SUPPLEMENTAL MATERIALS TO

Normalization of the host intestinal mucus systems requires long-term colonization

Malin E V Johansson[1], Hedvig E Jakobsson[1], Jessica Holmén-Larsson[1], André Schütte[1], Anna Ermund[1], Ana M Rodríguez-Piñeiro[1], Liisa Arike[1], Catharina Wising[1], Frida Svensson[1], Fredrik Bäckhed[2] and Gunnar C Hansson[1]

Pages:

SUPPLEMENTAL COMPLETE EXPERIMENTAL PROCEDURES

Animals

GF female mice (8-16 weeks) were bred in contamination-controlled flexible film isolators and were colonized with caecal flora from C57BL/6 mice known to have a normal mucus phenotype and treated as conventional animals onwards (Backhed et al., 2004). Conventionalization was performed using the same procedures in all mice. Each experiment and week were performed on a minimum of three mice. Experiments were repeated times. Convr C57BL/6 mice were treated with antibiotics in the drinking water for 3 weeks with ampicillin (1 g/l, Doktacillin, AstraZeneca), metronidazole (1 g/l, Sigma-Aldrich) vancomycin (0.5 g/l, Hospira) and neomycin trisulfate (1 g/l, Sigma-Aldrich) (Fu et al., 2011). Animals were housed in standardized conditions of temperature (21-22°C) and illumination (12 h light and 12 dark) with food and water provided ad libitum. After 2 to 8 weeks of conventionalization the mice were euthanized by isoflurane and cervical dislocation. All animal experimental procedures were approved by the Swedish Laboratory Animal Ethical Committee in Gothenburg.

Explant Tissue

Gastrointestinal tissue was dissected, flushed with ice-cold oxygenated (95% O_2, 5% CO_2) Krebs transport buffer and kept on ice, and the longitudinal muscle layer removed by blunt dissection. The intestinal explants were mounted in the horizontal perfusion chamber as described previously (Gustafsson et al., 2012b).

Mucus thickness measurements

The thickness of the intestinal mucus was measured as described previously (Gustafsson et al., 2012a; Gustafsson et al., 2012b). Briefly, the upper surface of the mucus, secreted by the mounted tissue, was visualized by addition of charcoal particles. The mucus thickness was determined by the distance between the epithelial surface and the mucus surface by a micropipette viewed through a stereomicroscope over time. The mucus that was not attached to the epithelial surface was removed by suctioning with a micropipette and the remaining small intestine or inner colon mucus was determined. Data are presented as mean ± SD.

Mucus penetrability assay

Mucus penetrability was measured as described previously (Gustafsson et al., 2012b; Johansson et al., 2014). Briefly, mouse colonic explants were mounted in the perfusion chamber and incubated for 20 min followed by the addition of a suspension of 2 μm green beads, 1 μm purple beads and 0.5 μm red beads (Fluospheres, Life technologies) to the apical surface. The beads were left to sediment through the mucus for 40 min after which the position of the beads in relation to the epithelium was analyzed by taking confocal images in XY stacks with an LSM 700 Axio Examiner Z1 confocal imaging system with a Plan- Apochromat 20x/1.0DIC water objective (Zeiss). Images were acquired using the ZEN 2010 software (Zeiss). The impenetrable mucus was determined as the distance to the 20 most penetrating beads. Distances from region of interest at the tissue surface to points indicating beads were measured using the Volocity 6.1.1 software (Perkin-Elmer). Data are presented as median with range.

Tissue fixation and immunostaining

Pieces of ileum or colon with fecal material were fixed in water free-Carnoy (methanol) and paraffin-embedded. Slides with sections were dewaxed and stained with hematoxylin and eosin, Alcian blue/ periodic acid–Schiff (PAS) or hybridized with 10 ng/µL of a general bacterial 16S rRNA gene probe (EUB 338) and immunostained for Muc2 using the MUC2C3 antisera and DNA by Hoechst 34580 (Life technologies) as previously described (Johansson et al., 2008). Immunostaining without FISH was performed after citric buffer antigen retrieval using the apoMuc2 antisera (Hansson et al., 1994). Images were obtained using a fluorescence microscope Eclipse E1000 with a Plan-Fluor 40x/0.75 DIC objective (Nikon) or an Axio Examiner Z1 LSM 700 confocal microscope and the ZEN 2010 software (Zeiss).

Proteomic analysis of mucus samples

The mucus samples removed after thickness measurements were solubilized in a guanidinium hydrochloride-based buffer, following the filter-aided sample preparation (FASP) method as described before (Rodriguez-Pineiro et al., 2013). Isotopically labelled peptides of Muc2, Clca1, Fcgbp and Agr2 were added to each sample as standards. Proteins were then digested with 10 ng/µl porcine trypsin (Promega), and the resulting peptides were eluted and cleaned in StageTips (Rappsilber et al., 2003). These samples were then analyzed by nano-reversed phase liquid chromatography (nRPLC) coupled to electrospray ionization-tandem mass spectrometry (ESI-MS/MS) in an LTQ-Orbitrap XL (Thermo Scientific), following the settings previously described for mucus analysis (Rodriguez-Pineiro et al., 2013). Results from the MS/MS experiments were analyzed with the MaxQuant 1.2.2.5 software (Cox et al., 2011). The databases used were the MaxQuant contaminant database, and the UniProt-SwissProt mouse database (version of March 2012, only reviewed sequences) completed with our public mucin

database (http://www.medkem.gu.se/mucinbiology/databases/). For the searches, we set Met oxidation and N-terminal acetylation as variable modifications, and Cys carbamidomethylation as fixed. The main search error was limited to 6 ppm, with a false discovery rate threshold of 0.01. Relative protein amounts were quantified in ppm relative to the total proteins detected in each sample. Data are presented as interquartile range boxplots with median and whiskers of one standard deviation. For Muc2, Clca1, Fcgbp and Agr2, the peak areas of isotopically labelled peptides and their endogenous light cognates were extracted with Skyline (version 2.5.0.6157) (Schilling et al., 2012) and quantified as ratios between the endogenous peptides and the standards.

