

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
23 December 2010 (23.12.2010)

PCT

(10) International Publication Number
WO 2010/147714 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2010/034254

(22) International Filing Date:
10 May 2010 (10.05.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/187,606 16 June 2009 (16.06.2009) US

(71) Applicant (for all designated States except US): **THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK** [US/US]; 412 Low Memorial Library, 535 West 116th Street, New York, NY 10027 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **LIPKIN, W., Ian** [US/US]; 45 West 105th Street, New York, NY 10025 (US). **HORNIG, Mady** [US/US]; 45 West 105th Street, New York, NY 10025 (US). **WILLIAMS, Brent, L.** [US/US]; 715 W 172nd St. -Apt. 22, New York, NY 10032 (US).

(74) Agents: **LOVE, Jane, M.** et al.; Wilmer Cutler Pickering Hale and Dorr LLP, 399 Park Avenue, New York, NY 10022 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) Title: AUTISM-ASSOCIATED BIOMARKERS AND USES THEREOF

(57) Abstract: Biomarkers for human autism are disclosed. Methods for treating, preventing, and diagnosing human autism and autism-related disorders are also disclosed.



WO 2010/147714 A1

AUTISM-ASSOCIATED BIOMARKERS AND USES THEREOF

[0001] This application claims the benefit of the filing date of U.S. Provisional Patent Application No. 61/187,606, filed June 16, 2009, the contents of which are hereby incorporated by reference.

[0002] This patent disclosure contains material that is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure as it appears in the U.S. Patent and Trademark Office patent file or records, but otherwise reserves any and all copyright rights.

[0003] This patent disclosure contains material that is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure as it appears in the U.S. Patent and Trademark Office patent file or records, but otherwise reserves any and all copyright rights.

GOVERNMENT SUPPORT

[0004] The work described herein was supported in whole, or in part, by National Institute of Health Grant No. U01 NS047537. Thus, the United States Government has certain rights to the invention.

BACKGROUND OF THE INVENTION

[0005] Autistic disorder is one of five pervasive developmental disorders defined in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision DSM-IV-TR (2000). Autistic disorder is a developmental disorder of the human brain that manifests during infancy or childhood and is characterized by behavioral and social abnormalities that appear to be developmentally based (for example, impairments in social interaction and communication). In addition, autism interferes with imagination and the ability to reason. Autism is frequently associated with other disorders such as attention deficit/hyperactivity disorder (AD/HD) and can be associated with psychiatric symptoms such as anxiety and depression. In the last decade, autism diagnoses have increased by 300% to 500% in the United States and many other countries. A means of prevention and treatment is needed for this health

crisis that addresses the underlying mechanisms leading to the development of autism versus those that merely address the symptoms.

[0006] Pervasive developmental disorders (PDDs) are also part of the Autism Spectrum Disorders (ASDs). PDD is used to categorize children who do not meet the strict criteria for Autistic Disorder but who come close, either by manifesting atypical autism or by nearly meeting the diagnostic criteria in two or three of the key areas. Some of these children meet criteria for the ASD known as Asperger's Disorder (ASP), wherein language capacities are relatively spared compared to children with Autistic Disorder. Others meet criteria for the PDDs known as Childhood Disintegrative Disorder, which begins at a slightly later age than the other ASDs, or Rett's Disorder, which is related to a mutation in a DNA methylation binding protein gene called MeCP2 and usually occurs in girls.

[0007] Many children with autism have gastrointestinal (GI) disturbances that affect their quality of life. Although some of these children have been investigated through GI immunopathology, molecular studies are lacking that characterize host gene expression or survey microflora using pyrosequencing methods.

SUMMARY OF THE INVENTION

[0008] The invention is based, at least in part, on the finding that decreased levels in sucrase isomaltase, maltase glucoamylase, lactase, GLUT2, and SGLT1 can serve as markers for human Autism Spectrum Disorders. Accordingly, in one aspect, the invention provides a method for detecting the presence of or a predisposition to autism or an autism spectrum disorder (ASD) in a human subject or a child of a human subject. The method comprises: (1) obtaining a biological sample from a human subject; and (2) detecting whether or not there is an alteration in the expression of a carbohydrate metabolic enzyme protein or a carbohydrate transporter protein in the subject as compared to a non-autistic subject. In one embodiment, the carbohydrate metabolic enzyme comprises sucrase isomaltase, maltase glucoamylase, lactase, or a combination thereof. In another embodiment, the carbohydrate transporter comprises GLUT2, SGLT1, or a combination thereof. In some embodiments, the method further comprises detecting a decrease in Bacteroidetes, an increase in the Firmicute/Bacteroidete ratios, an increase in cumulative levels of Firmicutes and Proteobacteria, an increase in Beta-

proteobacteria, and an increase in *Sutterella* sp. in the small or large intestine of the subject. In one embodiment, the detecting comprises detecting whether there is an alteration in the gene locus that encodes the carbohydrate metabolic enzyme protein or the carbohydrate transporter protein. In a further embodiment, the detecting comprises detecting whether expression of the carbohydrate metabolic enzyme protein or the carbohydrate transporter protein is reduced. In some embodiments, the detecting comprises detecting in the sample whether there is a reduction in the mRNA expression of the carbohydrate metabolic enzyme protein or the carbohydrate transporter protein. In some embodiments of the invention, the subject is a human embryo, a human fetus, or an unborn human child. In other embodiments, the sample comprises blood, serum, sputum, lacrimal secretions, semen, vaginal secretions, fetal tissue, skin tissue, small intestine tissue (e.g., the ileum), large intestine tissue (e.g., the cecum), muscle tissue, amniotic fluid, or a combination thereof.

[0009] An aspect of the invention provides a method for treating or preventing autism or an autism spectrum disorder in a subject in need thereof. The method comprises administering to the subject a therapeutic amount of a pharmaceutical composition comprising a functional carbohydrate metabolic enzyme molecule or a carbohydrate transporter molecule, thereby treating or preventing autism or an autism spectrum disorder. In a further embodiment, the administering comprises a subcutaneous, intra-muscular, intra-peritoneal, or intravenous injection; an infusion; oral, nasal, or topical delivery; or a combination of the delivery modes described. In some embodiments, the administering comprises delivery of a carbohydrate metabolic enzyme molecule or a carbohydrate transporter molecule to the alimentary canal or intestine of the subject. In other embodiments, the administering comprises feeding the human subject or child thereof a therapeutically effective amount of the carbohydrate metabolic enzyme molecule or a carbohydrate transporter molecule. In further embodiments, the administering occurs daily, weekly, twice weekly, monthly, twice monthly, or yearly.

[0010] In other aspects, the invention provides for a pharmaceutical composition comprising: a carbohydrate metabolic enzyme molecule or a carbohydrate transporter molecule molecule; and a pharmaceutically acceptable carrier.

[0011] An aspect of the invention provides for an isolated nucleic acid composition. In one embodiment, the composition comprises a nucleic acid molecule having at least about 80% identity to SEQ ID NO: 11, 12, 13, or 14. In one embodiment, the composition comprises a nucleic acid molecule having at least about 85% identity to SEQ ID NO: 11, 12, 13, or 14. In one embodiment, the composition comprises a nucleic acid molecule having at least about 90% identity to SEQ ID NO: 11, 12, 13, or 14. In one embodiment, the composition comprises a nucleic acid molecule having at least about 95% identity to SEQ ID NO: 11, 12, 13, or 14. In one embodiment, the composition comprises a nucleic acid molecule having at least about 98% identity to SEQ ID NO: 11, 12, 13, or 14. In one embodiment, the composition comprises a nucleic acid molecule having at least about 99% identity to SEQ ID NO: 11, 12, 13, or 14. In one embodiment, the composition is SEQ ID NO: 11, 12, 13, or 14.

[0012] An aspect of the invention provides for a diagnostic kit for detecting the presence of *Sutterella sp.* in a sample. In one embodiment, the kit comprises a nucleic acid molecule that specifically hybridizes to or a primer combination that amplifies a *Sutterella sp.* 16S nucleic acid sequence. In one embodiment, the nucleic acid molecule comprises a nucleic acid primer or nucleic acid probe. In another embodiment, the 16S nucleic acid sequence comprises at least about 80% of SEQ ID NO: 59 or SEQ ID NO: 60. In some embodiments, the 16S nucleic acid sequence comprises at least about 85% of SEQ ID NO: 59 or SEQ ID NO: 60. In further embodiments, the 16S nucleic acid sequence comprises at least about 90% of SEQ ID NO: 59 or SEQ ID NO: 60. In other embodiments, the 16S nucleic acid sequence comprises at least about 95% of SEQ ID NO: 59 or SEQ ID NO: 60. In another embodiment, the 16S nucleic acid sequence comprises at least about 98% of SEQ ID NO: 59 or SEQ ID NO: 60. In some embodiments, the 16S nucleic acid sequence comprises at least about 99% of SEQ ID NO: 59 or SEQ ID NO: 60. In further embodiments, the 16S nucleic acid sequence is SEQ ID NO: 59 or SEQ ID NO: 60. In one embodiment, the probe comprises a nucleotide sequence having SEQ ID NOS: 13 or 14 in Table 1, or the italicized nucleotide of sequence SEQ ID NO: 19. In a further embodiment, the probe comprises at least 10 consecutive nucleotide bases comprising SEQ ID NO: 19, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C nucleotide. In some embodiments, the probe comprises a reverse complement of SEQ ID NOS: 11, 12, 15,

16, 17, 18, or 19, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C nucleotide. In other embodiments, the primer comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS: 11, 12, 15, 16, 17, or 18, wherein, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C nucleotide. In one embodiment, the sample is from a human or non-human animal. In other embodiments, the sample comprises intestinal tissue (e.g., the small intestine or large intestine), feces, blood, skin, or a combination of the mentioned tissues.

[0013] An aspect of the invention provides for a diagnostic kit for determining whether a sample from a subject exhibits a presence of or a predisposition to autism or an autism spectrum disorder (ASD). In one embodiment, the kit comprising a nucleic acid primer that specifically hybridizes to an autism biomarker, wherein the primer will prime a polymerase reaction only when an autism biomarker is present. In another embodiment, the primer comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS: 11, 12, 15, 16, 17, or 18, wherein, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C nucleotide. In some embodiments, the autism biomarker is a carbohydrate transporter molecule, a carbohydrate metabolic enzyme molecule, or a gastrointestinal *Sutterella sp.* bacterium. In a further embodiment, the carbohydrate transporter molecule is GLUT2 or SGLT1. In other embodiments, the carbohydrate metabolic enzyme molecule is SI, MGAM, or LCT. In one embodiment, the sample is from a human or non-human animal. In other embodiments, the sample comprises intestinal tissue (e.g., the small intestine or large intestine), feces, blood, skin, or a combination of the mentioned tissues.

[0014] An aspect of the invention provides for a method of treating or preventing a disease associated with elevated levels of Beta-proteobacteria. The method of the invention comprises administering to a subject in need thereof a therapeutic amount of an antimicrobial composition

effective against Beta-proteobacteria for treating the disease. In one embodiment, the antimicrobial composition is an antibiotic, a probiotic agent, or a combination thereof. In another embodiment, the disease is ASD, autism, or a gastrointestinal disease. In a further embodiment, the gastrointestinal disease is diarrhea, inflammatory bowel disease, antimicrobial-associated colitis, or irritable bowel syndrome. In some embodiments, the diarrhea or inflammatory bowel diseases is ulcerative colitis or Crohn's disease. In one embodiment, the antibiotic comprises lincosamides, chloramphenicols, tetracyclines, aminoglycosides, beta-lactams, vancomycins, bacitracins, macrolides, amphotericins, sulfonamides, methenamin, nitrofurantoin, phenazopyridine, trimethoprim; rifampicins, metronidazoles, cefazolins, lincomycin, spectinomycin, mupirocins, quinolones, novobiocins, polymixins, gramicidins, antipseudomonals, or a combination of the stated antibiotics. In another embodiment of the invention, the probiotic agent comprises *Bacteroides*, *Prevotella*, *Porphyromonas*, *Fusobacterium*, *Sutterella*, *Bilophila*, *Campylobacter*, *Wolinella*, *Butyrovibrio*, *Megamonas*, *Desulfomonas*, *Desulfovibrio*, *Bifidobacterium*, *Lactobacillus*, *Eubacterium*, *Actinomyces*, *Eggerthella*, *Coriobacterium*, *Propionibacterium*, other genera of non-sporeforming anaerobic gram-positive bacilli, *Bacillus*, *Peptostreptococcus*, newly created genera originally classified as *Peptostreptococcus*, *Peptococcus*, *Acidaminococcus*, *Ruminococcus*, *Megasphaera*, *Gaffkya*, *Coprococcus*, *Veillonella*, *Sarcina*, *Clostridium*, *Aerococcus*, *Streptococcus*, *Enterococcus*, *Pediococcus*, *Micrococcus*, *Staphylococcus*, *Corynebacterium*, species of the genera comprising the Enterobacteriaceae and Pseudomonadaceae, or a combination of the listed probiotic agents.

