (4.6 x 250 mm, Thermo-Dionex) at a flow rate of 0.9 mL/min over 30 min or using an Acquity™ UPLC system, with fluorescence detector set at λ_{ex} 373 nm, λ_{em} 448 nm. For UPLC, samples were eluted over 10 min using Acquity UPLC® BEH C18, 1.7 µM column (2.1 x 50 mm, Waters) at a flow rate of 0.2 mL/min. The DMB-derivatized sialic acids were identified and quantified by comparing elution times and peak areas to known standards using Chromeleon® software. Data shown is from two independent experiments that were conducted ~three years apart. For the first experiment, VECs from 10 specimens were combined to obtain each pool as a greater biomass was required to complete all the analytical studies (shown in **Fig. 2-4**) on the same set of samples. The second experiment was conducted to specifically validate the preliminary findings of this experiment using a larger set of samples, for which VECs from 5 specimens were combined to obtain each pool.

2-Aminobenzamide (2-AB) labeling of purified *N*-glycans

The purified *N*-glycans were labeled with 2-AB as described earlier (*120*). Briefly, samples were dissolved in 10 µL solution of 0.44 M 2-AB in acetic acid-DMSO mixture (35:65 v/v) containing 1 M sodium cyanoborohydride. Samples were then incubated at 65 °C for 2.5 h. 2-AB labeled glycans were purified using GlycoClean S cartridge (GLYKO) following the manufacturer's protocol.

2-AB profiling with High-Performance Anion-Exchange Chromatography (HPAEC)-Fluorescence detection

Profiling of 2-AB labeled glycans was obtained using a Dionex CarboPac® PA1 (4 X 250 mm) anion exchange column along with a guard column (4 X 50 mm) at flow rate of 1 mL/min. Glycans were separated in 100 mM sodium hydroxide with a sodium acetate gradient of 0-250 mM in 0-75 min. Data was collected using the Dionex ICS-3000 HPLC system with Ultimate 3000 fluorescence detector (Dionex) set at λ_{ex} 330 nm at λ_{em} 420 nm, with sensitivity 7. The data was processed using Chromeleon™ software (Thermo Scientific). Glycans from RNAse B and bovine fetuin were processed in parallel and labeled as known standards with high mannose and complex type *N*-glycans.

Separation of *O*-glycans by charge

Anionic glycans were captured using liquid chromatography-Amino (LC-NH_2) cartridges (Supelco, Part no. 504483, 1 mL). Briefly, total *O*-glycans (reconstituted in water after purification as described above) were loaded on the cartridge and the column was washed with 3 mL of ultrapure water (Invitrogen) which was collected as flow-through. Anionic species were eluted with 2 mL of 100 mM ammonium acetate and lyophilized.

Per-*O*-methylation of glycans

Purified *O*-glycans, dried in glass reaction tubes, were dissolved in 350 µL of anhydrous DMSO by stirring for 1 h. To this solution was added a freshly prepared slurry (50 µL) of powdered NaOH in anhydrous DMSO, followed by immediate addition of 500 µL of methyl iodide. This reaction mixture was stirred vigorously at room temperature for 40 min. The reaction was quenched by adding 1 mL of ice-cold water followed by extraction of methylated glycans using chloroform. Finally, the chloroform was dried using a stream of nitrogen.

MALDI-TOF profiling of *O*-glycans

Permethylated glycans were dissolved in absolute methanol and mixed 1:1 with super-DHB matrix solution and spotted on the matrix-assisted laser desorption/ionization (MALDI) plate. The samples were air-dried and analyzed in positive-reflectron mode on a MALDI-qTOF mass spectrometer (Applied Biosystems QSTAR) (*121*). Structures were manually annotated and verified using the glycobioinformatics tool GlycoWorkBench (*122, 123*). The proposed assignments for the selected peaks were based on the knowledge of mammalian glycan biosynthesis pathways. The proposed structures were confirmed using MS/MS and GC-MS linkage analysis as described below.

MS/MS analysis of *O*-glycans

All spectra were acquired in the positive ionization mode over a range of 250-2000 amu on LTQ-XL Orbitrap Discovery Electrospray Ionization Mass Spectrometer (Thermo Scientific). Permethylated glycans were dissolved in absolute methanol, followed by addition of 1mM sodium hydroxide (NaOH), and direct injection into the mass spectrometer with a constant flow rate of 5 mL/min. The ion spray voltage was optimized at 3.7 kV with a sheath gas flow rate 12 mL/min. All MS and MS(n) spectra were collected in positive ionization mode with resolution of 30000. The other instrument parameters for positive-ion detection were: capillary temperature 145 °C; capillary voltage 47 V. Tandem mass spectrometry (MSn) was performed via collision-induced dissociation (CID) on selected parent ions with collision energy of 15-20%. Fragment ions are described based on the nomenclature proposed by Domon and Costello (*70*).

Glycan linkage analysis by GC-MS

To determine linkage positions of vaginal epithelial *O*-glycans, permethylated glycans (permethylation described above) were hydrolyzed with 4 N TFA at 100 °C for 4 h, followed by removal of acid using nitrogen flush. The sample was then reduced using ~5 mg sodium borodeuteride ($NaBD_4$) solution in 1 M ammonium hydroxide, and acetylated using pyridine and acetic anhydride (1:1 v/v) mixture at 100 °C for 1 h (*124*). Partially methylated acetylated alditol (PMAA) derivatives were analyzed by GC-MS (*125,126*) using Restek-5ms capillary column (30 m x 0.25 mm x 0.25 µm) on an Agilent

Technologies 7820A GC System 5975 Series Gas Mass Selective Detector (GC-MS). Identifications are achieved by using a combination of retention times and the mass fragmentation pattern in EI-MS mode.

Lectin analysis of VECs treated with *Gardnerella* sialidases

Prior to lectin staining, the cell pellet for each pooled sample (with ~100 µL of packed cell volume) was re-suspended in 1 mL wash buffer, divided equally into 4 tubes and centrifuged at 2000 x g for 2 min. Supernatants were removed, and the pellet was resuspended in 300 µL of SEB or SEB containing either similar amounts of enzyme activity of NanH2, NanH3, *A.u.* sialidase or similar volume of preparation from vector control. Cells were incubated in this buffer at 37 °C on a rotor to avoid clumping. After 1 h, supernatants were collected for sialic acid analysis by DMB-HPLC, and cells were washed with 1 mL of LSB. For staining with PNA, cells were incubated with 5 µg/mL of PNA-AlexaFluor647 (Invitrogen, L32460)) in 300 µL LSB in an ice bucket with intermittent mixing (every 10-15 min). After approximately 1h, PNA stained cells were washed with LSB, re-suspended in 150 µL LSB and analyzed by flow cytometry. About 10,000 events were acquired for each sample; PNA-AF647 binding was analyzed in the FL-4 channel on a BD FACSCalibur flow cytometer. For confocal microscopy, lectin-stained cells were processed for imaging as described above.

Lectin analysis of protein extracts derived from vaginal epithelial cells

Cells were isolated from individual vaginal swab eluates of women with and without BV as described above. Isolated VECs were frozen overnight at -80 °C and lyophilized. The freeze-dried samples were re-suspended in a lysis buffer containing 1:1 PBS / milliQ water with 2 % SDS and EDTA-free protease inhibitor cocktail (Roche, 11836170001), and incubated on ice for 20-25 min. Lysates were clarified by centrifugation at 15000 x g for 30 min at 4 °C. Protein was estimated using BCA assay. Equal amount of protein extracts (5 µg) derived from the VECs of women without or with BV were treated with sialidase (AUS) or PNGase F or with both sequentially (first AUS, followed by PNGase F). For sialidase treatment, 5 mU of enzyme (AUS from EY labs) per 20 µg protein was added to the protein extract and incubated for 1 h at 37 °C. For PNGase F, *N*-glycans were removed under denaturing conditions following supplier protocol (NEB). Equal amount of protein extracts from enzyme treated or untreated samples were resolved on a 4-15% Tris-glycine gel (Bio-Rad) under reducing and denaturing conditions, followed by transfer to a PVDF membrane in Towbin's buffer for 3 h (250 mA of constant current) at 4 °C. For lectin blotting, membranes were processed in glass chambers in lectin staining buffer 2 (LSB2-T: 10 mM HEPES, pH 7.5, 0.15 M sodium chloride, 0.1 mM calcium chloride, 0.01 mM manganese chloride, and 0.1% Tween-20). First, membranes were blocked with 2% fish gelatin (SIGMA) in LSB2-T at room temperature (RT) for 1 h. After one wash with LSB2-T, biotinylated lectin i.e. MAL-II (0.5 µg / mL), SNA (0.5 µg / mL, Vector Labs, B-1305), or ECA (1.0 µg / mL, Vector Labs, B-1145) was added directly to 10mL of LSB2-T in the glass chamber and incubated for 1 h at RT. After 5 washes with LSB2-T, membranes were incubated with HRP-conjugated streptavidin (1:50,000 dilution in 10 mL LSB2-T) for 1 h at RT. After 6 washes, lectin binding was detected by chemiluminescence. Images were acquired using Bio-Rad ChemiDoc.

LC-MS analysis of *N*- and *O*-glycans derived from VECs (UC Bank project)

VECs were derived from specimens collected in the UC Bank study. BV status was determined by clinical Amsel's criteria. A vaginal swab from a woman without BV and with BV was removed from the -80 °C freezer and placed in an ice bucket. In a biosafety cabinet, swabs were transferred to a cryotube on ice, cut ~3.5-4.0 cm, and immersed in 1 mL of 100 mM ammonium acetate buffer (pH 5.5) to allow for elution. After 1 h on ice, cryotubes containing swabs were shaken at 500 rpm for 10 min at room temperature, using a IKA MS 3 digital shaker. The individual swab eluates were transferred from the cryovials into microcentrifuge tubes and placed back on ice. Cells were isolated from the swab eluates by centrifugation at 1000 x g for 5 min at 4 °C and the cell pellets were freeze dried (lyophilized) before further processing.