Proteomic analysis of epithelial cells

Intestinal tissues were incubated in PBS containing 3 mM EDTA and 1 mM DTT at 37°C for 60 min. Thereafter epithelial cells were dissociated from the tissue by vigorous shaking in PBS and pelleting by centrifugation at 1000g for 5 min. Proteins were digested with LysC (Wako Chemicals USA, Inc.) according to FASP protocol (Wisniewski et al., 2009) and the resulting peptides were eluted and fractionated and cleaned in SAX StageTips to two fractions (pH11 and pH3) (Rappsilber et al., 2007). Samples were analyzed with EASY-nLC system (Thermo Scientific, Odense, Denmark) connected to a Q-Exactive (Thermo Scientific, Bremen, Germany). Peptides were separated with a 95-minute gradient: for pH3 fraction from 10 to 30% B (A: 0.1% formic acid, B: 0.1% formic acid/80% acetonitrile) and for pH11 fraction in 80 min from 5 to 20% B and in next 15 min up to 30% B, using a flow rate of 200 nl/min. MS raw files were processed with MaxQuant software version 1.3.0.5 (Cox et al., 2011) as described above. Protein abundances were calculated by dividing for each protein the summed peptide intensities

by the number of theoretically observable peptides (all fully LysC digested peptides between 800 and 2,500 Da) and normalizing to all proteins identified in the sample; the normalized values were multiplied with the number of protein copies in the mammalian cell - 10^9 (Milo, 2013).

Mucin oligosaccharide analysis by capillary-LC-MS/MS

Mucus from midcolon of GF, Convr, and conventionalized mice (2, 3, 4, 6 and 8 weeks after gavage) was scraped and the insoluble Muc2 mucin semipurified by 6M guanidinium hydrochloride (GuHCl) extraction to remove other components. The GuHCl insoluble fractions containing Muc2 of the different mice were reduced, alkylated and analyzed by Ag-PAGE composite gel electrophoresis (Holmen-Larsson et al., 2013). The gel was stained by Alcian blue to detect the negatively charged glycans, and the protein bands were blotted to a PVDF membrane. The stained Muc2 monomeric bands were excised and the *O*-glycans released by reductive β-elimination, desalted and analyzed by capillary-LC MS/MS in negative ion mode (Schulz et al., 2002; Holmen-Larsson et al., 2013). The glycans were detected as mono-, doubly or triply charged ions depending on the number of charged substituents.

Quantitative real-time PCR

Distal colon pieces of 10-30 mg with fecal material removed were snap-frozen in liquid nitrogen and stored in -80°C. Total RNA extraction was performed using the RNeasy kit (Qiagen) according to manufacturer´s instructions. The homogenization step was performed in a FastPrep®-24 Instrument (MP Biomedicals) using a 5-mm magnetic bead for 20 s. The RNA concentration was determined using a NanoDrop™ instrument (Thermo Scientific) and RNA integrity was assessed by electrophoresis in an Experion™ Automated Electrophoresis station

(Bio-Rad) using the Experion StdSens Analysis Kit (Bio-Rad). Total RNA (500 ng/20 uL reaction) was used for cDNA synthesis using a High Capacity cDNA synthesis kit (Life Technologies) according to the manufacturer´s protocol.

Samples were run in triplicates on a CFX96 system (Bio-Rad) using 1x SsoFast® EvaGreen™ supermix reagent (Bio-Rad) in a 20 uL reaction containing 25 ng of cDNA. Final primer concentrations were 150 nM for Muc2 and 300 nM for other genes. PCR cycling parameters were set to 95°C for 2 minutes, (95°C for 5s/Annealing for 5s/72°C for 10s) x40 followed by a Melt Curve 65°C to 95°C, 0.5°C increment. IL1β, IL10, IL13, IL17, and IFNgamma were tested together with Muc2. Primer sequences and annealing temperatures for IL1β, IL17, and Muc2 that showed robust expression differences can be found in Table S1. Data analysis was done using Bio-Rad CFX Manager 3.1 and samples were normalized (ΔΔCq) against GAPDH and *β*-Actin.

Sample collection for microbiota analysis

Using clean, sterile dissection tools, the ileum (Si8 only), distal colon and cecum were immediately removed. The lumen content from Si8 and distal colon was removed by gently squeezing out the intestinal content in a collection tube. The intestinal segment was then gently flushed with 2x1 ml sterile PBS. This was considered the mucus sample. The samples were immediately flash-frozen in liquid nitrogen and later stored at −80°C until analysis.

DNA extraction

DNA was extracted from 80 mg of caecal sample and 10-130 mg of lumen or mucosal tissue from each mice (Jakobsson et al., 2015).