[0015] An aspect of the invention provides for a method of detecting a *Sutterella sp.* in a sample. The method comprises: (a) selecting a *Sutterella sp.*-specific primer pair, wherein the primer pair mediates amplification of a polynucleotide amplicon of a selected, known length from a nucleic acid of a *Sutterella sp.*; contacting a nucleic acid from the sample with the *Sutterella sp.*-specific primer pair in a reaction mixture under conditions that promote amplification of a polynucleotide amplicon, wherein the primer pair will prime a polymerase reaction only when the nucleic acid of a *Sutterella sp.* is present; and detecting the amplicons, wherein the detection of an amplicon of a selected, known length is indicative of the sample containing the nucleic acid of a *Sutterella sp.* In one embodiment, the sample comprises intestinal tissue (e.g., the small intestine or large intestine), feces, blood, skin, or a combination of the listed tissues. In one embodiment, the primer pair comprises a forward primer and a reverse primer. In some

embodiments, the forward primer comprises SEQ ID NO: 11 or 17, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C nucleotide. In other embodiments, the reverse primer comprises SEQ ID NO: 12 or 18, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C nucleotide. In further embodiments, the forward primer comprises at least 10 consecutive nucleotide bases comprising SEQ ID NO: 17 or 19, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C. In some embodiments, the reverse primer comprises at least 10 consecutive nucleotide bases comprising SEQ ID NO: 18 or 19, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C nucleotide, wherein B is a T nucleotide, C nucleotide, or G nucleotide, wherein V is an A nucleotide, G nucleotide, or C nucleotide; wherein D is an A nucleotide, G nucleotide, or T nucleotide; and wherein K is a G nucleotide or T nucleotide.

BRIEF DESCRIPTION OF THE FIGURES

[0016] **FIG. 1** is a schematic depicting carbohydrate metabolizing enzymes (e.g., sucrase isomaltase, maltase glucoamylase, and lactase) and carbohydrate transporter proteins (e.g., GLUT2 and SGLT1) involved in carbohydrate metabolism, uptake, and absorption in the enterocytes of the ileum.

[0017] **FIG. 2** shows bar graphs depicting that carbohydrate metabolizing enzyme mRNAs are reduced in the ileum of ASD subjects. Graphs are shown for sucrase isomaltase (left), maltase glucoamylase (center), and lactase (right).

[0018] **FIG. 3** shows bar graphs depicting that carbohydrate transporter mRNAs are reduced in the ileum of ASD subjects. Graphs are shown for SGLT1 (Top) and GLUT2 (Bottom).

[0019] **FIG. 4** shows graphs depicting that mRNA for ileal inflammatory markers are increased in the ileum of ASD subjects. Graphs are shown for C1QA (Top Left), Resistin (Top Right), and IL17F (Bottom Left and Right).

[0020] **FIG. 5** shows bar graphs depicting the differences in bacteria phylum found in the ileum of ASD subjects. Changes at the phylum level were observed. Bar graphs show a decrease in Bacteroidetes (left) and increase in Firmicute/Bacteroidete ratios in ileum of AUT-GI children.

[0021] **FIG. 6** is a bar graph depicting the copy number of bacteroidetes found in the ileum of ASD subjects. Real-time PCR confirmed a decrease in Bacteroidete. Bacteroidete 16S rDNA copies (Normalized to Total Bacterial 16S rDNA).

[0022] **FIG. 7** is a schematic summarizing the interplay between expression levels of carbohydrate metabolic enzymes (e.g., sucrase isomaltase, maltase glucoamylase, and lactase), carbohydrate transporters (e.g., GLUT2 and SGLT1) and the population of bacteria in the ileum of ASD subjects.

[0023] **FIGS. 8A-B** are bar graphs showing the abundance of *Sutterella sp.* in the ileum (**FIG. 8A**) and cecum (**FIG. 8B**) of autism and control patients.

[0024] **FIGS. 8C-D** are bar graphs showing the abundance of *Sutterella sp.* sequences in the ileum (**FIG. 8C**) and cecum (**FIG. 8D**) of autism and control patients.

[0025] **FIGS. 8E-F** are bar graphs showing the abundance of *Sutterella sp.* sequences comprising the Beta-proteobacteria sequences in the ileum (**FIG. 8C**) and cecum (**FIG. 8D**) of autism and control patients.

[0026] **FIG. 9** is a photograph of an agarose gel showing the results of classical PCR experiments for the detection of *Sutterella*.

[0027] **FIG. 10A** is an amplification plot of *Sutterella sp.* through cycles of Real-time PCR experiments.

[0028] **FIG. 10B** is a standard curve graph showing the copy number of *Sutterella sp.* from Real-time PCR experiments.

[0029] **FIG. 11** is a photograph of an agarose gel showing the results of *Sutterella* detection in the ileum and cecum of patients using the V6-V8 *Sutterella sp.*-specific PCR.

[0030] **FIGS. 12A-B** are bar graphs showing the copy number of *Sutterella sp.* in the ileum (**FIG. 12A**) and cecum (**FIG. 12B**) of autism and control patients using the V6-V8 *Sutterella sp.*-specific PCR.

[0031] **FIGS. 12C-D** are bar graphs showing the copy number of *Sutterella sp.* in the ileum (**FIG. 12C**) and cecum (**FIG. 12D**) of autism and control patients using the V6-V8 *Sutterella sp.*-specific PCR.

[0032] **FIG. 13** is a sequence alignment for the V6-V8 region of *Sutterella sp.* obtained from biological samples of Autism patients 1, 3, 10, 11, and 12 (SEQ ID NO: 59), and Autism patients 5 and 7 (SEQ ID NO: 60).

[0033] **FIG. 14** depicts *Sutterella sp.* sequence clustering from the Operational Taxonomic Unit (OTU) analysis of V2 pyrosequencing reads.

[0034] **FIG. 15A** is a schematic depicting *Sutterella sp.* treeing analysis of the V6-V8 sequences.

[0035] **FIG. 15B** is a schematic depicting *Sutterella sp.* treeing analysis of the V2 sequence.

[0036] **FIG. 16** shows graphs that show quantitative real-time PCR analysis of disaccharidases, hexose transporters, villin and CDX2. Box-and-whisker plots displaying (**FIG. 16A**) SI (Mann-Whitney; $p=0.001$), (**FIG. 16B**) MGAM (Mann-Whitney; $p=0.003$), (**FIG. 16C**) LCT (Mann-Whitney; $p=0.032$), (**FIG. 16D**) SGLT1 (Mann-Whitney; $p=0.008$), (**FIG. 16E**) GLUT2 (Mann-Whitney; $p=0.010$), (**FIG. 16F**) Villin (Mann-Whitney; $p=0.307$), and (**FIG.**

16G) CDX2 (Mann-Whitney; $p=0.192$) mRNA expression relative to GAPDH mRNA in ileal biopsies from AUT-GI (AUT) and Control-GI (Control) patients. Box-and-whisker plots show the median and the interquartile (midspread) range (boxes containing 50% of all values), the whiskers (representing the 25th and 75th percentiles) and the extreme data points (open circles). *, $p < 0.05$, **, $p < 0.01$.

[0037] **FIG. 17** shows graphs depicting pyrosequencing analysis of intestinal microbiota. **(FIGS. 17A-B)** Phylum-level comparison of the average relative abundance of bacterial taxa in ileal **(FIG. 17A)** and cecal **(FIG. 17B)** biopsies from AUT-GI and Control-GI patients. **(FIGS. 17C-D)** Box-and-whisker plot displaying the distribution of Bacteroidetes as a percentage of total bacterial 16S rRNA V2 pyrosequencing reads from ileal **(C; Mann-Whitney, $p=0.012$)** and cecal **(FIG. 17D; Mann-Whitney, $p=0.008$)** biopsies from AUT-GI and Control-GI patients. **(FIGS. 17E-F)** Bacteroidete-specific quantitative real-time PCR analysis of ileal **(FIG. 17E; Mann-Whitney, $p=0.003$)** and cecal **(FIG. 17F; Mann-Whitney, $p=0.022$)** biopsies from AUT-GI and Control-GI patients. **(FIGS. 17G-H)** Heatmaps displaying abundance distributions (% of total sequence reads per patient) of Bacteroidetes classified at the family level in ileal **(FIG. 17G)** and cecal **(FIG. 17H)** biopsies from AUT-GI and Control-GI children (Bottom row displays cumulative levels of all family members by patient). *, $p < 0.05$, **, $p < 0.01$.

[0038] **FIG. 18** shows graphs of Firmicute abundance in AUT-GI and Control-GI children. **(FIGS. 18A-18B)** Box-and-whisker plots displaying the Firmicute/Bacteroidete ratio from pyrosequencing reads obtained from ileal **(FIG. 18A; Mann-Whitney, $p=0.026$)** and cecal **(FIG. 18B; Mann-Whitney, $p=0.032$)** biopsies of AUT-GI and Control-GI patients. **(FIGS. 18C-18D)** Box-and-whisker plots displaying the cumulative levels of members of the families Lachnospiraceae and Ruminococcaceae in ileal **(FIG. 18C; Mann-Whitney; $p=0.062$)** and cecal **(FIG. 18D; Mann-Whitney; $p=0.098$)** biopsies from AUT-GI and Control-GI children. **(FIGS. 18E-18F)** Heatmaps displaying abundance distribution (% of total sequence reads per patient) of family members in the class Clostridia in ileum **(FIG. 18E)** and cecum **(FIG. 18F)** of AUT-GI and Control-Gi children (Bottom row displays cumulative levels of all family members by patient). **(FIGS. 18G-18H)** Box-and-whisker plots displaying the cumulative abundance of Firmicutes and Proteobacteria from ileal **(FIG. 18G; Mann-Whitney, $p=0.015$)** and cecal **(FIG. 18H; Mann-Whitney, $p=0.007$)** biopsies from AUT-GI and Control-GI patients. **(FIGS. 18I-18J)**

Heatmaps displaying the abundance distribution (% of total sequence reads per patient) of Firmicutes and Proteobacteria by patient in ilea (**FIG. 18I**) and ceca (**FIG. 18J**) of AUT-GI and Control-GI children (Bottom row displays cumulative levels of Firmicutes and Proteobacteria by patient). *, $p < 0.05$, **, $p < 0.01$, †, $p < 0.1$ (trend).

[0039] **FIG. 19** shows graphs of the abundance of Proteobacteria in AUT-GI and Control-GI children. (**FIGS. 19A-19B**) Box-and-whisker plots displaying the phyla level abundance of Proteobacteria members in ilea (**FIG. 19A**; Mann-Whitney, $p=0.549$) and ceca (**FIG. 19B**; Mann-Whitney, $p=0.072$) of AUT-GI and Control-GI children obtained by pyrosequencing. (**FIGS. 19C-19D**) Box-and-whisker plots displaying the class level abundance of Betaproteobacteria members in ilea (**FIG. 19C**; Mann-Whitney, $p=0.072$) and ceca (**FIG. 19D**; $p=0.038$) of AUT-GI and Control-GI children. (**FIGS. 19E-19F**) Heatmaps displaying the abundance distribution (% of total sequence reads per patient) of family members within the classes Alpha-, Beta-, and Gammaproteobacteria in the ilea (**FIG. 19E**) and ceca (**FIG. 19F**) of AUT-GI and Control-GI children (Bottom row of each heatmap displays the cumulative levels of family members in each class by patient). *, $p < 0.05$, †, $p < 0.1$ (trend).