Glycans were isolated from the freeze-dried cell pellets (VECs), using all the material derived from each individual swab, and characterized using methods as described earlier (*67*), with some modifications. The freeze-dried cell pellet was re-suspended in lysis buffer (0.1 M HEPES, 0.25 M sucrose, 1 mL protease inhibitor cocktail (Calbiochem) and lysed through probe ultrasonication (Q700, QSonica). The cell lysates were centrifuged at 12000 x g for 10 min at 4 °C to remove cell debris and the nucleus associated faction. The supernatant was further ultracentrifuged at 200,000 x g for 45 min at 4 °C to isolate the cell membrane. Proteins from the cell membrane were quantitated using BCA assay, prior to denaturation. For *N*-glycan release, cell membrane proteins were denatured and incubated with peptide-N-glycosidase F (PNGase F, New England BioLabs), alongside a RNase B and serum standard, in boiling water bath for 18 h at 37 °C. The samples were ultracentrifuged at 200,000 x g for 45 min at 4°C to isolate the protein pellet and the released *N*-glycans. The supernatant containing *N*-glycans was subjected to solid phase extraction (SPE) clean up using porous graphitized carbon (PGC, Glygen) prior to analysis on an Agilent 6520 ChiP-QTOF-MS. The protein pellet after *N*-glycan release was used for extraction of *O*-glycans through reductive beta-elimination. The released *O*-glycans were enriched on PGC SPE cartridges and further cleaned up with HILIC SPE (Hilicon). The enriched *N*-glycan and *O*-glycan samples were reconstituted in water and injected into the nano-LC-MS/MS system (Thermo Fisher Scientific). Isomeric separation was performed using an Agilent PGC-Chip II (40 nL enrichment and 43 mm × 75 µm analytical column, particle size 5 µm) with a binary solvent system, and glycans were identified using in-house libraries for *N*- and *O*-glycan compounds utilizing MassHunter Qualitative Analysis Software B.07.00 (Agilent Technologies) as described earlier (*67*). Identified structures and compositions were confirmed through the tandem MS spectra. The compositions of the structures are reported based on the number of each residue type using the Hex_HexNAc_Fuc_NeuAc_0Sulfation format, where the Hex corresponds to the number of hexoses (i.e. mannose, galactose), HexNAc corresponds to the number of *N*-acetylhexosamines (i.e. *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc)), Fuc corresponds to the number of fucose residues, and NeuAc corresponds to the number of the neuraminic acid (i.e. sialic acid). The glycans were subsequently classified using an in-house classification system: high-mannose (HexaHexNAcb, where b = 2 denotes the chitobiose core and "a" refers solely to mannose residues), undecorated (HexaHexNAcb, where "a" refers to both mannose and galactose residues), neutral

(HexaHexNAcb, where “a” refers to only galactose residues in *O*-glycans), fucosylated (HexaHexNAcbFucc, where c ≥ 1), sialylated (HexaHexNAcbNeuAcd, where d ≥ 1), or sialofucosylated (HexaHexNAcbFuccNeuAcd, where both c and d ≥ 1). Relative abundance of each structure was determined through the peak areas extracted from the ion chromatograms. Relative abundances of glycan groups were calculated by adding the relative abundance of each individual glycan composition belonging to a specific glycan group. The glycomic profiles of No BV and BV VECs were separated by composition and glycan structure type to illustrate the differences in glycosylation.

VK2 experiments using recombinant NanH2 for RNA-seq

Cells were seeded in a 24-well tissue culture plate at a density of 10^5 cells / well, in 400 µL of complete cell growth medium, and cultured for ~48 h until they were about 80-90% confluent. Cells were then treated with enzyme or vector control for 1 h or 2 h prior to proceeding for RNA isolation. Specifically, spent cell culture media was replaced with fresh growth medium containing either 10 mU of truncated rNanH2 or vector control that were pre-mixed with 25 µg / mL of polymyxin B sulfate (SIGMA) at least 15 min prior to the addition to the cells. After 2 h from the first enzyme treatment, and 1 h from the second enzyme treatment in separate wells, cell culture supernatant was collected from each well and saved at -20 ˚C for analysis of released sialic acid by DMB-HPLC as described above. The cell monolayers were washed with sterile endotoxin-free PBS and RNA was isolated using RNeasy Micro Kit (Qiagen) following supplier instructions.

Library Preparation and Sequencing

Libraries were prepared using total RNA isolated from each well of VK2 cells treated with rNanH2 or the empty-vector preparation. RNA quality and integrity was determined using Agilent Tapestation 4200. All samples had RNA Integrity Number (RIN) greater than 9.5 and were used to generate RNA sequencing libraries using the Illumina® Stranded mRNA Prep (Illumina, San Diego, CA). Samples were processed following manufacturer’s instructions, starting with 500 ng of total RNA. Resulting libraries were multiplexed and sequenced with 100 basepair (bp) Paired End reads (PE100) to a depth of approximately 25 million reads per sample on an Illumina NovaSeq 6000. Samples were demultiplexed using bcl2fastq Conversion Software (Illumina, San Diego, CA).

RNA-Seq Data Quality Control and Processing

Quality control of the raw fastq files was performed using the software tool FastQC (*127*) v0.11.8. Sequencing reads were trimmed with Trimmomatic (*128*) v0.38 and aligned to the human genome (GRCh38.p13 (*129*)) using the STAR aligner (*130*) v2.5.1a. Read quantification was performed with RSEM (*131*) v1.3.0 and Gencode release 43 annotation (*132*). Each of the twelve samples (4 replicates for each condition) had >10 million unique reads and >90 % of the reads aligned uniquely to a gene for each sample. The R BioConductor packages edgeR (*133*) and limma (*134*) were used to implement the limma-voom (*135*) method for differential expression analysis. In brief, lowly expressed genes—those not having counts per million (cpm) ≥ 1 in at least 4 of the samples—were filtered out and then trimmed mean of M-values (TMM) (*136*) normalization was applied. The experimental design was modeled upon treatment timepoints. The voom method was employed to model the mean-variance relationship in the log-cpm, after which lmFit was

used to fit per-gene linear models and empirical Bayes moderation was applied with the eBayes function. Significance was defined by using an adjusted p-value cut-off of 0.05 after multiple testing correction (*137*) using a moderated t-statistic in limma. Functional enrichment of the differentially expressed genes was performed using fGSEA (*138*).

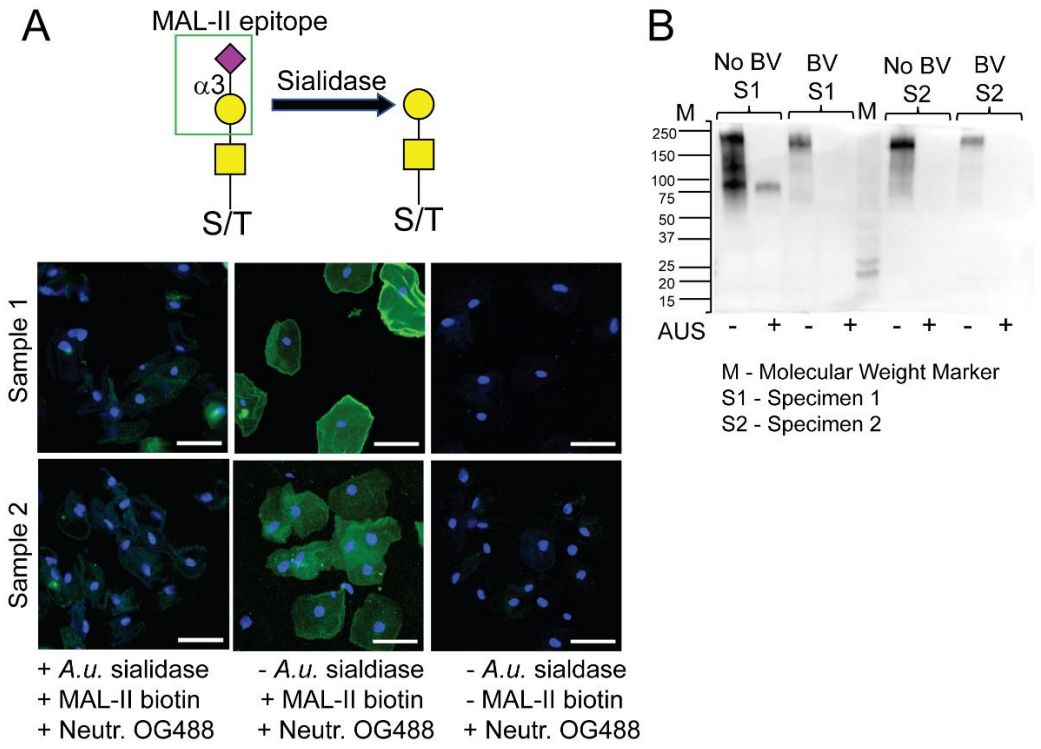

Fig. S1. Analysis of α2-3 linked sialic acids in VEC glycans using MAL-II lectin. (**A**) Confocal images of VECs from women without BV (Nugent 0-3) that were stained either with biotinylated-MAL-II lectin followed by neutravidin-OG488 (Neutr. OG488) or the neutravidin-OG488 alone (right). VECs pre-treated with exogenous sialidase (*A.u.* sialidase) were included to validate selective binding of MAL-II to sialic acids. Neutravidin-OG488 shows minimal non-specific binding to VECs in the absence of MAL-II lectin. Images shown are representative of multiple fields of view. Scale bars = 50 µm. (**B**) Analysis of α2-3 linked sialic acids in protein extracts of VECs derived from two individual women without BV (Nugent 0-3) and with BV (Nugent 7-10) by MAL-II lectin blotting. Equal amount of protein was loaded from each sample. Protein extracts treated with exogenous sialidase from *Arthrobacter ureafaciens* (AUS) were included as a control.