16S rRNA gene tag pyrosequencing

Bacterial 16S rRNA gene sequences were amplified from each sample using the primers 27F (5' AGAGTTTGATCCTGGCTCAG 3') with Titanium Adaptor B and 338R (5' TGCTGCCTCCCGTAGGAGT 3') with Titanium Adaptor A and a sample-specific barcode sequence consisting of twelve nucleotides targeting the V1-V2 hypervariable region of the 16S rRNA gene. Primers included unique error-correcting 12-base barcodes used to tag PCR products from different samples (Fierer et al., 2008). The 16S rRNA genes were amplified in a 50 µl reaction containing 1X PCR Buffer, 0.2 µM of each primer, 200 µM dNTP (PurePeak) DNA polymerase Mix (Pierce Nucleic Acid Technologies), 1.5U of FastStart Taq DNA Polymerase (Roche) and 10-300 ng of DNA template. The PCR conditions used were 95ºC for 3 min, 25 cycles of 95ºC for 20 s, 52ºC for 30 s and 72ºC for 1 min followed by 72ºC for 10 min. The distal colon samples were run for 28 cycles. Triplicate PCR reactions were performed for each sample. A negative PCR reaction without template was also included in each reaction. The triplicate PCR reactions were pooled and purified with AMPure beads (Becton Dickinson) using a Magnet Particle Separator (Invitrogen) The concentrations of the purified products were measured by Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen). The quality was assessed on a 1% agarose gel. The samples were amplified in PCR-mixture-in-oil emulsions and sequenced from the 338R primer using Roche 454 FLX and Titanium chemistry (Roche) at the Science for Life Laboratories (Solna, Sweden).

16S rRNA Gene Sequence Analysis

Post processing of pyrosequencing data was done using QIIME software (1.7.0) package (Caporaso et al., 2010a). For quality-filtering the raw data we discarded sequences < 200 bp or >1,000 bp, and sequences containing primer mismatches, uncorrectable barcodes, ambiguous

bases, or homopolymer runs in excess of 6 bases. The sequences were checked for chimeras using ChimeraSlayer (Haas et al., 2011) and assigned to operational taxonomic units (OTUs) using UCLUST (Edgar, 2010) with a 97% threshold of pairwise identity, and then classified taxonomically using Greengenes (GG) reference database (The GGs Databasefile version 13_5). The reads were aligned to the Greengenes Core reference alignment using PyNAST (Caporaso et al., 2010b). The GG taxonomies were used to generate summaries of the taxonomic distributions of OTUs across different levels (phylum, order, family, genus, and species). All OTUs that were observed fewer than 2 times were discarded. To standardize sequence counts across samples with uneven sampling, we randomly selected 1,000 sequences per sample (rarefaction) and used this as a basis to compare abundances of OTUs across samples. For phylogenetic tree-based analyses, each OTU was represented by a single sequence that was aligned using PyNAST (Caporaso et al., 2010b). A phylogenetic tree was built with FastTree (Price et al., 2009) and used for estimates of α-diversity (Chao, 1984), Shannon diversity (Hayek and Buzas, 1997), and β-diversity using unweighted UniFrac (Lozupone and Knight, 2005) . For all diversity measurements, means and standard errors for given categories were calculated from 100 iterations using a rarefaction of 1,000 sequences per sample.

Statistical analysis

For all mucus measurements data comparing two independent groups was analyzed using a two-tailed Mann-Whitney U test. Multiple independent group comparisons were performed using Kruskal-Wallis ANOVA with Dunn's correction for multiple comparisons. GraphPad Prism 6 was used for plotting the results. A P-value <0.05 (*) was regarded as statistically significant (**= P<0.01, ***= P<0.001, ****= P<0.0001). Only significant differences are marked in

graphs. The peptide ratios from the proteomics data were analyzed for significant differences by ANOVA with 250 randomizations and FDR correction at 1%. Multiscatter between all samples and correlation calculations using Pearson's correlation coefficient were performed. For the microbiota analysis, statistical significance testing for over- and under-representation of the bacterial lineages was made at phylum, class and genus (3% dissimilarity) levels. Significant differences were analyzed using the Wilcoxon rank-sum test, and P-values were converted to false discovery rate values (q-values) to correct for multiple testing in the R software (http://www.r-project.org/).

References

Backhed,F., Ding,H., Wang,T., Hooper,L.V., Koh,G.Y., Nagy,A., Semenkovich,C.F., and Gordon,J.I. (2004). The gut microbiota as an environmental factor that regulates fat storage. PNAS *101*, 15718-15723.

Caporaso,J.G., Kuczynski,J., Stombaugh,J., Bittinger,K., Bushman,F.D., Costello,E.K., Fierer,N., Pena,A.G., Goodrich,J.K., Gordon,J.I.et al. (2010a). QIIME allows analysis of high-throughput community sequencing data
1. Nat Methods *7*, 335-336.

Caporaso,J.G., Bittinger,K., Bushman,F.D., DeSantis,T.Z., Andersen,G.L., and Knight,R. (2010b). PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics *26*, 266-267.

Chao,A. (1984). Nonparametric Estimation of the Number of Classes in a Population. Scandinavian Journal of Statistics *11*, 265-270.

Cox,J., Neuhauser,N., Michalski,A., Scheltema,R.A., Olsen,J.V., and Mann,M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. *10*, 1794-1805.

Edgar,R.C. (2010). Search and clustering orders of magnitude faster than BLAST
1. Bioinformatics. *26*, 2460-2461.

Fierer,N., Hamady,M., Lauber,C.L., and Knight,R. (2008). The influence of sex, handedness, and washing on the diversity of hand surface bacteria
2. Proc Natl Acad Sci U. S. A *105*, 17994-17999.

Fu,J., Wei,B., Wen,T., Johansson,M.E.V., Xiaowei L., Bradford,E., Thomsson,K.A., McGee,S., Mansour,L., Tong,M.et al. (2011). Loss of intestinal core 1-derived *O*-glycans causes spontaneous colitis. J. Clin. Invest. *121*, 1657-1666.