[0040] **FIG. 20** shows schematics depicting factors that mediate GI disease in AUT-GI children. (**FIG. 20A**) Schematic representation of enterocyte-mediated digestion of disaccharides and absorption/transport of monosaccharides in the small intestine. Disaccharidase enzymes (SI, MGAM, and LCT) in the enterocyte brush border break down disaccharides into their component monosaccharides. The monosaccharides, glucose and galactose, are transported from the small intestinal lumen into the enterocyte by the sodium-dependent transporter SGLT1. On the basolateral enterocyte membrane, the facilitative transporter, GLUT2, transports glucose, galactose and fructose out of the enterocyte and into the circulation, thus regulating postprandial blood glucose levels. GLUT2 may also be transiently inserted into the apical enterocyte membrane, contributing a diffusive component to monosaccharide absorption in certain circumstances (Kellet et al., 2008). The expression levels of disaccharidases and hexose transporters may be controlled by the transcription factor CDX2. (**FIG. 20B**) In the normal small intestine, where expression of disaccharidases and hexose transporters are high, the majority, if not all, of disaccharides are efficiently digested and monosaccharides are absorbed

from the lumen. Thus, only complex polysaccharides reach the large intestine and serve as growth substrates for colonic bacteria. Those bacteria best suited for growth on polysaccharides (i.e., Bacteroidetes) outcompete other bacteria and dominate the colonic space. In the normal intestine, colonic (i.e., cecal) microbial community structure may be kept within a normal homeostatic range by the level of expression of disaccharidases and hexose transporters upstream in the small intestine. The constraint on bacterial structure regulated by ileal gene expression would constrain bacterial byproducts of fermentation such as SCFAs, and limit the growth of potential pathogens. **(FIG. 20C)** In the AUT-GI intestine, where expression of disaccharidases and hexose transporters are deficient, mono- and disaccharides accumulate in the lumen of the distal small intestine (ileum) and proximal colon (cecum), and may exert extraintestinal effects by reducing postprandial blood glucose. The presence of additional carbohydrate substrates in the lumen abrogates the growth advantage of bacteria best suited for growth on polysaccharides (i.e., Bacteroidetes) and promotes the growth of other bacteria. In ASD-GI this specifically manifests as an increase in Firmicute/Bacteroidete ratios, cumulative levels of Firmicutes and Proteobacteria, and in levels of Betaproteobacteria in both the ileum and cecum. The level of dysbiosis in the ileum and cecum may thus be controlled by the degree and type of deficiency of carbohydrate metabolism and transport in the small intestine. Within the intestine, malabsorbed monosaccharides can lead to osmotic diarrhea; non-absorbed sugars may also serve as substrates for intestinal microflora, that produce fatty acids and gases (methane, hydrogen, and carbon dioxide) and promote additional GI symptoms of bloating and flatulence. Additional effects of dysbiosis may manifest in changes in SCFAs that can reduce colonic pH, further inhibiting the growth of Bacteroidetes. Disruption of symbiotic relationships between the host and the intestinal microbial ecosystem as a result of dysbiosis may also play a fundamental role in development, distribution, activation and differentiation of immune cells within the intestine (Abt and Artis, 2009; Mazmanian et al., 2008), thus providing a framework for understanding previous reports of inflammatory indices in the AUT-GI intestine.

[0041] **FIG. 21** depicts lactase genotyping. **(FIG. 21A)** Representative agarose gel banding patterns observed for LCT-13910 and LCT-22018 polymorphisms. **(FIG. 21B)** Distribution of genotypes for 13910 and 22018 polymorphisms between AUT-GI (n=15) and Control-GI (n=7) patients. **(FIG. 21C)** Box-and-whisker plot displaying the distribution of LCT mRNA expression in all individuals (AUT-GI and Control-GI) with the homozygous adult-type hypolactasia

genotype (13910-C/C; 22018-G/G) compared to all individuals (AUT-GI and Control-GI) possessing at least one copy of the normal allele (13910-C/T; 22018-G/A and 13910-T/T; 22018-A/A); Mann-Whitney, $p=0.033$. (**FIG. 21D**) Distribution of LCT mRNA expression levels split by genotype and group (AUT-GI and Control-GI); Kruskal-Wallis, $p=0.097$. (**FIG. 21E**) Distribution of LCT mRNA expression for all patients possessing at least one copy of the normal (lactase persistence) allele for AUT-GI ($n=12$) and Control-GI ($n=6$); Mann-Whitney, $p=0.0246$. *, $p < 0.05$.

[0042] **FIG. 22** shows graphs depicting villin ratios. Disaccharidase or transporter mRNA/villin mRNA ratio for SI (**FIG. 22A**; Mann-Whitney, $p=0.001$), MGAM (**FIG. 22B**; Mann-Whitney, $p=0.001$), LCT (**FIG. 22C**; Mann-Whitney, $p=0.005$), SGLT1 (**FIG. 22D**; Mann-Whitney, $p=0.0008$), and GLUT2 (**FIG. 22E**; Mann-Whitney, $p=0.002$). *, $p < 0.05$, **, $p < 0.01$, ***, $p < 0.001$.

[0043] **FIG. 23** shows graphs of the diversity of AUT-GI and Control-GI phylotypes. (**FIGS. 23A-23B**) Rarefaction curves assessing the completeness of sampling from pyrosequencing data obtained for individual AUT-GI (red) and Control-GI (blue) subjects' ileal (**FIG. 23A**) and cecal (**FIG. 23B**) biopsies. The y-axis indicates the number of OTUs detected (defined at 97% threshold for sequence similarity), the x-axis the number of sequences sampled. (**FIGS. 23C-23D**) Rarefaction curves to estimate phylotype diversity, using the Shannon Diversity Index, from pyrosequencing data obtained for individual AUT-GI (red) and Control-GI (blue) subjects' ileal (**FIG. 23C**) and cecal (**FIG. 23D**) biopsies.

[0044] **FIG. 24** shows graphs depicting the distribution of pyrosequencing reads by patient. (**FIGS. 24A-24B**) Phylum level distribution of bacteria by patient obtained from 16S rRNA gene barcoded pyrosequencing for ilea (**FIG. 24A**) and ceca (**FIG. 24B**). (**FIGS. 24C-D**) Distribution of low abundance bacterial phyla obtained by barcoded pyrosequencing. By-patient distribution of low abundance bacterial phyla in ilea (**FIG. 24C**) and ceca (**FIG. 24D**) from AUT-GI (patients 1-15) and Control-GI (patients 16-22).

[0045] **FIG. 25** shows the OTU analysis of Bacteroidete phylotypes. (**FIGS. 25A-25B**) Heatmaps displaying abundance distributions (% of total sequence reads per patient) of the 12 most abundant Bacteroidete OTUs (defined at 97% threshold) in ileal (**FIG. 25A**) and cecal

(**FIG. 25B**) biopsies from AUT-GI and Control-GI children (Bottom row displays cumulative levels of all 12 OTUs by patient). (**FIGS. 25C-25D**) Box-and-whisker plots displaying the cumulative abundance of the 12 OTUs in ilea (**FIG. 25C**; Mann-Whitney, $p=0.008$) and ceca (**FIG. 25D**; Mann-Whitney, $p=0.008$) of AUT-GI and Control-GI children. (**FIG. 25E**) Greengenes- or microbial blast(*)-derived classification of representative sequences obtained from each Bacteroidete OTU. Color code denotes the family-level, Ribosomal Database-derived taxonomic classification of each OTU sequence.

[0046] **FIG. 26** shows graphs depicting order-level analysis of Firmicute/Bacteroidete ratio and confirmation by real-time PCR. (**FIGS. 26A-26B**) Box-and-whisker plot displaying the order-level distribution of the Clostridiales/Bacteroidales ratio from pyrosequencing reads obtained from ileal (**FIG. 26A**; Mann-Whitney, $p=0.012$) and cecal (**FIG. 26B**; Mann-Whitney, $p=0.032$) biopsies from AUT-GI and Control-GI patients. (**FIGS. 26C-26D**) Box-and-whisker plot displaying the Firmicute/Bacteroidete ratios obtained by real-time PCR for ilea (**FIG. 26C**; Mann-Whitney, $p=0.0006$) and ceca (**FIG. 26D**; Mann-Whitney, $p=0.022$) of AUT-GI and Control-GI children. *, $p < 0.05$, ***, $p < 0.001$.

[0047] **FIG. 27** shows graphs of the abundance of Firmicutes assayed by pyrosequencing and real-time PCR. (**FIGS. 27A-27B**) Box-and-whisker plots displaying the phyla level abundance of Firmicutes in the ilea (**FIG. 27A**; Mann-Whitney, $p=0.098$) and ceca (**FIG. 27B**; Mann-Whitney, $p=0.148$) of AUT-GI and Control-GI children obtained by pyrosequencing. (**FIGS. 27C-27D**) Box-and-whisker plots displaying the phyla level abundance of Firmicutes in the ilea (**FIG. 27C**; Mann-Whitney, $p=0.245$) and ceca (**FIG. 27D**; Mann-Whitney, $p=0.053$) of AUT-GI and Control-GI children obtained by real-time PCR. (**FIGS. 27E-27F**) Box-and-whisker plots displaying the abundance of Clostridiales from ileal (**FIG. 27E**; Mann-Whitney, $p=0.072$) and cecal (**FIG. 27F**; Mann-Whitney, $p=0.098$) biopsies from AUT-GI and Control-GI patients obtained by pyrosequencing.

[0048] **FIG. 28** shows genus-level distribution of members of the families *Ruminococcaceae* and *Lachnospiraceae*. (**FIGS. 28A-28B**) Heatmap representation of the individual patient distributions of *Ruminococcaceae* and *Lachnospiraceae* genus members in ileal (**FIG. 28A**) and cecal (**FIG. 28B**) biopsies from AUT-GI (Patients 1-15) and Control-GI

(Patients 16-22) patients. *, genus members contributing to the trend toward increased Firmicutes in AUT-GI children.

[0049] **FIG. 29** shows graphs depicting increases in inflammatory markers, such as C1Q, Resistin, CD163, Tweak, IL17F, and nNOS. These inflammatory markers may also serve as biomarkers for diagnosis of human Autism Spectrum Disorders, as well as for detecting the presence of or a predisposition to autism or an autism spectrum disorder.

DETAILED DESCRIPTION OF THE INVENTION

[0050] Autism, one of the ASDs, is mostly diagnosed clinically using behavioral criteria because few specific biological markers are known for diagnosing the disease. Autism is a neuropsychiatric developmental disorder characterized by impaired verbal communication, non-verbal communication, and reciprocal social interaction. It is also characterized by restricted and stereotyped patterns of interests and activities, as well as the presence of developmental abnormalities by 3 years of age (Bailey et al., (1996) J Child Psychol Psychiatry 37(1):89-126). Autism-associated disorders, diseases or pathologies can comprise any metabolic, immune or systemic disorders; gastrointestinal disorders; epilepsy; congenital malformations or genetic syndromes; anxiety, depression, or AD/HD; or speech delay and motor in-coordination.