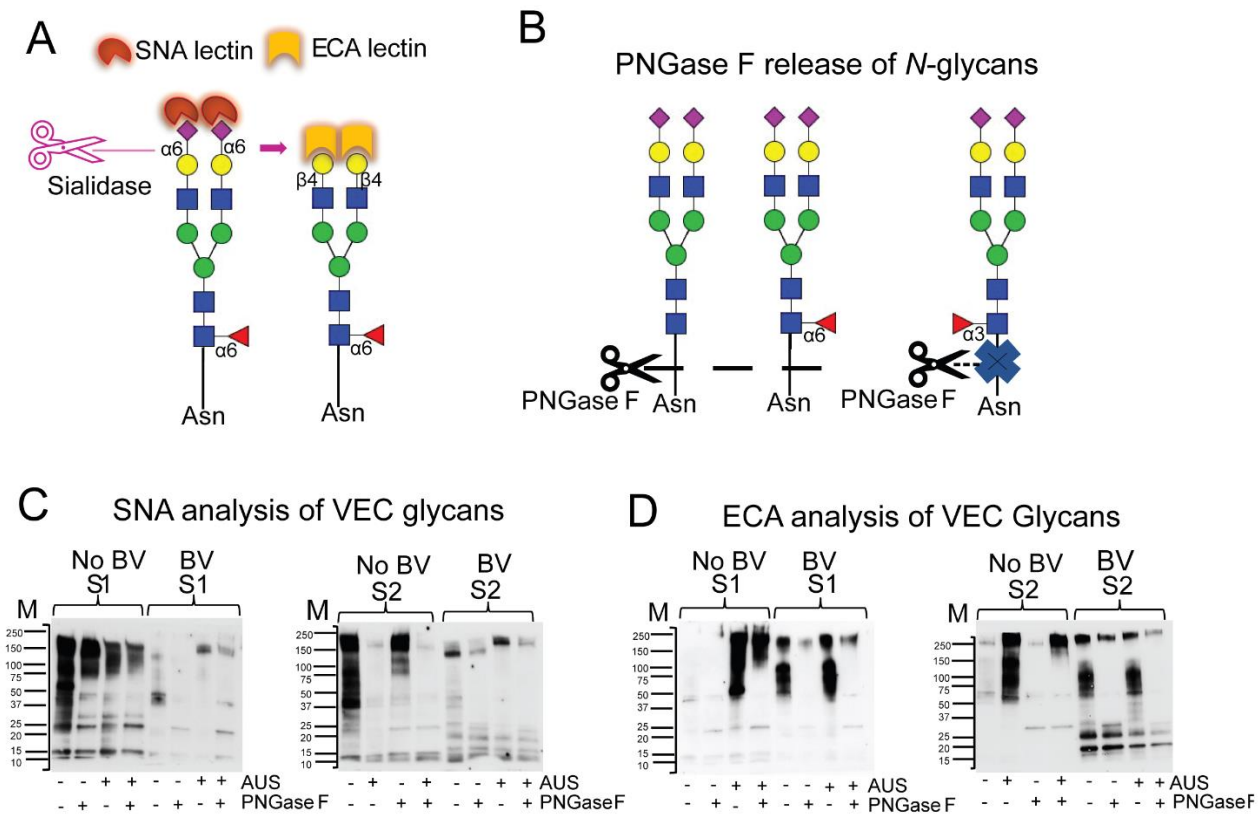

Fig. S2. Analysis of the sialylation of VEC *N*-glycans using SNA and ECA lectin. (**A**) Schematic showing SNA lectin binding to α2-6 linked sialic acids and ECA lectin binding to the Galβ1-4GlcNAc epitope that is exposed upon removal of sialic acid. Structures of different types of *N*-glycans are depicted following the NCBI Symbol Nomenclature for Glycans (yellow circle, galactose; green circle, mannose; blue square, *N*-acetylglucosamine; purple diamond, sialic acid; red triangle, fucose). (**B**) Schematic shows PNGase F releases non-fucoslayted and fucosylated *N*-glycans with α1-6 linked fucose, but not with α1-3 linked fucose. (**C** and **D**) Protein extracts of VECs derived from two individual women without BV (Nugent 0-3) and with BV (Nugent 7-10) were resolved on 4-15% Tris-glycine gel under reducing conditions and then transferred to PVDF membranes, followed by SNA lectin blotting (**C**) and ECA lectin blotting (**D**). Equal amount of protein was loaded from each sample. Protein extracts treated with exogenous sialidase from *Arthrobacter ureafaciens* (AUS) or PNGase F were included as controls.

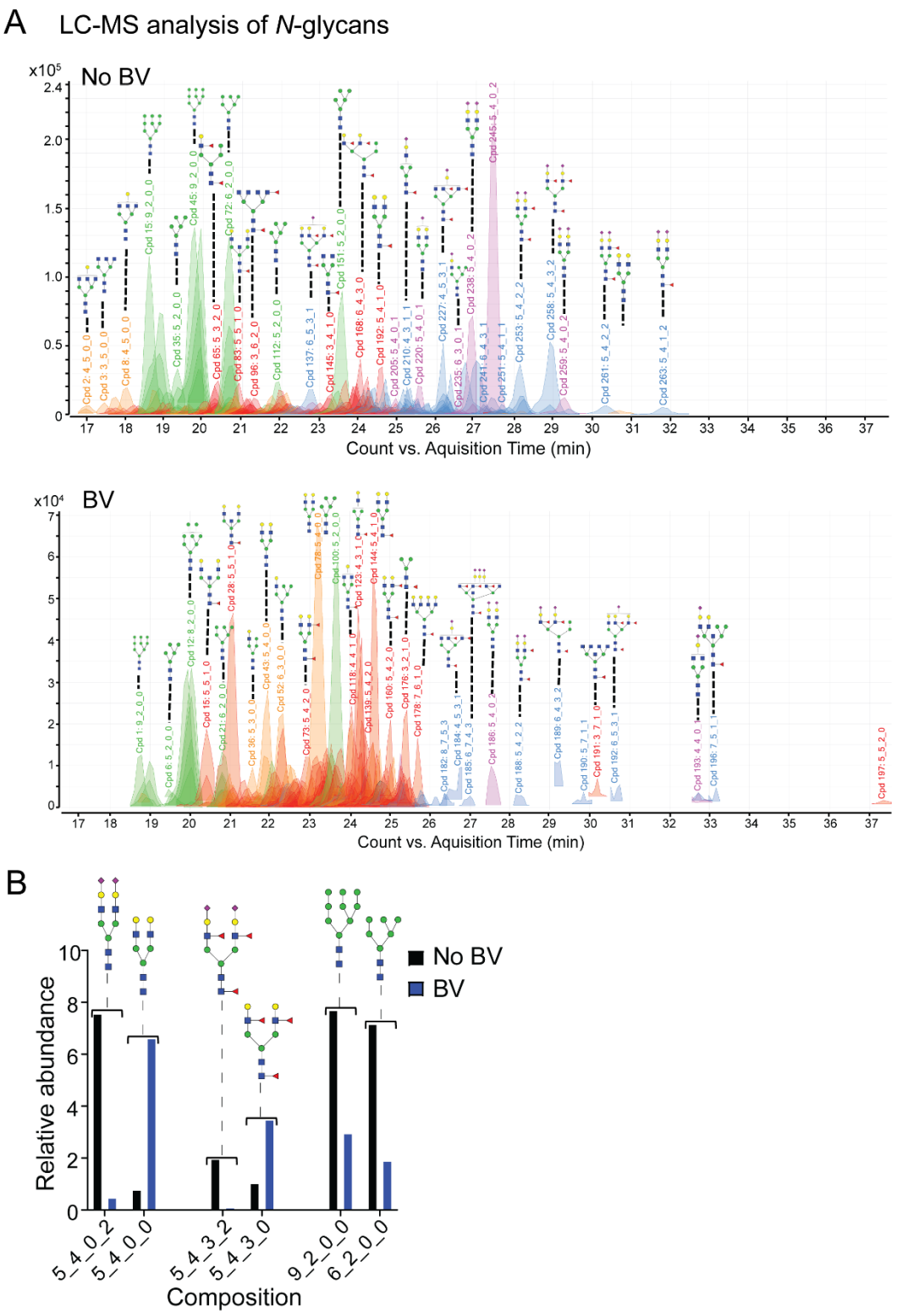

Fig. S3. Mass spectrometric profiling of the *N*-glycans derived from human VECs. (**A**) Chromatograms show structures and composition of VEC *N*-glycans derived from a woman without BV (upper panel) and another with BV (lower panel). Compound (Cpd) nomenclature reports the number of each type of residue in the glycan composition as [Hex]_[HexNAc]_[Fuc]_[Neu5Ac], where Hex refers to

the number of hexose (i.e. mannose or galactose), HexNAc refers to the number of *N*-acetylhexosamine (i.e. *N*-acetylglucosamine), Fuc refers to the number of fucose, and NeuAc refers to the number of neuraminic acid (i.e. sialic acid) residues. Glycans are classified as high mannose (green), fucosylated (red), sialylated (purple), sialofucosylated (blue), or undecorated (orange) as described in the methods. (**B**) Relative abundance of selected *N*-glycan structures identified in **A** are shown in the bar graph. Structures of different types of *N*-glycans are depicted following the NCBI Symbol Nomenclature for Glycans (yellow circle, galactose; green circle, mannose; blue square, *N*-acetylglucosamine; purple diamond, sialic acid; red triangle, fucose). A complete list of structures and composition, with their relative abundance, for these specimens is provided as a supplemental file.