Gustafsson,J.K., Ermund,A., Ambort,D., Johansson,M.E.V., Nilsson,H.E., Thorell,K., Hebert,H., Sjovall,H., and Hansson,G.C. (2012a). Bicarbonate and functional CFTR channel is required for proper mucin secretion and link Cystic Fibrosis with its mucus phenotype. J. Exp. Med. *209*, 1263-1272.

Gustafsson,J.K., Ermund,A., Johansson,M.E.V., Schutte,A., Hansson,G.C., and Sjovall,H. (2012b). An ex vivo method for studying mucus formation, properties and thickness in human colonic biopsies and mouse small and large intestinal explants. Am. J. Physiol. Gastrointest. Liver. Physiol. *302*, G430-G438.

Haas,B.J., Gevers,D., Earl,A.M., Feldgarden,M., Ward,D.V., Giannoukos,G., Ciulla,D., Tabbaa,D., Highlander,S.K., Sodergren,E.et al. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons
1. Genome Res *21*, 494-504.

Hansson,G.C., Baeckstrom,D., Carlstedt,I., and Klinga-Levan,K. (1994). Molecular Cloning of a cDNA Coding for a Region of an Apoprotein from the Insoluble Mucin Complex of Rat Small Intestine. Biochem. Biophys. Res. Commun. *198*, 181-190.

Hayek,L.C. and Buzas,M.A. (1997). Surveying Natural Populations. (New York: Columbia University Press).

Holmen-Larsson,J.M., Thomsson,K.A., Rodriguez-Pineiro,A.M., Karlsson,H., and Hansson,G.C. (2013). Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin *O*-glycan patterns reveal a regiospecific distribution. Am. J. Physiol. Gastroint. Liver Physiol. *305*, G357-G363.

Jakobsson,H.E., Rodriguez-Pineiro,A.M., Schütte,A., Ermund,A., Boysen,P., Sommer,F., Bäckhed,F., Hansson,G.C., and Johansson,M.E.V. (2015). The gut microbiota composition impairs the colon inner mucus layer barrier. EMBO Reports *16*, 164-177.

Johansson,M.E.V., Gustafsson,J.K., Holmen-Larsson,J., Jabbar,K.S., Xia,L., Xu,H., Ghishan,F.K., Carvalho,F.A., Gewirtz,A.T., Sjövall,H.et al. (2014). Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and in patients with ulcerative colitis. Gut *213*, 281-291.

Johansson,M.E.V., Phillipson,M., Petersson,J., Holm,L., Velcich,A., and Hansson,G.C. (2008). The inner of the two Muc2 mucin dependent mucus layers in colon is devoid of bacteria. Proc. Natl. Acad. Sci. USA *105*, 15064-15069.

Lozupone,C. and Knight,R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities
1. Appl Environ. Microbiol. *71*, 8228-8235.

Milo,R. (2013). What is the total number of protein molecules per cell volume? A call to rethink some published values. Bioessays *35*, 1050-1055.

Price,M.N., Dehal,P.S., and Arkin,A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix
1. Mol Biol Evol *26*, 1641-1650.

Rappsilber,J., Ishihama,Y., and Mann,M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal. Chem. *75*, 663-670.

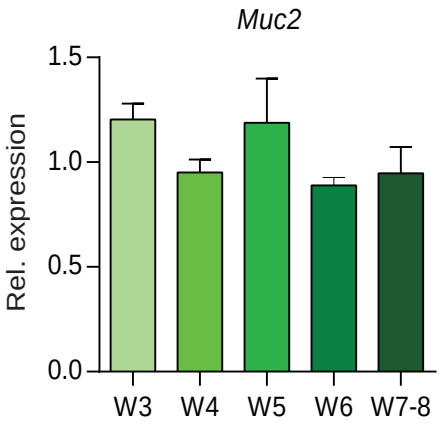
Rappsilber,J., Mann,M., and Ishihama,Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc. *2*, 1896-1906.

Rodriguez-Pineiro,A.M., Bergstrom,J.H., Ermund,A., Gustafsson,J.K., Schutte,A., Johansson,M.E.V., and Hansson,G.C. (2013). Studies of mucus in mouse stomach, small intestine, and colon. II. Gastrointestinal mucus proteome reveals Muc2 and Muc5ac accompanied by a set of core proteins. Am. J. Physiol. Gastroint. Liver Physiol. *305*, G348-G356.