[0051] Autism spectrum disorders (ASD) are defined by impairments in verbal and non-verbal communication, social interactions, and repetitive and stereotyped behaviors (DSM-IV-TR criteria, American Psychiatric Association, 2000). In addition to these core deficits, previous reports indicate that the prevalence of gastrointestinal symptoms ranges widely in individuals with ASD, from 9 to 91% (Buie et al., 2010). Macroscopic and histological observations in ASD include findings of ileo-colonic lymphoid nodular hyperplasia (LNH), enterocolitis, gastritis and esophagitis (Wakefield et al., 2000; Wakefield et al., 2005; Furlano et al., 2001; Torrente et al., 2002; Horvath et al., 1999). Associated changes in intestinal inflammatory parameters include higher densities of lymphocyte populations, aberrant cytokine profiles, and deposition of immunoglobulin (IgG) and complement C1q on the basolateral enterocyte membrane (Furlano et al., 2001; Ashwood and Wakefield, 2006). Functional disturbances include increased intestinal permeability (D'Eufemia et al., 1996), compromised sulfoconjugation of phenolic compounds (O'Reilly and Waring, 1993; Alberti et al., 1999), deficient enzymatic activity of disaccharidases

(Horvath et al., 1999), increased secretin-induced pancreatico-biliary secretion (Horvath et al., 1999), and abnormal Clostridia taxa (Finegold et al., 2002; Song et al., 2004; Parracho et al., 2005). Some children placed on exclusion diets or treated with the antibiotic vancomycin are reported to improve in cognitive and social function (Knivsberg et al., 2002; Sandler et al., 2000).

[0052] The gastrointestinal tract is exposed to an onslaught of foreign material in the form of food, xenobiotics, and microbes. The intestinal muco-epithelial layer must maximize nutritional uptake of dietary components while maintaining a barrier to toxins and infectious agents. Although some aspects of these functions are host-encoded, others are acquired through symbiotic relationships with microbial flora. Dietary carbohydrates enter the intestine as monosaccharides (glucose, fructose, and galactose), disaccharides (lactose, sucrose, maltose), or complex polysaccharides. Following digestion with salivary and pancreatic amylases, carbohydrates are further digested by disaccharidases expressed by absorptive enterocytes in the brush border of the small intestine and transported as monosaccharides across the intestinal epithelium. However, humans lack the glycoside hydrolases and polysaccharide lyases necessary for cleavage of glycosidic linkages present in plant cell wall polysaccharides, oligosaccharides, storage polysaccharides, and resistant starches. Intestinal bacteria encoding these enzymes expand our capacity to extract energy from dietary polysaccharides (Sonnenburg et al., 2008; Flint et al., 2008). As an end product of polysaccharide fermentation, bacteria produce short-chain fatty acids (butyrate, acetate, and propionate) that serve as energy substrates for colonocytes, modulate colonic pH, regulate colonic cell proliferation and differentiation, and contribute to hepatic gluconeogenesis and cholesterol synthesis (Wong et al., 2006; Jacobs et al., 2009). Indigenous microflora also mediate postnatal development of the muco-epithelial layer, provide resistance to potential pathogens, regulate development of intraepithelial lymphocytes and Peyer's patches, influence cytokine production and serum immunoglobulin levels, and promote systemic lymphoid organogenesis (O'Hara and Shanahan, 2006; Macpherson and Harris, 2004).

[0053] The prevalence of autism in the US is about 1 in 91 births and, largely due to changes in diagnostic practices, services, and public awareness. Autism is growing at the fastest pace of any developmental disability (10-17%) (Fombonne, E. (2003). The prevalence of autism.

JAMA 289(1): 87–9). Care and treatment of autism costs the U.S. healthcare system \$90B annually. Early detection and intervention may result in reducing life-long costs. In the last 5 years, federal funding for autism research rose by 16.1%. The Autism Society is currently lobbying Congress for \$37 million for autism monitoring and studies, another \$16.5 million for autism screening and academic research. At present, few tools outside psychiatric evaluation are available for diagnosing autism. While a causative link between GI abnormalities and pathology of autism has yet to be established, a correlation between the two disorders is relatively well established. Thus, technologies facilitating detection and treatment of abnormal gut flora in autistic patients has great potential utility for diagnosis and treatment.

[0054] The present invention provides the discovery and the identification of GLUT2 as well SGLT1 as biomarkers for human Autism Spectrum Disorders. The present invention provides for methods to use genes encoding carbohydrate metabolic enzyme molecules (such as sucrase isomaltase, maltase glucoamylase, and lactase) or carbohydrate transporter molecules, or a combination of the two, and corresponding expression products for the diagnosis, prevention and treatment of autism and autism spectrum disorders.

[0055] The methods of the invention are useful in various subjects, such as humans, including adults, children, and developing human fetuses at the prenatal stage.

[0056] The GLUT2 gene locus can comprise all GLUT2 sequences or products in a cell or organism, including GLUT2 coding sequences, GLUT2 non-coding sequences (e.g., introns), GLUT2 regulatory sequences controlling transcription and/or translation (e.g., promoter, enhancer, terminator).

[0057] A GLUT2 gene, also known as SLC2A2, encodes the glucose transporter 2 isoform. It is an integral plasma membrane glycoprotein of the liver, pancreatic islet beta cells, intestine, and kidney epithelium. GLUT2 mediates the bidirectional transport of glucose. In the context of the invention, the GLUT2 gene also encompasses its variants, analogs and fragments thereof, including alleles thereof (e.g., germline mutations) which are related to susceptibility to autism and/or autism spectrum disorders.

[0058] The SGLT1 gene locus can comprise all SGLT1 sequences or products in a cell or organism, including SGLT1 coding sequences, SGLT1 non-coding sequences (e.g., introns), SGLT1 regulatory sequences controlling transcription and/or translation (e.g., promoter, enhancer, terminator).

[0059] A SGLT1 gene, also known as SLC5A1, encodes the sodium/glucose co-transporter 1. The sodium dependent glucose transporter is an integral plasma membrane glycoprotein of the intestine. SGLT1 mediates glucose and galactose uptake from the intestinal lumen. Mutations in this gene have been associated with glucose-galactose malabsorption. In the context of the invention, the SGLT1 gene also encompasses its variants, analogs and fragments thereof, including alleles thereof (e.g., germline mutations) which are related to susceptibility to autism and/or autism spectrum disorders.

[0060] As used herein, “carbohydrate transport activity” means the ability of a polypeptide to bind a carbohydrate, such as glucose, to a transporter protein, and subsequently facilitate uptake of the carbohydrate from the serum or extracellular milieu into a cell (e.g., a liver cell, or pancreatic β -cell). Glucose transport activity can be measured as described by Hissin et al., 1982, *J. Clin. Invest.* 70(4): 780-90. In one embodiment, the carbohydrate transport activity is glucose transport activity, and the activity can be measured by determining glucose transport activity as described in Hissin as well as the ability to decrease extracellular or serum glucose levels. Non-limiting examples of a carbohydrate transporter include GLUT1, GLUT2, GLUT3, GLUT4, GLUT5, GLUT6, GLUT7, GLUT8, GLUT9, GLUT10, GLUT11, GLUT12, and HMIT (see Scheepers et al., JPEN J Parenter Enteral Nutr. 2004 Sep-Oct;28(5):364-71).

[0061] A sucrase isomaltase (SI) gene encodes a sucrase-isomaltase protein, which is a glucosidase enzyme, that is expressed in the intestinal brush border. The encoded protein is synthesized as a precursor protein that is cleaved by pancreatic proteases into two enzymatic subunits, sucrase and isomaltase. The two subunits heterodimerize to form the sucrose-isomaltase complex, which is essential for the digestion of dietary carbohydrates including starch, sucrose and isomaltose. Mutations in this gene are the cause of congenital sucrase-isomaltase deficiency. In the context of the invention, the SI gene also encompasses its variants,

analogs and fragments thereof, including alleles thereof (e.g., germline mutations) which are related to susceptibility to autism and/or autism spectrum disorders.

[0062] A maltase glucoamylase (MGAM) gene encodes a maltase-glucoamylase enzyme. It is localized to the brush border membrane and plays a role in the final steps of digestion of starch. The protein has two catalytic sites identical to those of sucrase-isomaltase, but the proteins are only 59% homologous. Both are members of glycosyl hydrolase family 31, which has a variety of substrate specificities. In the context of the invention, the MGAM gene also encompasses its variants, analogs and fragments thereof, including alleles thereof (e.g., germline mutations) which are related to susceptibility to autism and/or autism spectrum disorders.

[0063] A lactase (LCT) gene encodes a glycosyl hydrolase of family 1. The protein is integral to plasma membrane and has both phlorizin hydrolase activity and lactase activity.

[0064] As used herein, “carbohydrate metabolic enzyme activity” includes “sucrase isomaltase activity”, “maltase glucoamylase activity”, “lactase activity”, “sucrase activity”, “maltase activity”, “trehalase activity”, “amylase activity”, “cellulase activity”, “glucosidase activity”, “pullulanase activity”, “galactosidase activity”, “alpha-Mannosidase activity”, “glucuronidase activity”, “hyaluronidase activity”, “glycosylase activity”, “fucosidase activity”, “hexosaminidase activity”, “iduronidase activity”, or “maltase-glucoamylase activity”. “Sucrase isomaltase activity” means the ability of a polypeptide to catalyze the hydrolysis of sucrose to fructose and glucose and to enzymatically digest polysaccharides at the alpha 1-6 linkages. Sucrase and isomaltase activities can be measured as described by Dahlqvist, A. (1964) *Anal. Biochem.* 7, 18-25 and the enzyme assays described by Goda et al., *Biochem J.* 1988 February 15; 250(1): 41–46. “Maltase glucoamylase activity” means the ability of a polypeptide to enzymatically digest starch, releasing maltose and free glucose, as well as to catalyze the hydrolysis of the disaccharide maltose. Maltase and glucoamylase activities can be measured as described by Dahlqvist A. Specificity of the human intestinal disaccharidases and implications for hereditary disaccharide intolerance. *J Clin Invest.* 1962;41:463–9; Dahlqvist A. Assay of intestinal disaccharidases. *Scand J Clin Lab Invest.* 1984;44:169–72; and Quezada-Calvillo et al., *J. Nutr.* 137:1725-1733, July 2007. “Lactase activity” means the ability of a polypeptide to hydrolyze lactose to galactose and glucose. Lactase activity can be measured as described by

Dahlqvist A. Specificity of the human intestinal disaccharidases and implications for hereditary disaccharide intolerance. *J Clin Invest.* 1962;41:463–9; Dahlqvist A. Assay of intestinal disaccharidases. *Scand J Clin Lab Invest.* 1984;44:169–72; and Quezada-Calvillo et al., *J. Nutr.* 137:1725-1733, July 2007. “Trehalase activity” means the ability of a polypeptide to catalyze the conversion of the dissacharide trehalose (α -D-glucopyranosyl-1,1- α -D-glucopyranoside) to glucose.

[0065] SEQ ID NO: 1 is the human wild type amino acid sequence corresponding to the GLUT2 enzyme (residues 1-524) having GenBank Accession No. NP_000331:

```

1  MTEDKVTGTL VFTVITAVLG SFQFGYDIGV INAPQQVIIS HYRHVLGVPL DDRKAINNYV
61  INSTDELPTI SYSMNPKPTP WAEETVAAA QLITMLWSLS VSSFAVGGMT ASFFGGWLGD
121 TLGRIKAMLV ANILSLVGAL LMGFSKLGPS HILIIAGRSI SGLYCLISG LVPYIGEIA
181 PTALRGALGT FHQLAIVTGI LISQIIGLEF ILGNYDLWHI LLGLSGVRAI LQSLFFFCP
241 ESPRYLYIKL DEEVKAKQSL KRLRGYDDVT KDINEMRKER EEASSEQKVS IIQLFTNSSY
301 RQPILVALML HVAQQFSGIN GIFYYSTSIF QTAGISKPVY ATIGVGAVNM VFTAVSVFLV
361 EKAGRRSLFL IGMSGMFVCA IFMSVGLVLL NKFSWMSYVS MIAIFLVVSF FEIGPGPIPW
421 FMVAEFFSQG PRPAALATAA FSNWTCNFIV ALCFQYIADF CGPYVFFLFA GVLLAFTLFT
481 FFKVPETK GK SFEETIAEFQ KKS GSAHRPK AAVEMKFLGA TETV