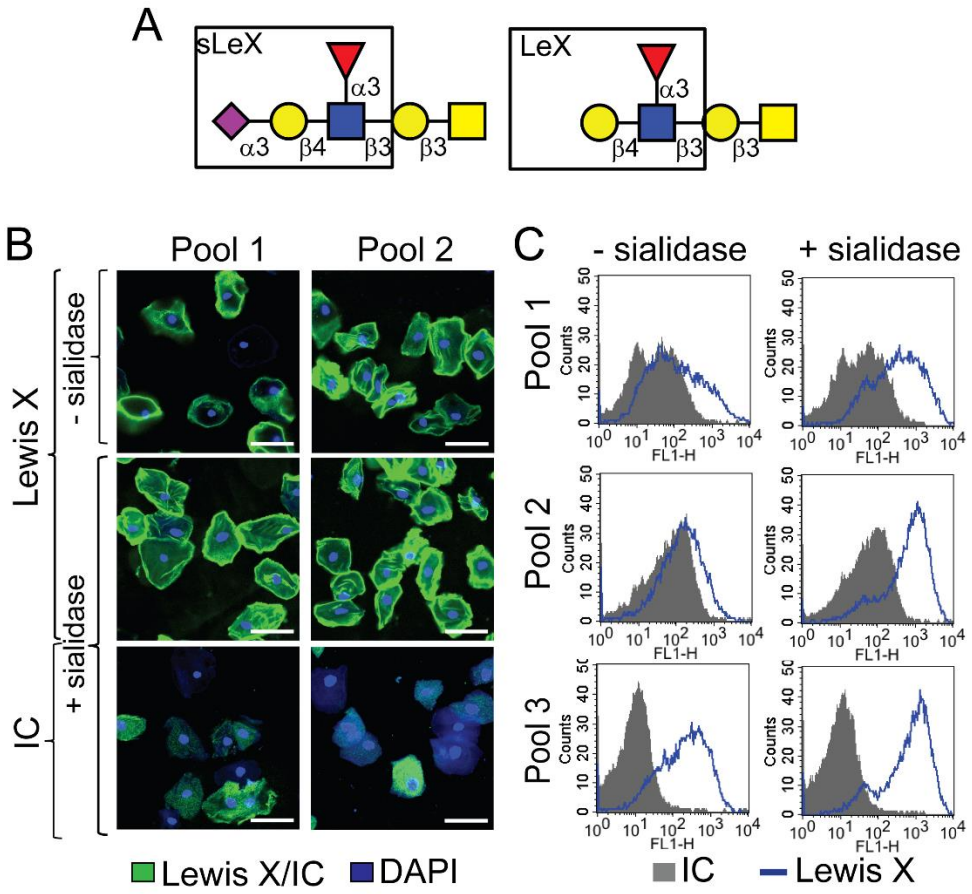

Fig. S4. Analysis of Lewis X (CD15) antigen on vaginal epithelial cells. Cells from the same pool of VECs, isolated from vaginal specimens of women without BV (Nugent 0-3), were treated with exogenous sialidase (*A.u.* sialidase) or buffer, followed by staining with anti-Lewis X or isotype control. (**A**) Structures of sialyl Lewis X (sLeX) and Lewis X (LeX) antigens. (**B**) Representative confocal images of two separate pools of VECs. Green = Lewis X or Isotype Control (IC), Blue = DAPI. Scale bars = 50 µm. (**C**) Histograms show that binding of the Lewis X antibody to VECs is enhanced upon treatment with exogenous sialidase (*A.u*. sialidase). Grey = IC, Blue = anti-CD15. Data is representative of 3 independent experiments. A total of N=18 specimens from individual women were used to generate the data in **B**-**C**.

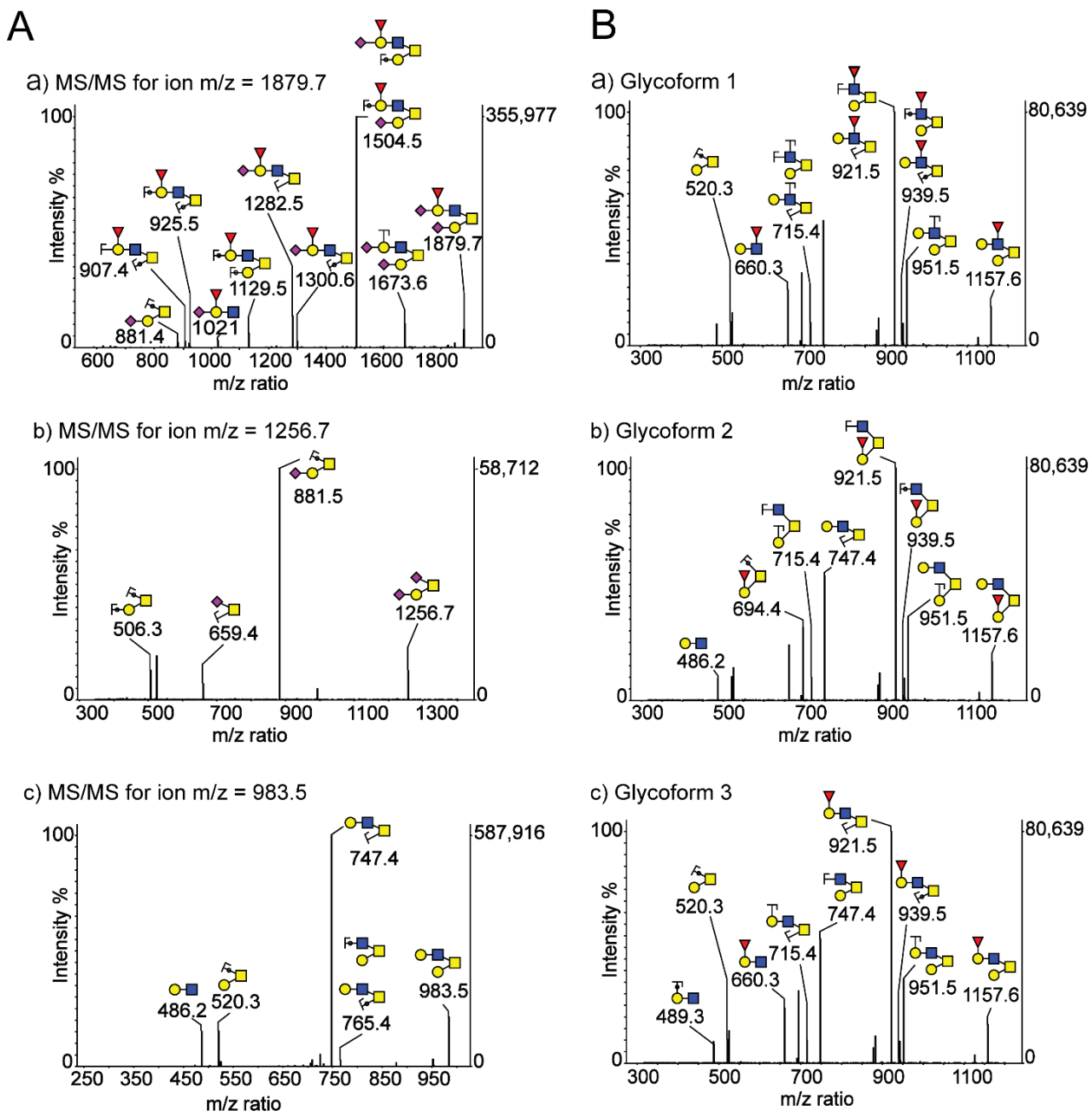

Fig. S5. MS/MS spectra of *O*-glycans present in the vaginal epithelial glycocalyx. Representative MS/MS data for permethylated *O*-glycans indicate presence of core-1 and core-2 structures. (**A**) MS/MS for ion m/z ~a) 1878, b) 1257, and c) 983, indicates presence of di-sialylated and fucosylated core-2 *O*-glycan structure, di-sialylated core-1 *O*-glycan structure, and asialo core-2 containing *O*-glycans respectively. (**B**) MS/MS for ion m/z ~1158 indicates presence of three different glycoforms - a, b, and c, with fucose attached either to *N*-acetylglucosamine or galactose residue. Structures are depicted following the NCBI Symbol Nomenclature for Glycans (red triangle, fucose; yellow circle, galactose; blue square, *N*-acetylglucosamine; yellow square, *N*-acetylgalactosamine; purple diamond, sialic acid). The circles superimposing the bonds mean that the oxygen atoms of these glycosidic bonds are included in the fragment, i.e. they are Y-ions or C-ions.