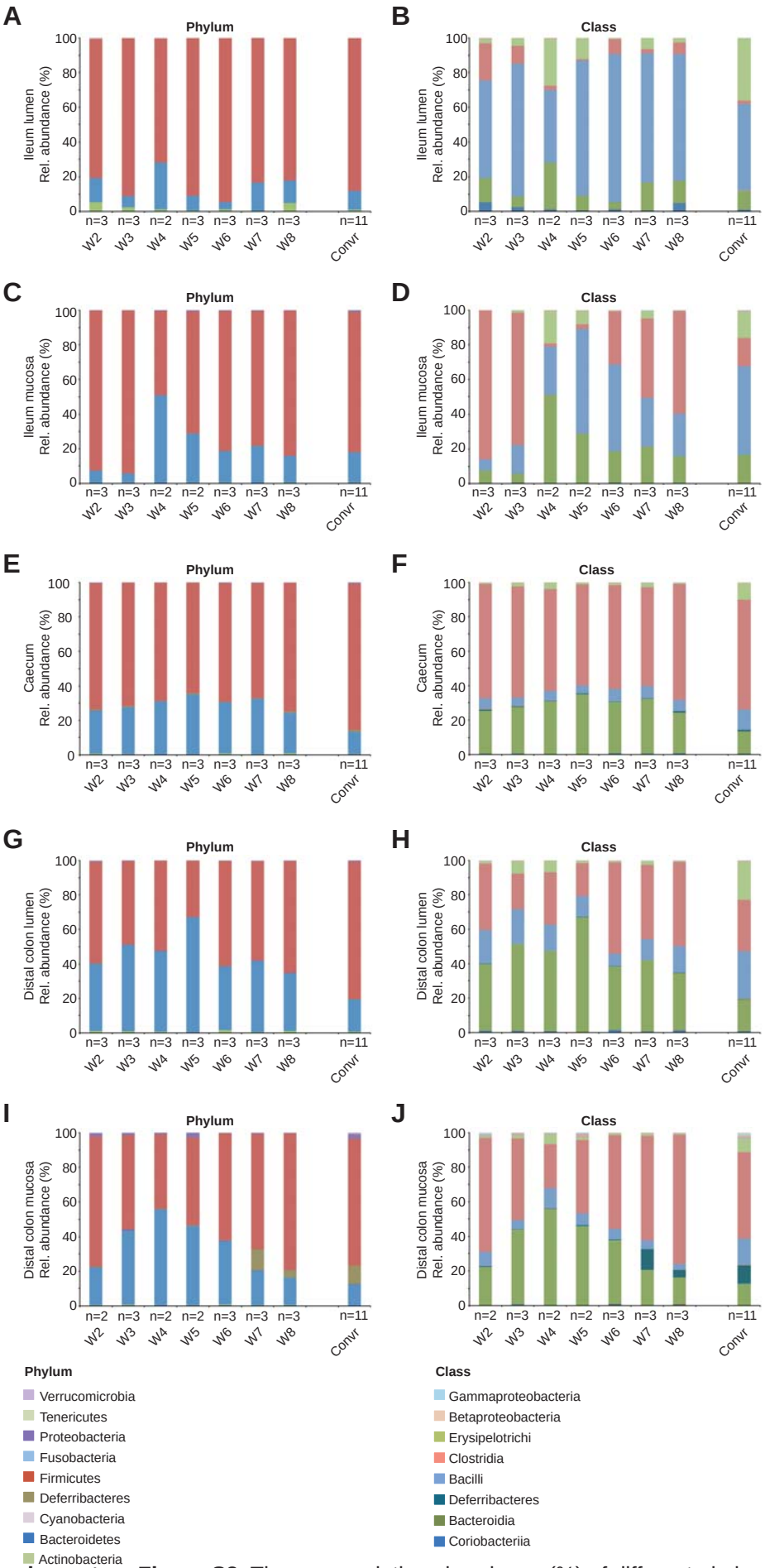
Schilling,B., Rardin,M.J., MacLean,B.X., Zawadzka,A.M., Frewen,B.E., Cusack,M.P., Sorensen,D.J., Bereman,M.S., Jing,E., Wu,C.C.et al. (2012). Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. Mol Cell Proteomics *11*, 202-214.

Schulz,B.L., Packer,N., and Karlsson,N.G. (2002). Small-scale analysis of O-linked oligosaccharides from glycoproteins and mucins separated by gel electrophoresis. Anal. Chem. *74*, 6088-6097.

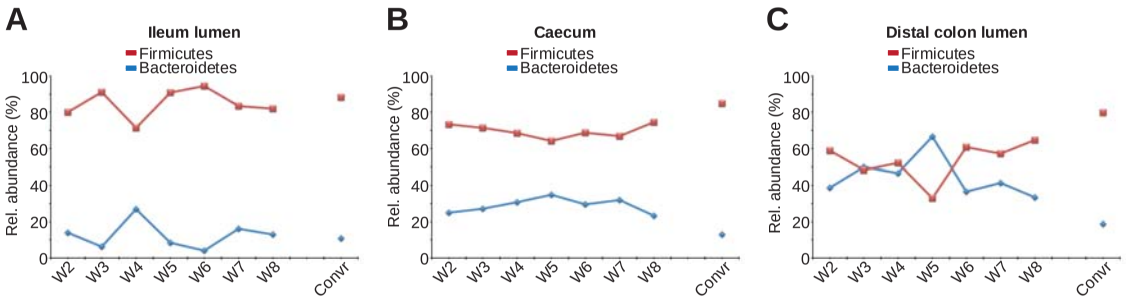
Wisniewski,J.R., Zougman,A., Nagaraj,N., and Mann,M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359-362.