```

[0066] SEQ ID NO: 2 is the human wild type nucleic acid sequence corresponding to the GLUT2 enzyme (bps 1-3439) having GenBank Accession No. NM_000340:

```

1  tctggtttgt aacttatgcc taaggacact gctcccatTT tctttcctag tggaaacaaag
61  gtattgaagc cacaggttgc tgaggcaaag cacttattga ttagattccc atcaatatcc
121 agctgccgct gagaagatta gacttggaact cttaggtctg gtagcccaa ctcctccctc
181 tccttgctcc tctcctgca atgcataact aggcctaggc agagctgcga ataaacaggc
241 aggagctagt caggtgcatg tgccacactc acacaagacc tggaaattgac aggactccca
301 actagtacaa tgacagaaga taaggtcact gggaccctgg ttttactgt catcactgct
361 gtgctgggtt ccttccagtt tggatatgac attggtgtga tcaatgcacc tcaacaggta
421 ataatatctc actatagaca tgttttgggt gttccactgg atgaccgaaa agctatcaac
481 aactatgtta tcaacagtac agatgaactg cccacaatct catactcaat gaacccaaaa
541 ccaaccctt gggctgagga agagactgtg gcagctgctc aactaatcac catgctctgg
601 tccctgtctg tatccagctt tgcagttggt ggaatgactg catcattctt tgggtgggtg
661 cttggggaca cacttgggaag aatcaaagcc atgttagtag caaacattct gtcattagtt
721 ggagctctct tgatggggtt ttcaaaattg ggaccatctc atatacttat aattgctgga
781 agaagcatat caggactata ttgtgggcta atttcaggcc tggttcctat gtatatcggt
841 gaaattgctc caaccgctct caggggagca cttggcactt ttcacagct ggccatcgtc
901 acgggcattc ttattagtcg gattattggt cttgaattta tcttgggcaa ttatgatctg
961 tggcacatcc tgcttggcct gtctgggtgt cgagccatcc ttcagtctct gctactcttt
1021 ttctgtccag aaagccccag atacctttac atcaagttag atgagggaagt caaagcaaaa
1081 caaagcttga aaagactcag aggatatgat gatgtcacca aagatattaa tgaaatgaga
1141 aaagaaaagag aagaagcatc gagtgaagcag aaagtctcta taattcagct cttcaccaat
1201 tccagctacc gacagcctat tctagtggca ctgatgctgc atgtggctca gcaattttcc
1261 ggaatcaatg gcatttttta ctactcaacc agcatttttc agacggctgg tatcagcaaa
1321 cctgtttatg caaccattgg agttggcgct gtaaacatgg ttttactgct tgtctctgta
1381 ttcttgtggt agaaggcagg gcgacgttct ctctttctaa ttggaatgag tgggatgttt
1441 gtttgtgcca tcttcatgtc agtgggactt gtgctgctga ataagttctc ttggatgagt

```

```

1501 tatgtgagca tgatagccat ctctctcttt gtcagcttct ttgaaattgg gccaggccccg
1561 atcccctggt tcatggtggc tgagtttttc agtcaaggac cacgtcctgc tgcttttagca
1621 atagctgcat tcagcaattg gacctgcaat ttcattgtag ctctgtgttt ccagtacatt
1681 gcggacttct gtggacctta tgtgtttttc ctctttgctg gagtgtcctt ggcctttacc
1741 ctgttcacat tttttaaagt tccagaaacc aaaggaaagt cttttgagga aattgctgca
1801 gaattccaaa agaagagtgg ctacgcccac aggccaaaag ctgctgtaga aatgaaattc
1861 ctaggagcta cagagactgt gtaaaaaaaaa aaccctgctt tttgacatga acagaaacaa
1921 taagggaacc gtctgttttt aaatgatgat tccttgagca ttttatatcc acatctttaa
1981 gtattgtttt atttttatgt gctctcatca gaaatgtcat caaatattac caaaaaagta
2041 tttttttaag ttagagaata tttttttgat ggtaagactg taattaagta aaccaaaaag
2101 gctagtttat tttgttacac taaagggcag gtggttctaa ttttttagc tctgttcttt
2161 ataacaaggt tcttctaaaa ttgaagagat ttcaacatat ctttttttta acacataact
2221 agaaacctga ggatgcaaca aatatttata tttttgaata tcattaaatt ggaattttct
2281 taccatata tcttatgtta aaggagatat ggctagtggc aataagttcc atgttaaaat
2341 agacaactct tccatttatt gcactcagct tttttcttga gtactagaat ttgtattttg
2401 cttaaaattt tacttttggt ctgtattttc atgtggaatg gattatagag tatactaaaa
2461 aatgtctata gagaaaaact ttcatttttg gtaggcttat caaaatcttt cagcactcag
2521 aaaagaaaac catttttagtt cctttattta atggccaaat ggtttttgca agatttaaca
2581 ctaaaaaggt ttcacctgat catatagcgt gggttatcag ttaacattaa catctattat
2641 aaaaccatgt tgattccctt ctggtacaat cctttgagtt atagtttgc tttgttttta
2701 attgaggaca gcctggtttt cacatacact caaacaatca tgagtcagac atttggtata
2761 ttacctcaaa ttcctaataa gtttgatcaa atctaattga agaaaatttg aagtaaagga
2821 ttgatcactt tgtaaaaaat attttctgaa ttattatgtc tcaaaataag ttgaaaaggt
2881 agggtttgag gattcctgag tgtgggcttc tgaaacttca taaatgttca gcttcagact
2941 tttatcaaaa tccctattta attttcttg aaagactgat tgttttatgg tgtgttccta
3001 acataaaaata atcgtctcct ttgacatttc ctcttttgct ttagctgtat acagattcta
3061 gccaaactat tctatggcca ttaactaacac gcattgtaca ctatctatct gcctttacct
3121 acataggcaa attggaaata cacagatgat taaacagact ttagettaca gtcaatttta
3181 caattatgga aatatagttc tgatgggtcc caaaagctta gcagggtgct aacgtatctc
3241 taggctgttt tctccacca ctggagcact gatcaatcct tcttatgttt gctttaatgt
3301 gtattgaaga aaagcacttt ttaaaaagta ctctttaaga gtgaaataat taaaaaacac
3361 tgaacatttg ctttgttttc taaagttggt cacatatatg taatttagca gtccaaagaa
3421 caagaaattg tttcttttc

```

[0067] SEQ ID NO: 3 is the human wild type amino acid sequence corresponding to the SGLT1 enzyme (residues 1-664) having GenBank Accession No. NP_000334:

```

1 MDSSTWSPKT TAVTRPVETH ELIRNAADIS IIVIIYFVVM AVGLWAMFST NRGTVGGFFL
61 AGRSMVWWPI GASLFASNIG SGHFVGLAGT GAASGIAIGG FEWNALVLVV VLGWLFVPIY
121 IKAGVVTMPE YLRKRFGGQR IQVYLSLLSL LLYIFTKISA DIFSGAIFIN LALGLNLYLA
181 IFLLLAITAL YTITGGGLAAV IYTDLTQTVI MLVGLSLITG FAFHEVGGYD AFMEKYMKAI
241 PTIVSDGNTT FQEKCYTPRA DSFHIFRDPL TGDLPWPGFI FGMSILTLWY WCTDQVIVQR
301 CLSAKNMSHV KGGCILCGYL KLMPMFIMVM PGMISRILYT EKIACVVPSE CEKYCGTKVG
361 CTNIAYPTLV VELMPNGLRG LMLSVM LASL MSSLTSIFNS ASTLFTMDIY AKVRKRASEK
421 ELMIAGRLFI LVLIGISIAW VPIVQSAQSG QLFYDIQSIT SYLGPPIAAV FLLAIFWKRV
481 NEPGAFWGLI LGLLIGISRM ITEFAYGTGS CMEPSNCPTI ICGVHYLYFA IILFAISFIT
541 IVVISLLTKP IPDVHLYRLC WSLRNSKEER IDLDAEEENI QEGPKETIEI ETQVPEKKKG
601 IFRRAYDLFC GLEQHGAPKM TEEEEKAMKM KMTDTSEKPL WRTVLNVNGI ILVTVAVFCH
661 AYFA

```

[0068] SEQ ID NO: 4 is the human wild type nucleic acid sequence corresponding to the SGLT1 enzyme (bps 1-5061) having GenBank Accession No. NM_000343:

```

1  ccccatctgc aggcagctc ttacctgcgc ggccgcgcgc ccagccaaca gctcagccgg
61  gtgctccttc ctgggctcca cgcgccgagc tgcttcctga cgggtgcagc gcaaggcatc
121  gcaggggccc cgcgctactg ccctgctccc tcaaagtccc aggtcccctc ccctggtgct
181  gatcattaac caggaggccg tataaggagc tagcggccct ggcgagaggg aaggacgcaa
241  cgctgccacc atggacagta gcacctggag cccaagacc accgcggtca cccggcctgt
301  tgagaccac gagctcatte gcaatgcagc cgatatctcc atcatcgta tctacttcgt
361  ggtagtgatg gccgtcggac tgtgggctat gttttccacc aatcgtggga ctgttgaggg
421  cttcttcctg gcaggccgaa gtatgggtgt gtggccgatt ggagcctccc tctttgctag
481  taacattgga agtggccact ttgtggggct ggccgggact gggcgagctt caggcatcgc
541  cattggaggc ttgaaatgga atgccttggt tttggtggtt gtgctggggt ggctggttgt
601  ccccatctat attaaggctg ggggtggtgac aatgccagag tacctgagga agcgggttgt
661  aggccagcgg atccaggctc acctttccct tctgtccctg ctgctctaca ttttcaccaa
721  gatctcggca gacatcttct cgggggccat attcatcaat ctggccttag gcctgaatct
781  gtatttagcc atctttctct tattggcaat cactgccctt tacacaatta cagggggcct
841  ggcggcggtg atttacacgg acaccttgca gacggtgatc atgctggtgg ggtctttaat
901  cctgactggg tttgcttttc acgaagtggg aggctatgac gccttcattg aaaagtacat
961  gaaagccatt ccaaccatag tgtctgatgg caacaccacc tttcaggaaa aatgctacac
1021  tccaagggcc gactccttcc acatcttccg agatcccctc acgggagacc tcccatggcc
1081  tgggttcata tttgggatgt ccatecttac cttgtggtac tgggtgcagc atcaggctat
1141  tgtgcagcgc tgccctctcag ccaagaatat gtctcacgtg aagggtggct gcctcctgtg
1201  tgggtatcta aagctgatgc ccattgttcat catggtgatg ccaggaatga tcagccgcat
1261  tctgtacaca gaaaaaattg cctgtgtcgt cccttcagaa tgtgagaaat attgcggtac
1321  caaggttggc tgtaccaaca tcgcttatcc aaccttagtg gtggagctca tgcccaatgg
1381  actgcgaggc ctgatgctat cagtcatgct ggccctccctc atgagctccc tgacctccat
1441  cttcaacagc gccagcacc ctttcacat ggacatctac gccagggtcc gcaagagagc
1501  atctgagaaa gagctcatga ttgccggaag gttgtttatc ctggtgctga ttggcatcag
1561  catcgcttgg gtgcccattg tgcagtcagc acaaagtggg caactcttcg attacatcca
1621  gtccatcacc agttacttgg gaccacccat tgcggctgtc ttcctgcttg ctattttctg
1681  gaagagagtc aatgagccag gagccttttg gggactgatc ctaggacttc tgattgggat
1741  ttcacgtatg attactgagt ttgcttatgg aaccgggagc tgcattggagc ccagcaactg
1801  tcccacgatt atctgtgggg tgcactactt gtactttgcc attatcctct tcgccatttc
1861  tttcatcacc atcgtggtea tctccctcct caccaaacc attccgatg tgcattctta
1921  ccgtctgtgt tggagcctgc gcaacagcaa agaggagcgt attgacctgg atgcggaaga
1981  ggagaacatc caagaaggcc ctaaggagac cattgaaata gaaacacaga ttcctgagaa
2041  gaaaaaagga atcttcagga gagcctatga cctattttgt gggctagagc agcagcgtgc
2101  acccaagatg actgaggaag aggagaaagc catgaagatg aagatgacgg acacctctga
2161  gaagcctttg tggaggacag tgttgaacgt caatggcatc atcctggtga ccgtggctgt
2221  cttttgccat gcataatttg cctgagtcct accttttget gtagatttac catggctgga
2281  ctcttactca ccttccttta gtctcgtcct gtggtgttga agggaaatca gccagttgta
2341  aattttgccc aggtggataa atgtgtacat gtgtaattat aggctagctg gaagaaaacc
2401  attagtttgc tgttaattta tgcatttgaa gccagtgtga tacagccatc tgtacctact
2461  ggagctgcag aagggaagtc cactcagtea catccagaaa aaggcagact aagaatcaga
2521  agccatgtga ttgatgtctg acgtgagtc gtctcaggtg gattccgggt gtcagtgtgg
2581  tttataatcc ttgaatattg ttttagaaac ttttgtctcc ctggttctct ccaacttttc
2641  tgtccgtcct cctccccatt ttttttttaa aagaaagctg ttttccctc atcatatccc
2701  tcttgagttt tgccctggact ttccctctca agtgtgtcaa tcaggtaaac tgaggaatgc
2761  atgggaagctg aggatggagc ttgatgggct ccctgtcctg ggtgtttgct ctctgaagtg
2821  gaggcctgag gaaggtagta cttccacaaa agggagggac ccgggccccca gcctcaagct
2881  agtgggggag gcagatagcc tgaatccagg ggattttctg ggcttcttaa aatgtccatt
2941  gtgagttccc cgtgtttggg attccactca ttttggcatt cacagtgcct ggaatgtctt
3001  agattttcag caatgcgtgt tgaataaatg aatgacatag gcattttatt ttaaatcttt
3061  gcttgctttt tacatgagcc tggcccttag ttaacctttt cttgtggcta cacaagtat