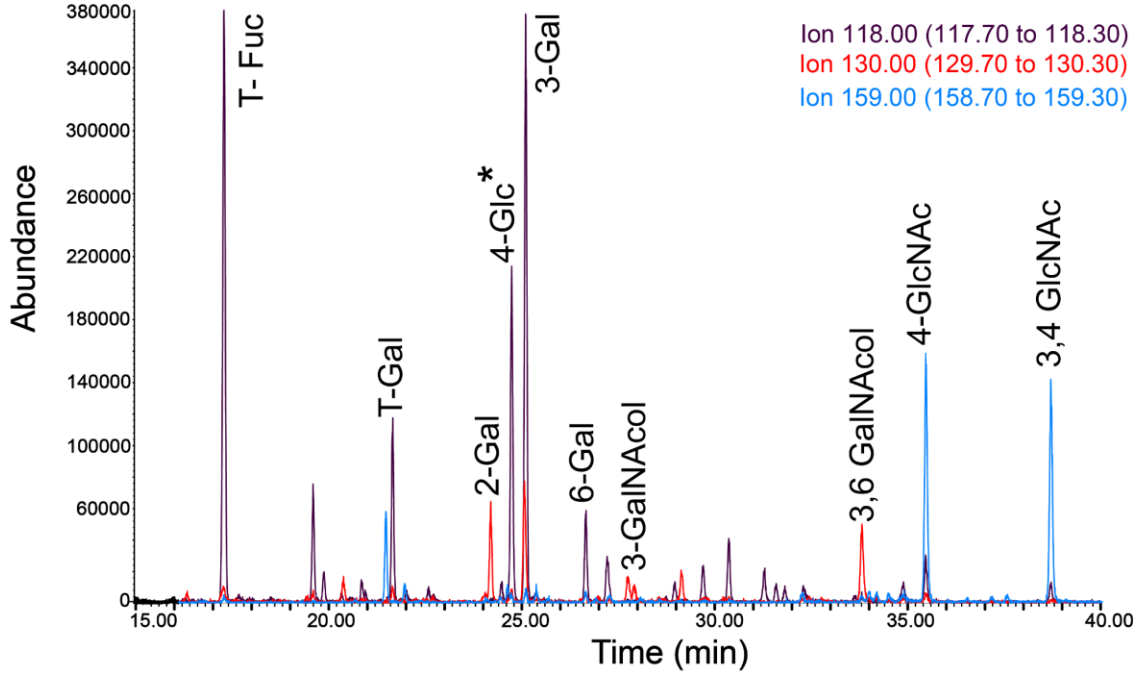

Fig. S6. GC-MS linkage analysis confirms the annotation of selected precursor ions from the VEC *O*-glycome. Gas chromatogram of charge-separated *O*-glycans from a pool of No BV ('normal') VECs. Colors indicate ion scans to identify peaks containing monosaccharide fragment ions of a given m/z as indicated in the legend in the top-right.. Exact elution times for each peak are indicated in table S1. *4-Glc - indicates a contaminating peak from glycogen present in vaginal specimens.

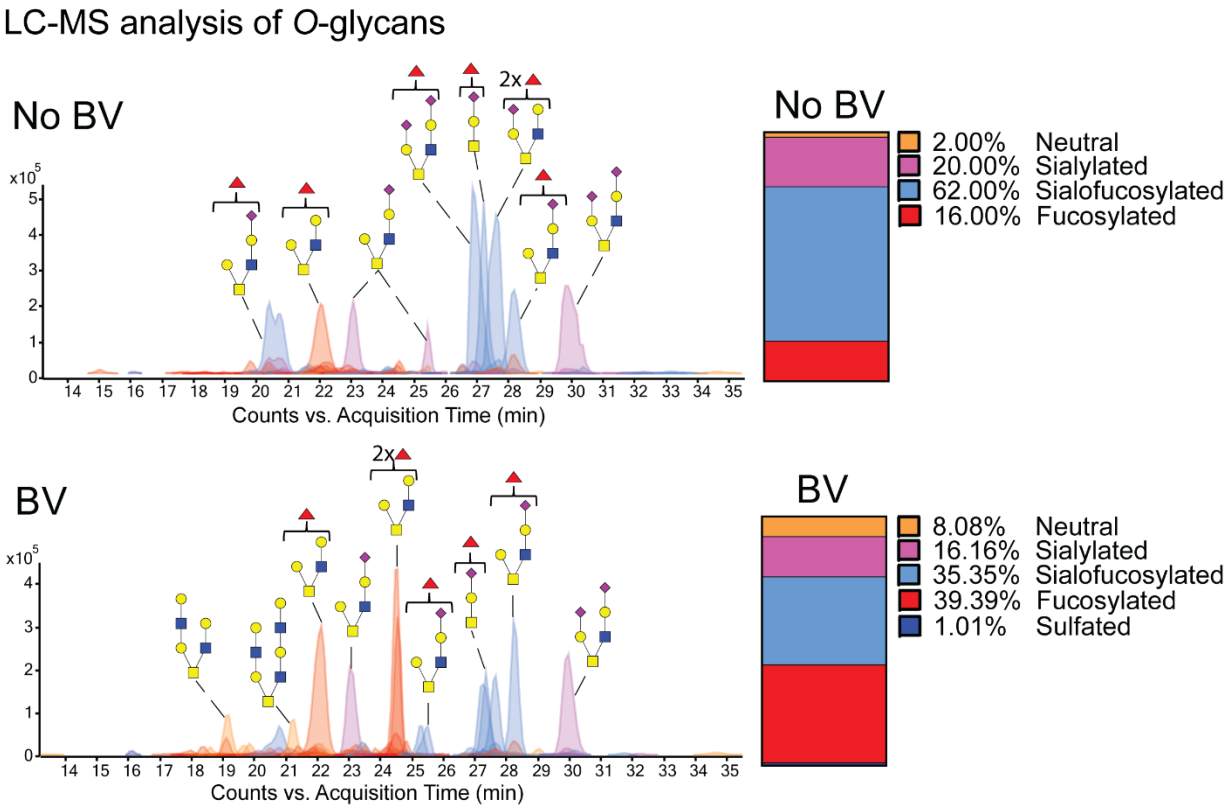

Fig. S7. Mass spectrometric profiling of the *O*-glycans derived from human VECs. Chromatograms show probable structures (based on monosaccharide composition) of VEC *O*-glycans derived from a woman without BV (upper panel) and with BV (lower panel). Relative abundances of *O*-glycan groups are shown next to each chromatogram. Structures of different types of *O*-glycans are depicted following the NCBI Symbol Nomenclature for Glycans (red triangle, fucose; yellow circle, galactose; blue square, *N*-acetylglucosamine; yellow square, *N*-acetylgalactosamine; purple diamond, sialic acid). A complete list of *O*-glycan structures and composition, with their relative abundance, for these specimens is provided as a supplemental file.

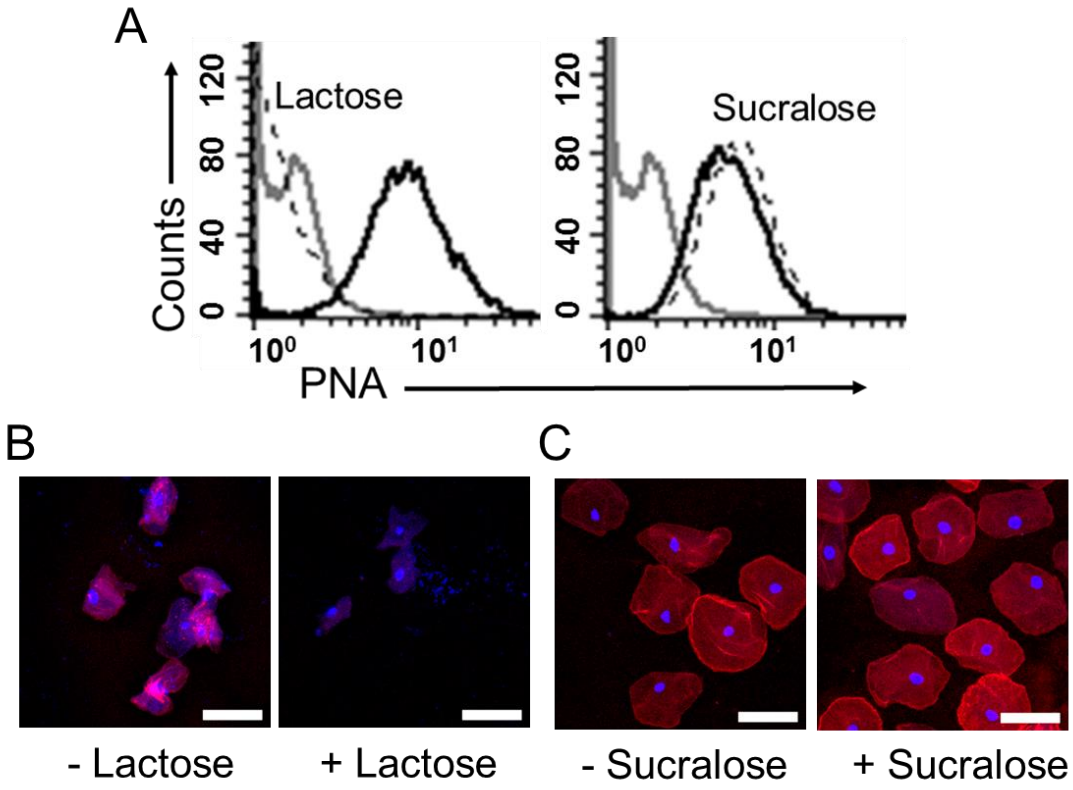

Fig. S8. Verification of PNA selectivity for recognition of galactose containing glycans. VECs isolated from vaginal swabs of women with BV were treated with exogenous sialidase (A.u. sialidase), followed by staining with PNA-rhodamine in the presence of either 10mM lactose or sucralose. (**A**) Histograms show binding of PNA to VECs in the presence (dotted line) or absence (solid black line) of competing saccharide. Grey = Unstained VECs. (**B**) Representative confocal images of VECs stained with PNA in presence and absence of lactose. (**C**) Representative confocal images of VECs stained with PNA in presence and absence of sucralose. Red = PNA, Blue = DAPI. Scale bars = 50 µm.