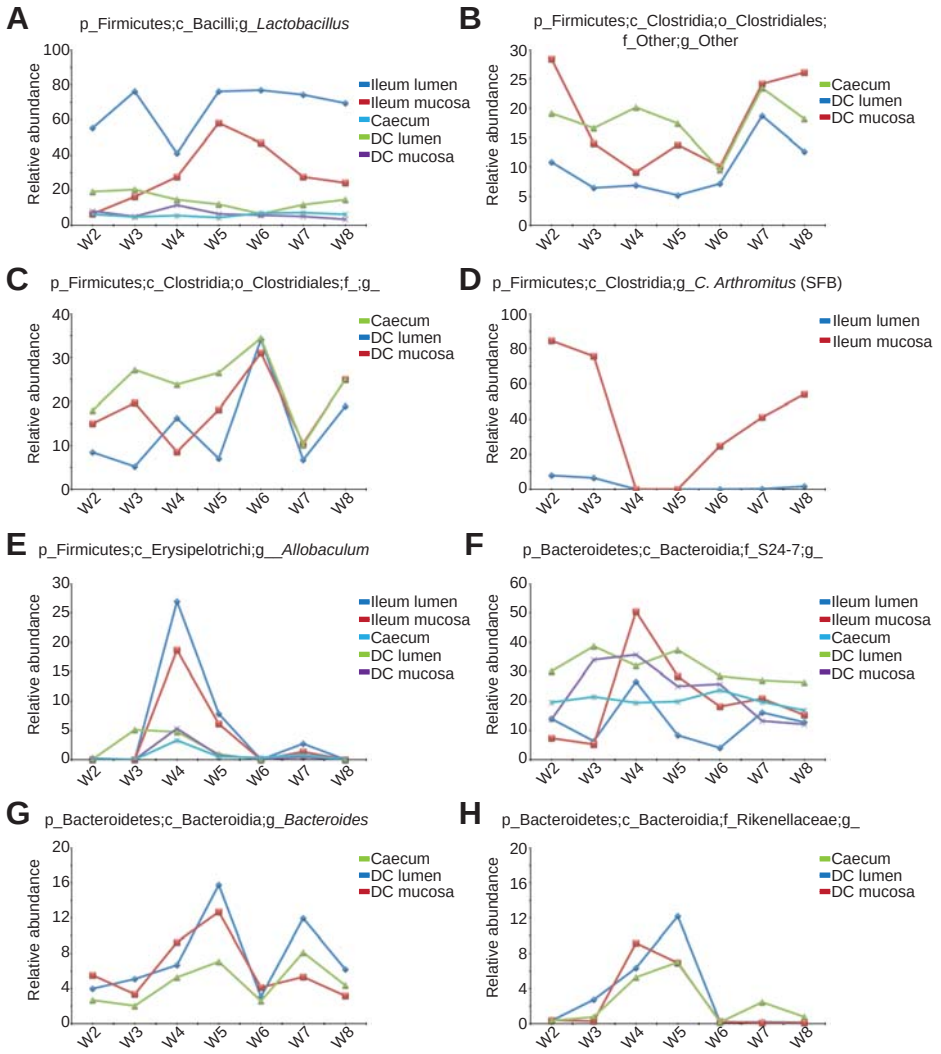
Supplementary Figure S1. Relative Muc2 expression analyzed by qPCR normalized to actin and GAPDH of colon samples during conventionalization. Convr mice were set to 1. Samples with an RQI value below 7.9 were excluded from analysis which excluded results from week 2. Supplemental to Fig. 4.



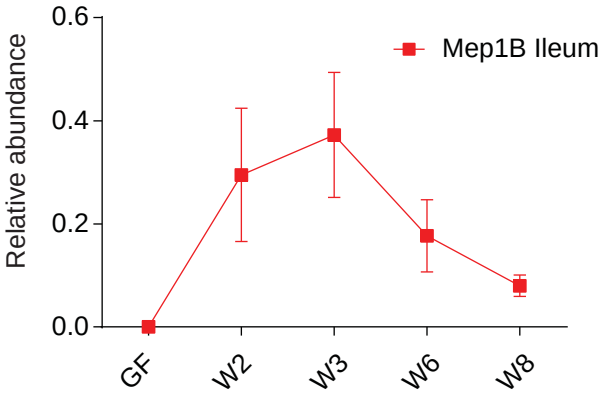
Supplementary Figure S2. The mean relative abundance (%) of different phyla and classes found along the gut during 2-8 weeks of colonization as well as in conventionally raised mice (Convr). (A-D) Ileum lumen and mucosa. (E-F) Caecum. (G-J) Distal colon lumen and mucosa. Supplement to Fig. 6.



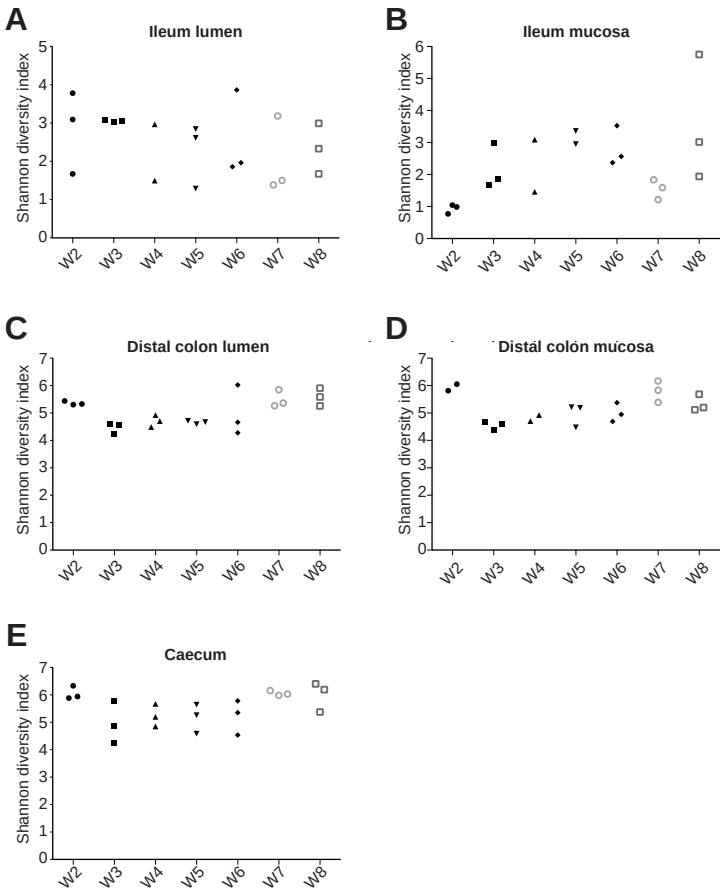
Supplementary Figure S3. The mean relative abundance (%) of the dominant phyla Firmicutes and Bacteroidetes in (A) ileum lumen, (B) caecum, and (C) distal colon lumen at 2-8 weeks following colonization. n=2-3 mice/time point. See Table S5, S7, and S9 for more details on exact number of mice per time point. Supplement to Fig. 6.