```

```

3121 gctcactggt tactaatgac ttgggatgca tttgtcaaac tgattatatt agttttctag
3181 ggatgccata acaaagtagc acagaccaga tggtcgaagc agcagacatt tatttttctca
3241 cagttctaga ggctagaagt tggaggccaa gatgtcagca gggttggttt cttctgaggc
3301 ctctctcctt ggttgcagat ggtcatatct cactctgtct tccgtggcct tccttttgtc
3361 tgtgtcctaa atctactctt ctgataagga catcagtcac attggaatag gaccaccctt
3421 aatgtcttca ttttaatcac ctctttaaag cccctacctc caaatacagt cacactgtga
3481 gaaactgagg gttaggaagt cagcaagtga gtcttgaaga gatactaaac aaaccacaa
3541 cacagataaa gtatgcattt tggagatttc caagccagag tctcccgtga aaaaggtaaa
3601 cggaagcagt tattgtgcag caaaaggaaa aagaattaca aactgaacgt atgtaggtga
3661 ggcaaggcag ggtagggcag ggcctttggg taggctgac agagggtttt tcaacaataa
3721 atcaatggga atgcatttgt tgctcccagg accctggcac cttgactctg gtactatagc
3781 atgtcagcaa atacaagcaa agcccaacac tctgatttgc atttatgcca atctaaacta
3841 tccggtgttt agtttgattt tttgagtgc ggttcattca aggaccaggt tcccttgtgc
3901 tcagggtgaa gtagaaccag aaaacatcgt tatccattcc cagaagtttt ggaagagcct
3961 tggtagaaaa gcagaagctg ctttgaccgt gaaaatattt gactcctatc agtttttggg
4021 caggagaaga tatccaccta gaccaacctg aggagaaggc tcagagtaca gatatacccc
4081 gagcaacgtg atcaatgtcc ttgaaccttc atttttcatc tgaaaacaga gacataaatg
4141 cctggctcac agattttaat gttatacatt gacagcattt atcagtataa catttattta
4201 aataagtagg tgctcaatag gtgttggtct tctaacttgt ctacatccca tccccattcc
4261 agggctctca gaattgaagg agagatgttg tatcactgtt agaaggctgc tttgggacat
4321 tctgcagcag ggaggaggga ctgtcaaccc ctacaccatg accaccaagt tcctcacctt
4381 ggctgagtc ctaaaactct ctgaacctca ggttcctcca agcataatgc agacttcaca
4441 gagctgttgt aaagattagg tgagggtcaat tgatactgct taaaaggccc ggtccgtaga
4501 aaatgcccaa taaacattac tgctttcccc ctcaccctac tgcctgaaaa aatattacac
4561 ctgtgagact gactttgaga accagtgtgg gtggggagtt gtgcatataa actatttaat
4621 gagtaccaa caaaaagtc aagcttgtaa aatatcaggc cttgccccag aaagacaaat
4681 accacatgat ctactgata tgtagaatct taaaaagtca aactcagaag cagagagtag
4741 aatgatggtt atcaagggtt gggggaggga gggactgggg agatgttggg caaatgatac
4801 aaaggtttag ttaggtggaa taagttcaga aatcaattg tacaatgat caattatagt
4861 taatagcaat ataacatata cttgaaaatt gctgagagta gtgtgagtgt tctaccacaa
4921 aaaaatatgt gcagtaatag atgttaatta ccttaattta gtcatttcac aatatgtaca
4981 tatataaaaa tatgttgtat gccatgagta tatataatta ttatttgtga atttaaaaaa
5041 taaaaataat ttccaaaaaa a

```

[0069] SEQ ID NO: 5 is the human wild type amino acid sequence corresponding to the sucrase isomaltase (SI) enzyme (residues 1-1827) having GenBank Accession No. NP_001032:

```

1 MARKKFSGLE ISLIVLFVIV TIIAIALIVV LATKTPAVDE ISDSTSTPAT TRVTTNPSDS
61 GKCPNVLNDP VNVRIICIPE QFPTEGICAQ RGCCWRPWND SLIPWCFFVD NHGYNVQDMT
121 TTSIGVEAKL NRIPSPTLFG NDINSVLFTT QNQT PNRFRF KITDPNNRRY EVPHQYVKEF
181 TGPTVSDTLY DVKVAQNPFIS IQVIRKSNGK TLFDTSIGPL VYSDQYLQIS TRLPSDYIYG
241 IGEQVHKRFR HDLSWKTWPI FTRDQLPGDN NNNLYGHQTF FMCIEDTSGK SFGVFLMNSN
301 AMEIFIQPTP IVTYRVGTGGI LDFYILLGDT PEQVVQQYQQ LVGLPAMPAY WNLGFQLSRW
361 NYKSLDVVKE VVRRNREAGI PFDTQVTDID YMEDKKDFTY DQVAFNGLPQ FVQDLHDHGQ
421 KYVILDPAL SIGRRANGTT YATYERNGTQ HVWINESDGS TPIIGEVWPG LTVYPDFTNP
481 NCIDWWANEC SIFHQEVQYD GLWIDMNEVS SFIQGSTKGC NVNKLNYPPF TPDILDKLMY
541 SKTICMDAVQ NWGKQYDVHS LYGYSMIAIT EQAVQKVFPN KRSFILTRST FAGSGRHAHAH
601 WLGDNTASWE QMEWSITGML EFSLFGIPLV GADICGFVAE TTEELCRRWM QLGAFFPFSR
661 NHNSDGYEHQ DPAFFGQNSL LVKSSRQYLT IRYTLLPFLY TLFYKAHVFG ETVARPVLHE
721 FYEDTNSWIE DTEFLWGPAL LITPVLKQGA DTVSAYIPDA IWYDYESGAK RPWRKQVRDM
781 YLPADKIGLH LRGGYIPIQ EPDVTTTASR KNPLGLIVAL GENNTAKGDF FWDGETKDT
841 IQNGNYILYT FSVSNNTLDI VCTHSSYQEG TTLAFQTVKI LGLTDSVTEV RVAENNQPMN
901 AHSNFTYDAS NQVLLIADLK LNLGRNFSVQ WNQIFSENER FNCYPADLA TEQKCTQRGC

```

```

961 VVRTGSSLSK APECYFPRQD NSYSVNSARY SSMGITADLQ LNTANARIKL PSDPISTLRV
1021 EVKYHKNDML QFKIYDPQKK RYEVVPVLNI PTTPISTYED RLYDVEIKEN PFGIQIRRRS
1081 SGRVIWDSWL PGFAFNDQFI QISTRLPSEY IYGFGEVEHT AFKRDLNWNW WGMFTRDQPP
1141 GYKLNSYGFH PYMALEEEG NAHGVFLNLS NAMDVTFQPT PALTYRTVGG ILDFYMFLLG
1201 TPEVATKQYH EVIGHFVMPA YWALGFQLCR YGYANTSEVR ELYDAMVAAN IPYDVQYTDI
1261 DYMERQLDFT IGEAFQDLPO FVDKIRGEGM RYIIILDPAI SGNETKTYPA FERGGQNDVF
1321 VKWPNTNDIC WAKVWPDLPN ITIDKTLTED EAVNASRAHV AFPDFFRTST AEWWAREIVD
1381 FYNEKMKFDG LWIDMNEPSS FVNGTTTNQC RNDELNYPPY FPELTKRTDG LHFRTICMEA
1441 EQILSDGTSV LHYDVHNLYG WSQMKPTHDA LQKTTGKRGV VISRSTYPTS GRWGGHWLGD
1501 NYARWDNMDK SIIGMMEFSL FGMSYTGADI CGFFNNSEYH LCTRWMQLGA FYPYSRNHNI
1561 ANTRRQDPAS WNETFAEMSR NILNIRYTL PYFYTMHEI HANGGTVIRP LLHEFFDEKP
1621 TWDIFKQFLW GPAFMVTPVL EPYVQTVNAY VPNAWFDYH TGKDIGVRGQ FQTFNASYDT
1681 INLHVRGGHI LPCQEPQNT FYSRQKHKML IVAADDNQMA QGSLFWDDGE SIDTYERDLY
1741 LSVQFNLNQT TLTSTILKRG YINKSETRLG SLHVWGKGT PVNAVTLTYN GNKNSLPFNE
1801 DTTNMILRID LTTHNVTLEE PIEINWS

```

[0070] SEQ ID NO: 6 is the human wild type nucleic acid sequence corresponding to the sucrose isomaltase (SI) enzyme (bps 1-6023) having GenBank Accession No. NM_001041:

```

1 ttattttggc agccttatcc aagtctggta caacatagca aagagaacag gctatgaaat
61 aagatggcaa gaaagaaatt tagtggattg gaaatctctc tgattgtcct ttttgtcata
121 gttactataa tagctattgc ctttaattgtt gtttttagcaa ctaagacacc tgctgttgat
181 gaaattagtg attctacttc aactccagct actactcgtg tgactacaaa tccttctgat
241 tcaggaaaat gtccaaatgt gttaaatgat cctgtcaatg tgagaataaa ctgcattcca
301 gaacaattcc caacagaggg aatttgtgca cagagaggct gctgctggag gccgtggaat
361 gactctctta ttcttgggtg cttcttcggt gataatcatg gttataacgt tcaagacatg
421 acaacaacaa gtattggagt tgaagccaaa ttaaacagga taccttcacc tacactatct
481 ggaaatgaca tcaacagtgt tctcttcaca actcaaaatc agacacccaa tcgtttccgg
541 ttcaagatta ctgatccaaa taatagaaga tatgaagttc ctcatcagta tgtaaaagag
601 tttactggac ccacagtttc tgatacgttg tatgatgtga aggttgccca aaaccattt
661 agcatccaag ttattaggaa aagcaacggt aaaactttgt ttgacaccag cattggtccc
721 ttagtgtact ctgaccagta cttacagatc tcaaccgctc ttccaagtga ttatatttat
781 ggtattggag aacaagttca taagagattt cgtcatgatt tatcctggaa aacatggcca
841 atttttactc gagaccaact tctgtgtgat aataataata atttatacgg ccatcaaaca
901 ttctttatgt gtattgaaga tacatctgga aagtcattcg gtgtttttt aatgaatagc
961 aatgcaatgg agatttttat ccagcctact ccaatagtaa catatagagt taccggtggc
1021 attctggatt ttacatcct tctaggagat acaccagaac aagtagttca acagtatcaa
1081 cagcttggtg gactaccagc aatgccagca tattggaatc ttggattcca actaagtcgc
1141 tggaaattata agtcactaga tgtagtgaaa gaagtggtaa ggagaaaccg ggaagctggc
1201 ataccatttg atacacaggt cactgatatt gactacatgg aagacaagaa agactttact
1261 tatgatcaag ttgcgtttta cggactccct caatttgtgc aagatttgca tgaccatgga
1321 cagaaatatg tcatcatctt ggaccctgca atttccatag gtcgaaatgc caatggaaca
1381 acatatgcaa cctatgagag gggaaacaca caacatgtgt ggataaagta gtcagatgga
1441 agtacaccaa ttattggaga ggtatggcca ggattaacag tataccctga tttcactaac
1501 ccaaactgca ttgattgggt ggcaaagtga tgcagtattt tccatcaaga agtgcaatat
1561 gatggacttt ggattgacat gaatgaagtt tccagcttta ttcaaggttc aacaaaagga
1621 tgtaattgaa acaaattgaa ttatccaccg ttactcctg atattcctga caaactcatg
1681 tattccaaaa caatttgcag ggatgctgtg cagaactggg gtaaacagta tgatgttcat
1741 agcctctatg gatacagcat ggctatagcc acagagcaag ctgtacaaaa agtttttctc
1801 aataagagaa gcttcattct taccgctca acatttgcgt gatctggaag acatgctgct
1861 cattggttag gagacaatac tgcttcatgg gaacaaatgg aatggtctat aactggaatg
1921 ctggagttca gtttgttttg aatacctttg gttggagcag acatctgtgg atttgtggct
1981 gaaaccacag aagaactttg cagaagatgg atgcaacttg ggcatttta tccattttcc
2041 agaaaccata attctgacgg atatgaacat caggatcctg catttttttg gcagaattca