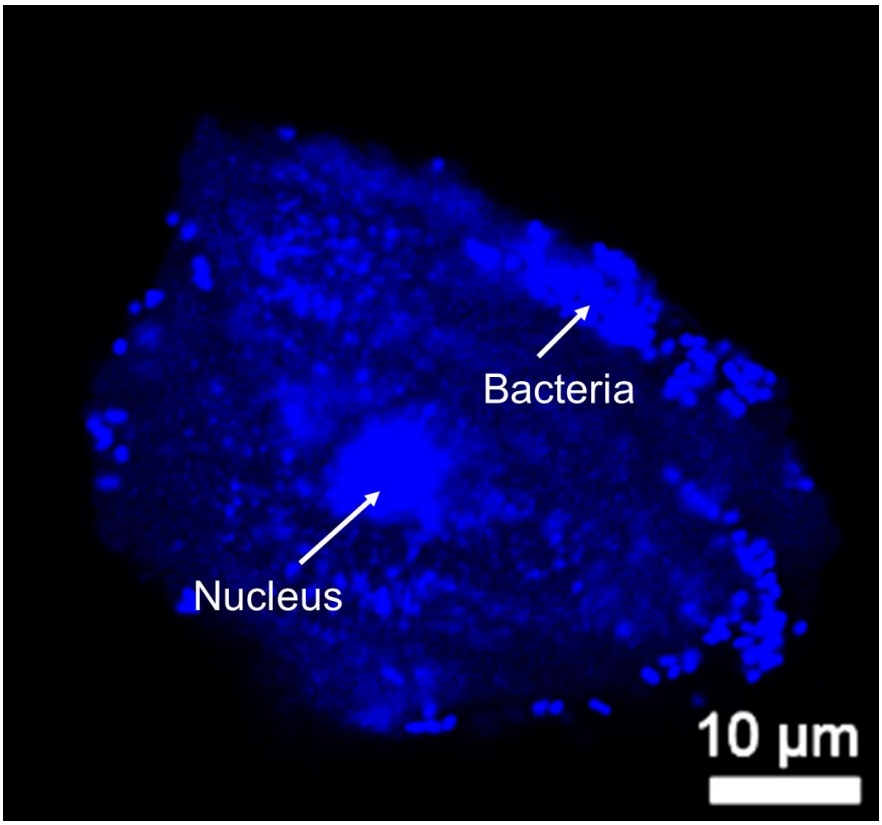

Fig. S9. Visualization of bacteria attached to VECs (Clue cell phenotype). Confocal image of DAPI (blue) stained VEC from a BV specimen. Nucleus is observed in the center. Blue puncta distributed around the nucleus and close to the cell membrane are bacterial cells stained with DAPI, giving a phenotype similar to that observed in wet mounts of women with BV.

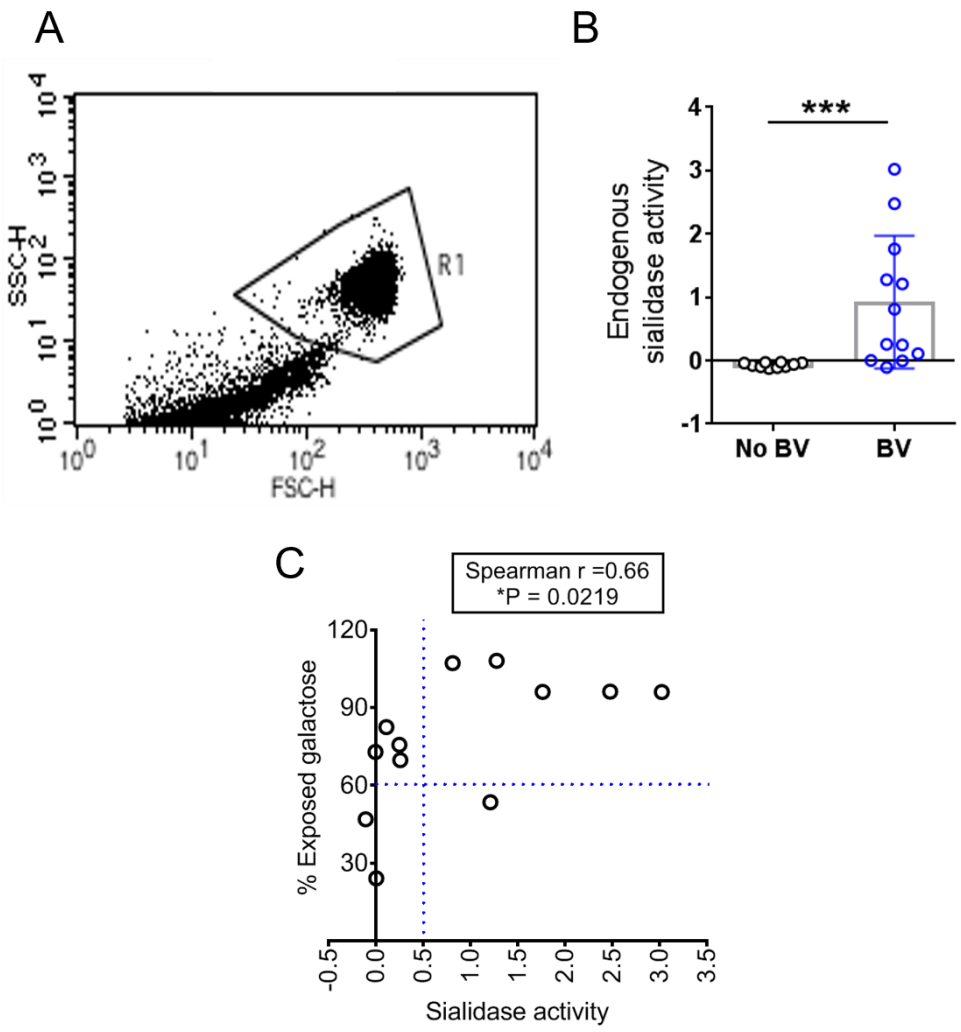

Fig. S10. Galactose exposure on VECs correlates with sialidase activity in BV. (**A**) Dot plot showing flow cytometry gating strategy for selection of VECs based on forward (FSC-H) and side scatter (SSC-H). (**B**) Endogenous sialidase activity in individual vaginal swab eluates estimated using 4MU-Sia assay. Statistical analysis by Mann–Whitney U, ***P<0.001. (**C**) Graph showing correlation analysis for galactose exposure and sialidase activity observed in vaginal specimens of individual women. Samples that were pooled from multiple subjects were not included in this analysis.