Supplementary Figure S4. The mean relative abundance (%) of a few representative genera within the gut at 2-8 weeks following colonization. (A) The genus *Lactobacillus* along the gut. (B) Other genus within the Firmicutes phylum. (C) Unknown genus within the Firmicutes phylum. (D) The genus *Candidatus Arthromitus* (Segmented filamentous bacteria). (E) The genus *Allobaculum* along the gut. (F) Unknown genus within the Bacteroidetes phylum. (G) The genus *Bacteroides* in caecum and distal colon. (H) Unknown genus within the Bacteroidetes phylum in caecum and distal colon. n=2-3/time point, see Tables S5-S9 for details regarding exact number of mice per time point. Supplement to Fig. 6.



Supplementary Figure S5. Relative abundance (mean ± SEM) in ileum of the protein meprin 1 beta (Mep1B) during colonization of germ free mice. The level of Mep1B in each sample was calculated as a proportion of the total abundance of the proteins identified. GF: germ free; W: week. Supplement to Fig. 4.



Supplementary Figure S6. Shannon diversity index in ileum (A-B), distal colon (C-D) and caecum at 2-8 weeks following colonization. 2-3 mice/time point. See Table S5-S9 for details regarding the exact number of mice per time point. Supplement to Fig. 6.

Supplementary Table S1. qPCR primers used.
Supplement to Experimental Procedure

Gene	Species	Forward Primer (5´- 3)	Reverse primer (5´- 3´)	Amplicon size	Annealing temp.	Intron spanning
GAPDH	Mouse	GGAGAAACCTGCCAAGTATG	GGAGTTGCTGTTGAAGTCG	124bp	61,4	primer
β-Actin	Mouse	AACGAGCGGTTCCGATGC	GTAGTTTCATGGATGCCACAGG	175bp	61,4	product
Muc2	Mouse	GACTGCCGAGACTCCTACAAG	GATGCACTCATGGTGGAGCTG	276bp	63,3	product
IL-1β	Mouse	TGCCACCTTTTGACAGTGATG	TGA TGT GCT GCT GCG AGA TT	138bp	63,3	primer

Supplementary Table S2. Relative amounts of proteins found in the mucus during colonization of GF mice. Supplement to Fig. 4.

See Separate Excel sheet.

Supplementary Table S3. Structure of *O* -linked oligosaccharides from the Muc2 mucin of mouse midcolon. Supplement to Fig. 5.

Label	Proposed oligosaccharide sequence/composition[a,b]	Animal			Observed ions	
		Germ-free	Colonization 3 weeks	Wild-type conv.	$[M-H^+]^-$	$[M-2H^+]^{2-}$
Neutral		+	+			
384	Gal-GalNAcol	+	+	+	384,1	
530	Fuc-Gal-GalNAcol	+	+	+	530,3	
587	Gal-(**HexNAc-6**)GalNAcol	+	+	+	587,2	
733	Fuc-Gal-(**HexNAc-6**)GalNAcol	+	+	+	733,3	
749	Gal-(**Gal-4GlcNAc-6**)GalNAcol		+	+	749,3	
895	Gal-(**Fuc-Gal-3/4GlcNAc-6**)GalNAcol	+	+	+	895,3	
1041	Fuc-Gal-(**Fuc-Gal-GlcNAc-6**)GalNAcol	+	+	+	1041,3	
One acidic residue						
667	Gal-(**SO_3^--GlcNAc-6**)GalNAcol	+	+	+	667,3	
675	NeuAc-Gal-GalNAcol	+	+	+	675,3	
813	Fuc-Gal-(**SO_3^--GlcNAc-6**)GalNAcol	+	+	+	813,3	
829	Gal-(**Gal-(SO_3^-)GlcNAc-6**)GalNAcol	+		+	829,1	
975	Gal-(**Fuc-Gal-(SO_3^--)GlcNAc-6**)GalNAcol		+	+	975,3	
1040	$[NeuAc]_1$ Gal-(**Gal-GlcNAc-6**)GalNAcol		+		1040,3	
1243a	$[NeuAc]_1$ Gal-(**HexNAc-Gal-GlcNAc-6**)GalNAcol	+	trace	trace	1243,4	621,5
1243b	$[NeuAc]_1$ HexNAc-Gal-3(**Gal-GlcNAc-6**)GalNAcol	+	trace	trace	1243,4	621,5
1899	$[NeuAc]_1$ HexNAc-Gal-GlcNAc-Gal-(**HexNAc-Gal-GlcNAc-6**)GalNAcol		+	+		949,5
Multiple acidic residues						
1120	NeuAc-Gal-(**SO_3^--Gal-GlcNAc-6**)GalNAcol		trace	+		559,8
1323	$[NeuAc]_1$ HexNAc-Gal-(**Gal-(SO_3^--)GlcNAc-6**)GalNAcol	+			1323,5	661,3
1469	Fuc-Gal-(**GalNAc-(NeuAc)Gal-(SO_3^--)GlcNAc6-**)GalNAcol[c]	+		+		734,2
1477	Gal-(**GlcNAc-(SO_3^--)Gal-GlcNAc-Gal-(SO_3^--)GlcNAc-6**)GalNAcol		+			738,3
1526	$[NeuAc]_1$ GlcNAc-(**SO_3^--Gal-GlcNAc-Gal-GlcNAc-6**)GalNAcol		+	+		762,8
1535	[NeuAc] Gal-(**GalNAc-(NeuAc)Gal-GlcNAc-6**)GalNAcol[c]	+	+	+		766,8
1565	$[NeuAc]_1$ Gal-(**SO_3^--Gal-GlcNAc-Gal-(SO_3^-)GlcNAc-6**)GalNAcol			+		782,3
1681	[NeuAc] Fuc-Gal-(**GalNAc-(NeuAc)Gal-GlcNAc-6**)GalNAcol[c]	+	+	+		839,9
1711	$[NeuAc]_1$ Fuc-Gal-(**SO_3^--Gal-GlcNAc-Gal-(SO_3^-)GlcNAc-6**)GalNAcol		+	+		855,5
1884a	$[NeuAc]_2$ $[Fuc]_1$ $[HexNAc]_3$ $[Gal]_2$ GalNAcol	+				941,5
1884b	[NeuAc] GalNAc-(NeuAc)Gal-GlcNAc-(**Fuc-Gal-4GlcNAc-6**)GalNAcol[c]	+		trace		941,5
2103	GalNAc-(NeuAc)Gal-GlcNAc-Gal-(**GalNAc-(NeuAc)Gal-GlcNAc-6**)GalNAcol[c]	+				1050,9
2126	[NeuAc] Fuc-Gal-(**GalNAc-(NeuAc)Gal-GlcNAc-Gal-(SO_3^-)GlcNAc-6**)GalNAcol[c]	+	+	+		1062,5
2249	[NeuAc] Fuc-Gal-GlcNAc-(**GalNAc-(NeuAc)Gal-GlcNAc-Gal-GlcNAc-6**)GalNAcol[c]	+				1124,0