```

2101 ctttttggtta aatcatcaag gcagtatatta actattcgct acaccttatt acccttcctc
 2161 tacactctgt ttataaaagc ccatgtgttt ggagaaacag tagcaagacc agttcttcat
 2221 gagttttatg aggatacga cagctggatt gaggacactg agtttttgtg gggccctgca
 2281 ttacttatta ctctgttct aaaacaggga gcagatactg tgagtgccta catccctgat
 2341 gctatttggt atgattatga atctggtgca aaaaggccat ggaggaaaca acgggttgat
 2401 atgtatcttc cagcagacaa aataggatta catcttagag gaggttatat catccccatt
 2461 caagaaccag atgtaacaac aacagcaagc cgtaagaatc ctctaggact tatagtcgca
 2521 ttaggtgaaa acaacacagc caaaggagac tttttctggg atgatggaga aactaaagat
 2581 acaatacaaaa atggcaacta catattatat acattttcag tttctaataa cacattagat
 2641 attgtgtgca cacattcatt atatacaggaa ggaactacct tagcatttca gactgtaaaa
 2701 atccttgggt tgacagacag gtttacagaa gttagagtgg cggaaaataa tcaaccaatg
 2761 aacgctcatt ccaatttcac ttatgatgct tctaaccagg ttctcctaatt tgcagatctc
 2821 aaacttaatc ttggaagaaa ctttagtggt caatggaatc aaattttctc agaaaatgaa
 2881 agatttaatt gttatccaga tgcagatttg gcaactgaac aaaagtgcac acaacgtggc
 2941 tgtgtatgga gaacgggttc ttctctatcc aaagcacctg agtgttactt tcccagacaa
 3001 gataactctt attcagtcac ctacgtctgc tattcatcca tgggtataac agctgacctc
 3061 caactaaata ctgcaaatgc cagaataaag ttacctctg accccatctc aactcttcgt
 3121 ttggagggtga aatatcaca aaatgatatt ttgcagttta agatttatga tccccaaaag
 3181 aagagatatg aagtaccagt accgttaaac attccaacca cccaataaag tacttatgaa
 3241 gacagacttt atgatgtgga aatcaaggaa aatccttttg gcatccagat tgcacggaga
 3301 agcagtggaa gagtcatttg ggattcttgg ctgcctggat ttgcttttaa tgaccagttc
 3361 attcaaatat cgactcgctt gccatcagaa tatatatatg gttttgggga agtggaacat
 3421 acagcattta agcgagatct gaactggaat acttggggaa tgttcacaag agaccaacc
 3481 cctggttaca aacttaattc ctatggattt catccctatt acatggctct ggaagaggag
 3541 ggcaatgctc atggtgtttt ctactcaac agcaatgcaa tggatgttac attccagcca
 3601 actcctgctc taacttaccg tacagttgga gggatcttgg atttttatat gttttgggc
 3661 ccaactccag aagttgcaac aaagcaatac catgaagtaa ttggccatcc agtcatgcca
 3721 gcttattggg ctttgggatt ccaattatgt cgttatggat atgcaaatat ttcagaggtt
 3781 cgggaattat atgacgctat ggtggctgct aacatcccct atgatgttca gtacacagac
 3841 attgactaca tggaaaggca gctagacttt acaattgggtg aagcattcca ggaccttct
 3901 cagtttggtg acaaaaataag aggagaagga atgagataca ttattatcct ggaatccaga
 3961 atttcaggaa atgaaacaaa gacttaccct gcatttgaaa gaggacagca gaatgatgtc
 4021 tttgtcaaat ggccaaacac caatgacatt tgttgggcaa aggtttggcc agatttgccc
 4081 aacataacaa tagataaaaac tctaacggaa gatgaagctg ttaatgcttc cagagctcat
 4141 gtagctttcc cagatttctt caggacttcc acagcagagt ggtgggcccag agaaattgtg
 4201 gacttttaca atgaaaagat gaagtttgat ggtttgtgga ttgatatgaa tgagccatca
 4261 agttttgtaa atggaacaac tactaatcaa tgcagaaatg acgaactaaa ttatccacct
 4321 tatttcccag aactcacaac aagaactgat ggattacatt tcagaacaat ttgcatggaa
 4381 gctgagcaga ttcttagtga tggaaacatca gttttgcatt acgatgttca caatctctat
 4441 ggatggtcac agatgaaacc tactcatgat gcattgcaga agacaactgg aaaaagaggg
 4501 attgtaattt ctctgtccac gtatcttact agtgagcagat ggggaggaca ctggcttgga
 4561 gacaactatg cagcatggga caacatggac aaatcaatca ttggtatgat ggaatttagt
 4621 ctgtttggaa tgtcatatac tggagcagac atctgtggtt ttttcaacaa ctcagaatat
 4681 catctctgta cccgctggat gcaacttgga gcattttatc catactcaag gaatcacaac
 4741 attgcaaata ctagaagaca agatcccgtt tcctggaatg aaacttttgc tgaaatgtca
 4801 aggaatatct taaatattag atacacctta ttgccctatt tttacacaca aatgcatgaa
 4861 attcatgcta atggtggcac tgttatccga ccccttttgc atgagttctt tgatgaaaaa
 4921 ccaacctggg atatatccaa gcagttctta tggggtccag catttatggt taccocagta
 4981 ctggaacctt atgttcaaac tgtaaatgcc tacgtcccca atgctcgggt gtttgactac
 5041 catacaggca aagatatattg cgctcagagga caatttcaaa catttaatgc ttcttatgac
 5101 acaataaacc tacatgtccg tgggtggcac atcctaccat gtcaagagcc agctcaaaac
 5161 acattttaca gtcgacaaaa acacatgaag ctcatgtttg ctgcagatga taatcagatg
 5221 gcacagggtt ctctgttttg ggtgatgga gagagtatag acacctatga aagagacct
 5281 tatttatctg tacaatttaa tttaaaccag accaccttaa caagcactat attgaagaga
 5341 ggttacataa ataaaagtga aacagggctt ggtatccctc atgtatggg gaaaggaaact
 5401 actcctgtca atgcagttac tctaacgtat aacggaaata aaaattcgct tctttttaat
 5461 gaagacacta ccaacatgat attacgtatt gatctgacca cacacaatgt tactctagaa

```

5521 gaaccaatag aaatcaactg gtcatgaaga tcaccatcaa ttttagttgt caatgggaaa
5581 aaacaccagg atttaagttt cacagcactt acaattttcc ctcttcactt ggttcttgta
5641 ctctacaaaa tatagctttc ataacatcga aaagttatth tgtagcgtac atcaatgata
5701 atgctaattt tattatagta atgtgacttg gattcaattt taaggcatat ttaacaaaat
5761 ttgaatagcc ctatttatcc ttgttaagta tcagctacaa ttgtaaaacta gttactaaac
5821 atgtatgtaa atagctaaga tataatttaa acgtgatttt taaattaaat aaaattttta
5881 tgtaattata tatactatat ttttctcaat gtttagcaga ttttaagatat gtaacaacaa
5941 ttattttgaag atttaattac ttcttagtat gtgcatttaa ttagaaaaag agaataaaaa
6001 atgtaagtgt aaaaaaaaaa aaa

```

[0071] SEQ ID NO: 7 is the human wild type amino acid sequence corresponding to the maltase glucoamylase (MGAM) enzyme (residues 1-1857) having GenBank Accession No. NP_004659:

```

1 MARKKLLKFT TLEIVLSVLL LVLFIISIVL IVLLAKESLK STAPDPGTTG TPDPGTTGTP
61 DPGTTGTTHA RTTGPPDPGT TGTPVSAEC PVVNELERIN CIPDQPPTKA TCDQRGCCWN
121 PQGAVSVFPC YYSKNHSYHV EGNLVNTNAG FTARLKNLPS SPVFGSNVDN VLLTAEYQTS
181 NRFHFKLTDQ TNNRFEVPHE HVQSFSGNAA ASLTYQVEIS RQPFSEIKVTR RSNNRVLFDS
241 SIGPLLFADQ FLQLSTRLPS TNVYGLGEHV HQQYRHDMMW KTWPIFNRTD TPNGNGTNLY
301 GAQTFFLCLE DASGLSFGVF LMNSNAMEVV LQPAPAITYR TIGGILDFYV FLGNTPEQVV
361 QEYLELIGRP ALPSYWALGF HLSRYEYGTI DNMREVVERN RAAQLPYDVQ HADIDYMDER
421 RDTYDSVDF KGFPEFVNEL HNNGQKLVII VDPAISNNSS SSKPYGPYDR GSDMKIWNVS
481 SDGVTPPLIGE VWPGQTVFPD YTNPNCAVWW TKEFELFHNQ VEFDGIWIDM NEVSNFVDGS
541 VSGCSTNNLN NPPFTPRILD GYLFCCKTLCM DAVQHWGKQY DIHNLYGYSM AVATAEAAKT
601 VFPNKRSEIL TRSTFAGSGK FAAHWLGDNT ATWDDLRSWI PGVLEFNLFG IPMVGPDICG
661 FALDTPPEELC RRWMQLGAFY PFSRNHNGQG YKDQDPASFG ADSLLLNSSR HYLNIRYTL
721 PYLYTLFFRA HSRGDTVARP LLHEFYEDNS TWDVHQQFLW GPGLLITPVL DEGAEKVMAY
781 VPDVWYDYE TGSQVRWRKQ KVEMELPGDK IGLHLRGGYI FPTQQPNTTT LASRKNPLGL
841 IIALDENKEA KGELFWDNGE TKDTVANKVY LLCEFSVTQN RLEVNISQST YKDPNNLAFN
901 EIKILGTEEP SNVTVKHNGV PSQTSPTVTY DSNLKVAIIT DIDLLGEAY TVEWSIKIRD
961 EEKIDCYPDE NGASAENCTA RGCIVEASNS SGVPFCYFVN DLYSVSDVQY NSHGATADIS
1021 LKSSVYANAF PSTPVNPLRL DVTYHKNEML QFKIYDPNKN RYEVVPVPLNI PSMPSSTPEG
1081 QLYDVLIKKN PFGIEIRRKS TGTIIWDSQL LGFTFSDMFI RISTRPLSKY LYGFGETEHR
1141 SYRRDLEWHT WGMFSRDQPP GYKKN SYGVH PYMGLEEDG SAHGVLLNS NAMDVTFQPL
1201 PALT YRTTGG VLDYFVFLGP TPELVTQOYT ELIGRPVMPV YWSLGFQLCR YGYQNDSEIA
1261 SLYDEMVAQ IPYDVQYSDI DYMERQLDFT LSPKFAGFPA LINRMKADGM RVILILDPAI
1321 SGNETQPYPA FTRGVEDDVF IKYPNDGDIV WGVVWPDFPD VVVGSLDWD SQVELYRAYV
1381 AFPDFFRNST AKWWKREIEE LYNNPQNPER SLKFDGMWID MNEPSSFVNG AVSPGCRDAS
1441 LNHPYPMPHL ESRDRGLSSK TLCMESQOIL PDGSLVQHYN VHNLYGWSQT RPTYEAVQEV
1501 TGQRGVVITR STFPSSGRWA GHWLGDNATA WDQLKKSIIIG MMEFSLFGIS YTGADICGFF
1561 QDAEYEMCVR WMQLGAFYFP SRNHNTIGTR RQDPVSWDVA FVNISRTVLQ TRYTLPLPYL
1621 TLMHKAHTG VTVVRPLLHE FVSDQVTWDI DSQFLLGPAF LVSPVLERNA RNVTAFFPRA
1681 RWYDYTYGVD INARGEWKTL PAPLDHINLH VRGGYILPWQ EPALNTHLSR QKFMGFKIAL
1741 DDEGTAGGWL FWDDGQSIDT YGKGLYYLAS FSASQNTMQS HIIFNYYITG TNPLKLGYLE
1801 IWGVGSVPVT SVSISVSGMV ITPSFNNDPT TQVLSIDVTD RNISLHNFTS LTWISTL