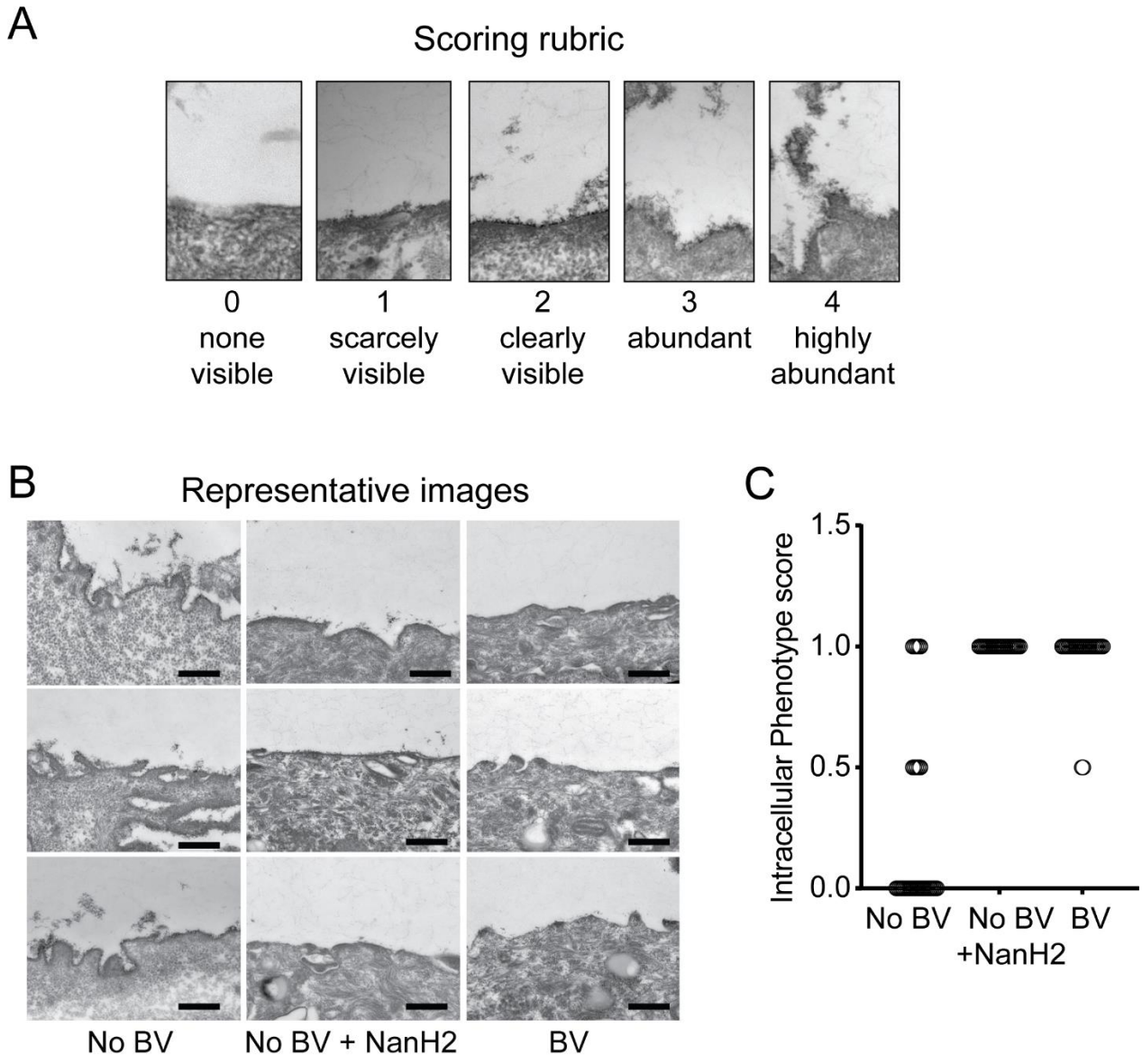

Fig. S11. Rubric for vaginal epithelial glycocalyx scoring and representative images. (**A**) Rubric provided to three blinded observers for glycocalyx scoring. (**B**) Representative transmission electron microscopy (TEM) images of VECs acquired in a blinded fashion from (i) No BV specimens (left), (ii) No BV specimens - treated with *Gardnerella* NanH2 sialidase (middle), and (iii) BV specimens (right). In No BV specimens, glycocalyx appears as a fuzzy layer close to the epithelial cell membrane. Images were acquired in a blinded fashion and scored by three observers blinded to the BV-status of the sample. Scoring of the glycocalyx is shown in **Fig. 6D**, **E**. All images were acquired at 25,000 X. Scale bars are 500 nm. (**C**) Analysis of the intracellular compartment structure**.** An image was scored as 1 if greater than 50% of the intracellular compartment had a threaded appearance and scored as 0 otherwise. Scoring was done by two scorers blinded to the BV-status and experimental treatment of the specimen. A total of ten TEM images were scored from each specimen. Data points in the graph represent the average of two scores for each image. For (i) No BV; n=40 images from n=4 specimens, (ii) BV; n=30 images from n=3 specimens, and (iii) No BV treated with NanH2; n=30 images from n=3 specimens. A total of N=7 specimens were used to generate these data.

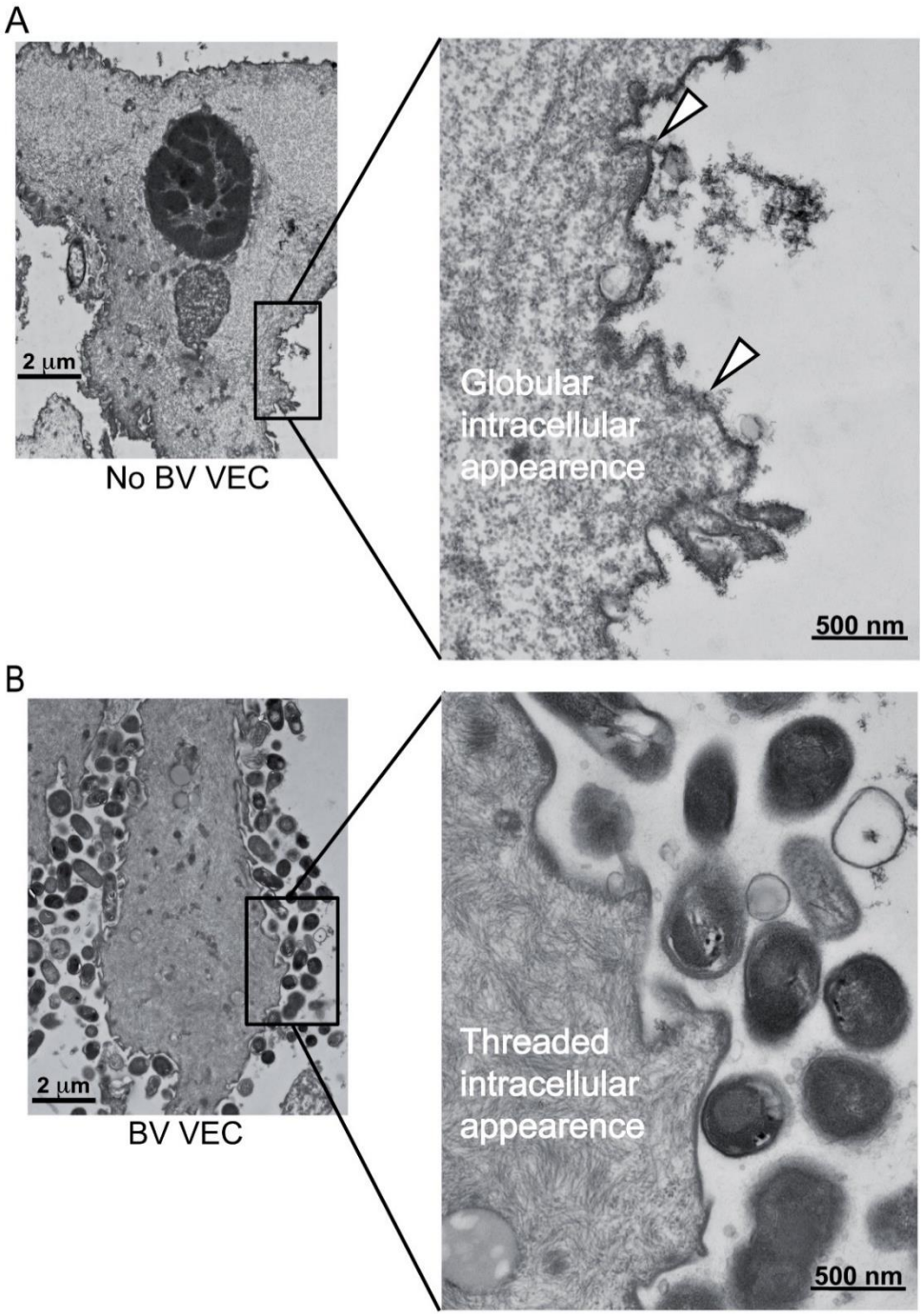

Fig. S12. Visualization of the vaginal epithelial glycocalyx in women with and without BV. Representative transmission electron microscopy (TEM) images of vaginal epithelial cell (VEC) stained with ruthenium red and osmium tetroxide. TEM image of a VEC from (**A**) a woman without BV (Nugent 0-3, referred to as No BV) and (**B**) a woman with BV (Nugent 7-10), with bacteria of varied morphologies around the cell. Arrowheads (Triangles) point to the glycocalyx observed on one VEC from the woman without BV. Text (white) in the images (right panel) describes the appearance of the intracellular compartment as globular (A) or threaded (B). Data is representative of two independent experiments. Scale bar for images in the left panel is 2 μm and for those in the right panel is 500 nm.

ggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactat
aggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatata***CCATGG*****CTACTTCAGAAAAATCCAGTAAACC**CAATAG
TTCAGTTTTGCAAATATCACTTAAAAGAACTGATGGCCTTAAAGATAATGTCCAATTAAATAGTAGGCTAACATATACGGTTACAT
ATACAAATAAGGGTTCACAACCAATAACTGCGTATCCGTATGAAAGTAATATTTCCGATATTAGAGTTGGGGATTCAAACAATTG
CCGCTGGGCAAATCTTAAGCCAGGTGAAACTCAGCAGTGCGGCTTTGGGTACCATGTTGTAACCGAAGAGGATGCCAAAGCGG
GTAAATTCACACCTAAAACAGTTTGGAAAGTTTCAAAAGATAGGGACGGTAACCAAGTAATCGAAAATAACATTGTTGTTGAAGA
CAAGCCTGTTACTGGTATAAAGAATACAGAATTTGACTCATCTTTGGGTCTTAAAACGGTAAAGCTTACAACTAGAACCAATAGC
GTTTGCTACAGGATTCCAGCAATAACACAGACTGCAAATGGTTGGATTTTAGCAGCATACGATAAGAGGCCTGGTAATTGCTCC
GATGCACCTATGCCAAACTCCATTGTTCAGCGAATTTCAAAAGATGGTGGTAAAACATTTGAAGCTGAAACTGTTGTTAAGCAAG
GGCATTATGGCGATGAAAATACTAGATGGGGATATTCGGATCCGTCTTATGTGCTAGACCGTGAAACTAATGAAATTTTCCTGTT
TTCTGTAAAGTCTTACAGAAACGGCTGGGGTGGGTCCAGCGCTGGAGTAGATCCAGAAAACAGGAACGTTTTACACGCTTCAG
TTACTAGCTCCAAGGACAATGGCGTCACTTGGAGCACTCCAAAAATAATTACTAAAGATGTGACTGCAGATGTTGGGCATTGGA
GCTCTAGGTTTGCGGCATCTGGTCACGGCATTCAGCTTCG***CCACGG***CAAGTATAAGGGTCGTTTGTTGCAGCAGTATACAATAA
ACAACTCGGGAAGGATTCAAGCAGTAACGGTTTACTCCGACGATCATGGTAAGACTTGGAAGGTTGGCAAGCCTGTTGGGTTTA
GAAGTGAAAAATCACATATGGACGAAAACAAGGTTGTAGAACTTTCTGATGGACGAGTTATGTTGAATTCTAGGCCTCAGAATTT
TAGGCATCGTCTTGTTGCAATATCAAATGATGGCGGCGAAACTTATGGCGAAGTAAAAGAAGAGATGCAGCTTCCAGACCCTGC
AAACAATGCTCAGATAACTAGGGCATACCCAGATGCGCCTATGGGGTCTCCAGAAGCCAAGGTTCTTGTGTATTCTTCTTCTTC
GGCGTATGGGCGTAGTGACGGCTTAATTCGCGTATCGTTTGATGATGGTGCAACTTGGGATACGGGCAAATTGTTTAAGTCTGG
TCCTATGGCTTATTCTTGCATTGAAGTTTTGAACAAGCAGCATGGTGGCGGATTTGGACTTTTATATGAGGGCGATAATAGCAAT
ATTATTTACACTCATATATCTGAAGATTGGCTTGGTTATACTCCTGTAACCGCAGATAATAGCGTTATTGACGTTCCGAAAAATGC
CAAAAATATTGAAGTTAAGATTAAGAATCTTGGAAGCCATGATTATTCCGACTTAAAGCTTTCTATGAAGGATTTAGGTAAACAAT
CTGAATTAAATAAAGATTGGAAGATTCAAGATAAGACTGGTATTTCTGTAAAGCCTTTTGCTAACGCAAACATAAAGTTAGATATT
GCAATGCCTAAGGATGTTAAAGAAGGAACGACCGTTTCGATTCCAGTGCGATTTGCGTCTGCAAAATATGCAGTGGGAGATTCT
TTGCCAGGAATTCAGTCTGTGATTAATATTCGCGTTGTAAAGGCGGTGGAAAGCCATAGTGTTGTGCCTTCTGTTCCTTCCGTTG
ATGTAAATCCTTCTGTGGATACTCCATCTGCAGAAAAACCTTCTGCAGAAAATCCTTCGCAAAATCCGCCATCGGAAACAATTAC
ACCTACCGAGCCTTCGCAAAATCCAGCGCCTGCAAGAGAACAGGCGCCACAGGTTAGCAATCAATCTAGGAATGAAGCTAATA
CTCAGAAAGCTGATGAGCAAAAGTCTGAGCAAAAGAAAGAGGATTCTAAAGCAAAATACGATGTTGTTCCAGCTCCTCCAGTCA
GCGTTGATCAGTTGGATAAATCCCTAAAGGGAGCTATTACTGTTGGGAATAATAATCAGATTGTTGCTGGTGAGCCTGTAAAGAT
TTCTGGAAAGATTTCCAACAAGGAGTTTGCGCAACAGTTGCAAAAGAATGGTTCAGCAAAGGCCTATGCGTTTATATATTCTACG
CCTAAAGCCTTAAAGGACTCTGATGGCAATTCGTTTATAACGGTTCGTGGATCTGCAGACAATGCACAATTTTCTGTTGATATTC
CAGACGGATATTCTGGAAAGCATACAATCGTGCTTGTAAACGAGAAGGGCGAACAGCTTGCCTGGAATGAGGTAACAGTTAGC
GCAAATCCAGTTGCGGACAAGAAGGATAGTAGTAGTGAAAAATCGGATTTGAACCGTAATGCATTTA**GTAAAGTTTCAACGGGA**
AGTTCA*GTCGAC*aagcttgcggccgcactcgag**caccaccaccaccaccac**tgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctga
gcaataactagcataaccccttggggcctctaaacgggtcttgaggggtttttgctgaaaggaggaactatatccgga