[a] Proposed structures were interpreted from LC-MS[n] analysis data collected in the negative ion mode.

Assumptions: Hex residues are Gal, deoxyHex are Fuc, HexNAc within chain are GlcNAc, and HexNAcol are GalNAcol.

[b] Residues attached to C6 of GalNAcol are highlighted in bold.

[c]The NeuAc is based on glycosyltransferse data localized on the Gal together with GalNAc as in the Sd[a]/Cad epitope.

Supplementary Table S4. Abundance of glycosyltransferases found in Convr and GF mice as determined by proteomics. Supplement to Fig. 5.

	Proximal colon						Distal colon					
	Convr			GF			Convr			GF		
Protein name	Mean abundance[1]	RSD[2]	Count[3]	Mean abundance	RSD	Count	Mean abundance	RSD	Count	Mean abundance	RSD	Count
St6galnac6				58685	0,30	4	69431	0,72	5	88884	0,81	6
St6galnac2	3351	0,46	6				4592	0,81	4			
St6gal1	16530	0,40	7	6236	0,33	4	19406	0,24	8	5445	0,44	6
St3gal6	114253	0,37	7	57641	0,53	8	309723	0,37	9	197647	0,39	8
St3gal4	19784	0,55	9	27404	0,77	8	37200	0,44	8	53558	0,38	7
Gal3st2							19373	0,57				
Fut2	33437	1,00	8	62347	0,35	6	46142	1,12	3	67819	0,62	7
Chst4	16850	0,38	7				18030	0,18	7			
B4galnt2	269012	0,35	9	320339	0,30	8	72860	0,43	9	100273	0,22	8
B4galnt1	31454	0,66	7	58512	0,23	8				16196	0,22	5
B3gnt7	36757	0,86	7				257791	0,39	9	102275	0,47	8
B3gnt3	36742	0,62	4	58811	0,86	6	75364	0,61	8	65235	0,48	8
B3galt5	51809	0,58	9	91419	0,51	8	66770	0,31	9	113060	0,29	8
B4galt4	20143	0,70	5	12134	0,73	3						
B4galt1	27703	0,43	8				16795	0,74	9			
Gcnt3	295144	0,30	9	188909	0,18	8	103159	0,61	9	49264	0,51	8
C1galt1c1	210157	0,41	9	138282	0,36	8	93298	0,62	9	107057	0,43	8
C1galt1	249526	0,19	8	193224	0,27	8	115762	0,63	9	93028	0,31	8
Galnt12	59746	0,59	9	84443	1,00	6	51544	0,59	9	13709	0,38	7
Galnt10	15516	0,35	4				11332	1,31	3	3854	0,34	5
Galnt7	422633	0,35	9	370400	0,24	8	255888	0,44	9	229190	0,16	7
Galnt6	35615	0,56	9	19908	0,91	7	13317	0,39	9	15366	0,53	8
Galnt5	102431	0,81	9	62123	0,76	7	46789	0,83	7	23854	1,32	5
Galnt4	99581	0,34	9	51433	0,52	7	48556	0,25	9	34040	0,14	7
Galnt3	130580	0,61	9	106448	0,48	8	36376	0,60	8	59772	0,43	7
Galnt2	2099	0,38	4									
Galnt1	27517	0,42	9	27457	0,61	8	23675	0,34	8	29990	0,51	8

[1]Mean abundance, molecules/cell.

[2]Relative standard deviation.

[3]Amount of samples used for abundance calculation.

Supplementary Table 5. The mean relative abundance of dominant phyla, classes and genera in ileum lumen samples obtained from colonized mice. <1 means less than 0.0001. Supplement to Fig. 6.

See separate Excel sheet.