Fig. S13. Sequence of the truncated *nanH2* gene. Sequence of the truncated *nanH2* gene (black) cloned in pET28a vector (flanking sequence in gray lowercase). The bold underlined parts in black are the sequences of the polymerase chain reaction primer. The bold and italicized part in black (shaded gray) indicates the mutated NcoI site in the *nanH2* gene. The bold underlined part in gray indicates the sequence of His tag.

Table S1. Characteristics of the study population.

Characteristic	**Overall**	**BV**	**No BV**	**P-value**
	N=258	n=97	n=161	
Age (mean years ± SD)	25.7 ± 6.2	25.7	25.7 -	0.9519
Race				<0.001
Black	130 (50.4%)	66 (68.0%)	64 (39.8%)	
White	111 (43.0%)	27 (27.8%)	84 (52.2%)	
Other	17 (6.6%)	4 (4.1%)	13 (8.1%)	
Marital status				0.7683
Single, never married	143 (55.4%)	56 (57.7%)	87 (54.0%)	
Married or living with partner	95 (36.8%)	33 (34.0%)	62 (38.5%)	
Separated, divorced, or widowed	20 (7.8%)	8 (8.2%)	12 (7.5%)	
Education				<0.001
<= High School	85 (33%)	43 (44.3%)	42 (26.1%)	
Some college	102 (39.5%)	41 (42.3%)	61 (37.9%)	
College graduate	71 (27.5%)	13 (13.4%)	58 (36.0%)	
Nulligravid	89 (34.5%)	18 (18.6%)	71 (44.1%)	<0.001
Nulliparous	122 (47.3%)	32 (33.0%)	90 (55.9%)	<0.001
Sexually transmitted infection				
Chlamydia trachomatis	7 (2.8%)	5 (5.2%)	2 (1.3%)	0.0644
Neisseria gonorrheae	3 (1.2%)	3 (3.1%)	0	0.0257
Trichomonas vaginalis	7 (2.7%)	6 (6.3%)	1 (0.6%)	0.0080
Prior diagnosis of sexually transmitted infection	92 (36.7)	42 (43.3%)	50 (31.1%)	0.0482
Prior diagnosis of BV	67 (26.0%)	29 (29.9%)	38 (23.6%)	0.4920
Currently pregnant	32 (18.9%)	16 (20.3%)	16 (17.8%)	0.6834
Lifetime number of sexual partners (mean ± SD)	8.6 ± 12.0	9.5 (1.4)	8.1 (0.9)	0.3424

Table S2. Definition of the constituents of *O*-glycans of the vaginal epithelium. We identified the monosaccharide components and linkages of the *O*-glycans of VECs from women without BV by GC-MS analysis of partially methylated alditol acetates (PMAAs). Data is representative of 2 independent experiments. A total of N=10 No BV specimens, combined to form one VEC pool, were used to generate these data. VEC pool generated from these specimens was also used for studies reported in **Fig. 1C**, **1D** and to confirm structures of data reported in **Fig. 3**.

Elution time (min)	**Characteristic fragment ions**	**Assignment[a]**	**Probable parent *O*-glycan**
17.288	118,131,162,175	terminal-Fuc	Fucosylated *O*-glycans
21.655	118, 161, 205	terminal-Gal	Core-1/Core-2 *O*-glycan/ T-antigen
24.205	190, 161	2-Gal	Fucosylated galactose
25.105	118,129,161,234	3-Gal	Sialylated T-antigen
26.664	118, 129,162,189,233	6-Gal	Likely Sialylated
27.759	130, 246, 290	3-GalNAc-itol	Core 1 *O*-glycan
33.798	130, 246, 318	3,6-GalNAc-itol	Core 2 *O*-glycan
35.467	117, 159, 233	4-GlcNAc	Extended core-2 *O*-glycan
38.692	117, 159, 346	3,4-GlcNAc	Lewis antigens

[a]The respective PMAAs were identified by elution time and their characteristic fragment ions (*126*).

Table S3. Genes differentially expressed in VK2 cells after exposure to *Gardnerella* sialidase.

Genes differentially expressed in the VK2 cells after one-hour NanH2 treatment				
Gene name	Description	logCPM	logFC	Adj. P value
SLC30A1	Solute Carrier Family 30 Member 1	7.906	-0.4511	0.00003344
SLC20A1	Solute Carrier Family 20 Member 1	8.953	0.2881	0.0001952
MT1E	Metallothionein 1E	6.317	-0.6862	0.0003299
MMP1	Matrix Metallopeptidase 1	5.153	0.5362	0.002997
KLF4	KLF Transcription Factor 4	5.948	0.3741	0.006785
EGR1	Early Growth Response 1	5.292	0.5104	0.007241
DDIT4	DNA Damage Inducible Transcript 4	8.476	-0.2974	0.007241
SOX9	SRY-Box Transcription Factor 9	6.792	0.3695	0.01413
EREG	Epiregulin	6.771	0.3688	0.01572
LDLR	Low Density Lipoprotein Receptor	9.098	0.2121	0.0178
MAFB	MAF BZIP Transcription Factor B	5.927	0.2718	0.02127
RHOB	Ras Homolog Family Member B	8.469	0.2362	0.02199
FOXQ1	Forkhead Box Q1	4.368	0.452	0.02199
TRIB1	Tribbles Pseudokinase 1	5.892	0.3109	0.02199
BMP2	Bone Morphogenetic Protein 2	5.872	0.3501	0.0261
LAMB3	Laminin Subunit Beta 3	10.85	0.1631	0.0261
HS3ST1	Heparan Sulfate-Glucosamine 3-Sulfotransferase 1	3.422	0.4724	0.02723
IER2	Immediate Early Response 2	6.966	0.2494	0.02759
PPP1R15A	Protein Phosphatase 1 Regulatory Subunit 15A	5.355	0.4338	0.03325
DUSP4	Dual Specificity Phosphatase 4	7.597	0.258	0.03325
JAG1	Jagged Canonical Notch Ligand 1	7.807	0.2305	0.0381
GADD45A	Growth Arrest And DNA Damage Inducible Alpha	5.962	0.2743	0.04119
DUSP5	Dual Specificity Phosphatase 5	6.496	0.4038	0.04119
RAP2B	Ras-Related Protein Rap-2b	7.147	0.1872	0.04119
ACTG1	Actin Gamma 1	10.63	0.1382	0.04256

ACTB	Actin Beta	11.76	0.1737	0.043
Genes differentially expressed in the VK2 cells after two-hour NanH2 treatment				
Gene name	Description	logCPM	logFC	Adj. P value
TNFAIP3	TNF Alpha Induced Protein 3	5.69	0.7026	0.0004522
NFKBIA	NFKB Inhibitor Alpha	5.34	0.517	0.001026