```

[0072] SEQ ID NO: 8 is the human wild type nucleic acid sequence corresponding to the maltase glucoamylase (MGAM) enzyme (bps 1-6513) having GenBank Accession No. NM_004668:

```

1 attgctaagc catccttcag acagagaggg agcggctgca agaggtaatg agagatggca
61 agaaaaaagc tgaaaaaatt tactactttg gagattgtgc tcagtgttct tctgcttggtg
121 ttgtttatca tcagtattgt totaattgtg ctttttagcca aagagtcact gaaatcaaca
181 gccccagatc ctgggacaac tggtagccca gatcctggga caactggtac cccagatcct
241 ggaacaactg gtaccacaca tgctaggaca acgggtcccc cagatcctgg aacaactggt
301 accactcctg tttctgctga atgtccagtg gtaaatgaat tggaacgaat taattgcatc
361 cctgaccagc cgccaacaaa ggccacatgt gaccaacgtg gctgttgctg gaatccccag
421 ggagctgtaa gtgttccctg gtgctactat tccaagaatc atagctacca tgtagagggc
481 aaccttgtca acacaaatgc aggattcaca gcccggttga aaaatctgcc ttcttcacca
541 gtgtttggaa gcaatgttga caatgttctt ctacacagcag aatatcagac atctaactgt
601 ttccacttta agttgactga ccaaaccaat aacaggtttg aagtgtccca cgaacacgtg
661 cagtccttca gtggaaatgc tgctgttctt ttgacctacc aagttgaaat ctccagacag
721 ccatttagca tcaaagtgc cagaagaagc aacaatcgtg tttgtttga ctcgagcatt
781 gggccctac tgtttgctga ccagttcttg cagctctcca ctgactgcc tagcactaac
841 gtgtatggcc tgggagagca tgtgcaccag cagtatcggc atgatatgaa ttggaagacc
901 tggcccatat ttaacagaga cacaactccc aatggaaacg gaactaattt gtatggtgcg
961 cagacattct tcttggtgct tgaagatgct agtggattgt cctttggggt gtttctgatg
1021 aacagcaatt ccattggagg tgtccttcag cctgcgccag ccatcactta ccgcaccatt
1081 gggggcattc tcgacttcta tgtgttcttg ggaaacactc cagagcaagt tgttcaagaa
1141 tatctagagc tcattgggag gccagccctt ccctcctact gggcgcttgg atttcacctc
1201 agtcgttacg aatatggaac cttagacaac atgagggagc tcgtggagag aaatcgcgca
1261 gcacagctcc cttatgatgt tcagcatgct gatattgatt atatggatga gagaagggac
1321 ttcacttatg attcagtgga ttttaaaggc ttccctgaat ttgtcaacga gttacacaat
1381 aatggacaga agcttgatc cattgtggat ccagccatct ccaacaactc ttcctcaagt
1441 aaaccctatg gcccatatga caggggttca gatatgaaga tatgggtgaa tagttcagat
1501 ggagtgactc cactcattgg ggaggtctgg cctggacaaa ctgtgtttcc tgattatacc
1561 aatcccaact gtgctgtttg gtggacaaag gaatttgagc tttttcaca tcaagtagag
1621 tttgatggaa tctggattga tatgaatgaa gtctccaact ttgttgatgg ttcggtctca
1681 ggatgttcca caaacaacct aaataatccc ccattcactc ccagaatcct ggatgggtac
1741 ctgtttctga agactctctg tatggatgca gtgcagcact ggggcaagca gtagacatt
1801 cacaatctgt atggctactc catggcggtc gccacagcag aagctgccaa cctgtgttct
1861 cctaataaga gaagcttcat tctgaccctg tctacctttg cgggctctgg caagtttgca
1921 gcacattggt taggagacaa cactgccacc tgggatgacc tgagatggtc catccctggc
1981 gtgcttgagt tcaacctttt tggcatccca atgggtgggtc ctgacatatg tggctttgct
2041 ttggacaccc ctgaggagct ctgtaggcgg tggatgcagt tgggtgcatt ttatccgttt
2101 tctagaaatc acaatggcca aggtctacaag gaccaggatc ctgcctcctt tggagctgac
2161 tccctgctgt tgaattcctc caggcactac cttaacatcc gctatactct attgccctac
2221 ctatacaccc tcttcttccg tgctcacagc cgaggggaca cgggtggccag gccctttttg
2281 catgagttct acgaggacaa cagcacttgg gatgtgcacc aacagttctt atggggggcc
2341 ggcctcctca tcaactccag tctggatgaa ggtgcagaga aagtgatggc atatgtgcct
2401 gatgctgtct ggtatgacta cgagactggg agccaagtga gatggaggaa gcaaaaagtc
2461 gagatggaac ttctctggga caaaattgga cttcaccttc gaggaggcta catcttcccc
2521 acacagcagc caaatacaac cactctggcc agtcgaaaga accctcttgg tcttatcatt
2581 gccctagatg agaacaaaga agcaaaagga gaacttttct gggataatgg gaaacgaag
2641 gatactgtgg ccaataaagt gtatctttta tgtgagtttt ctgtcactca aaaccgcttg
2701 gaggtgaata tttcacaatc aacctacaag gacccaata atttagcatt taatgagatt
2761 aaaattcttg ggacggagga acctagcaat gttacagtga aacacaatgg tgtcccaagt
2821 cagacttctc ctacagtcac ttatgattct aacctgaagg ttgccattat cacagatatt
2881 gatcttctcc tgggagaagc atacacagtg gaatggagca taaagataag ggatgaagaa
2941 aaaatagact gttaccctga tgagaatggt gcttctgccg aaaactgcac tgcccgtggc
3001 tgtatctggg aggcattccaa ttcttctgga gtcccttttt gctattttgt caacgacctc
3061 tactctgtca gtgatgttca gtataatttc catggggcca cagctgacat ctcttaag
3121 tcttccgttt atgccaatgc ctccccctcc acaccgtga accccttcg cctggatgtc
3181 acttaccata agaattgaaat gctgcagttc aagatttatg atcccaaca gaatcggtat
3241 gaagttccag tccctctgaa cataccagc atgccatcca gcaccctga ggtcaactc
3301 tatgatgtgc tcattaagaa gaatccattt gggattgaaa ttcgccgga ggtacaggc
3361 actataattt gggactctca gctccttggc tttaccttca gtgacatgtt tatccgcac

```

```

3421 tccacccgcc ttccctccaa gtacctctat ggctttgggg aaactgagca caggctcctat
3481 aggagagact tggagtggca cacttggggg atgttctccc gagaccagcc cccagggtag
3541 aagaagaatt cctatgggtg ccacccctac tacatggggc tggaggagga cggcagtgcc
3601 catggagtgc tcctgctgaa cagcaatgcc atggatgtga cgttccagcc cctgcctgcc
3661 ttgacatacc gcaccacagg gggagtctct gacttttatg tgttcttggg gccgactcca
3721 gagcttgta cccagcagta cactgagttg attggccggc ctgtgatggg acctacttgg
3781 tctttggggg tccagctgtg tcgctatggc taccagaatg actctgagat cgccagcttg
3841 tatgatgaga tgggtggctgc ccagatccct tatgatgtgc agtactcaga catcgactac
3901 atggagcggc agctggactt caccctcagc cccaagtttg ctgggtttcc agctctgata
3961 aatcgcatga aggtgatgg gatgggggtc atcctcattc tggatccagc catttctggc
4021 aatgagacac agccttatcc tgccttctac cggggcgtgg aggatgacgt cttcatcaaa
4081 tacccaaata atggagacat tgtctgggga aaggtctggc ctgattttcc tgatgttggt
4141 gtgaatgggt ctctagactg ggacagccaa gtggagctat atcgagctta tgtggccttc
4201 ccagactttt tccgtaattc aactgccaa ggggtgaaga gggaaataga agaactatac
4261 aacaatccac agaataccaga gaggagcttg aagtttgatg gcatgtggat tgatatgaat
4321 gaaccatcaa gcttcgtgaa tggggcagtt tctccaggct gcaggagcgc ctctctgaac
4381 caccctccct acatgccaca tttgtagtcc agggacaggg gcctgagcag caagaccctt
4441 tgtatggaga gtcagcagat cctccagac ggctccctgg tgcagcacta caacgtgcac
4501 aacctgtatg ggtgggtcca gaccagaccc acatacgaag ccgtgcagga ggtgacggga
4561 cagcgagggg tcgtcatcac ccgctccaca tttccctctt ctggccgctg ggcaggacat
4621 tggctgggag acaacacggc cgcattggat cagctgaaga agtctatcat tggcatgatg
4681 gagttcagcc tcttcggcat atcctatacg ggagcagata tctgtgggtt ctttcaagat
4741 gctgaatatg agatgtgtgt tcgctggatg cagctggggg ctttttacc cttctcaaga
4801 aaccacaaca ccattgggac caggagacaa gaccctgtgt cctgggatgt tgcttttgtg
4861 aatatttcca gaactgtcct gcagaccaga tacaccctgt tgccatatct gtataccttg
4921 atgcataagg cccacacgga gggcgctact gttgtgcggc ctctgctcca tgagtttgtg
4981 tcagaccagg tgacatggga catagacagt cagttcctgc tgggcccagc cttcctgggtc
5041 agccctgtcc tggagcgtaa tgccagaaat gtcactgcat atttcctag agcccgtgg
5101 tatgattact acacgggtgt ggatattaat gcaagaggag agtgaagac cttgccagcc
5161 cctcttgacc acattaatct tcactccgtt gggggctaca tcctgcctg gcaagagcct
5221 gcaactgaaca cccacttaag ccgccagaaa ttcattgggt tcaaaattgc cttggatgat
5281 gaaggaaactg ctgggggctg gctcttcttg gatgatgggc aaagcattga tacctatggg
5341 aaaggactct attacttggc cagcttttct gccagccaga atacgatgca aagccatata
5401 attttcaaca attacatcac tggtaaaaa cctttgaaac tgggctacat tgaaatctgg
5461 ggagtgggca gtgtccccgt taccagtgtc agcatctctg tgagtggcat ggtcataaca
5521 ccctccttca acaatgaccc cagcacacag gtattaagca tcgatgtgac tgacagaaac
5581 atcagcctac ataattttac ttcattgacg tggataagca ctctgtgaat ttttacagca
5641 agattctaac taactatgaa tgactttgaa actacttata cttcatactc ataaaaatta
5701 ttgtgtgttg ctaatttgtt catacccact attggtgaaa tatttctgtt aattttgtta
5761 tatgtttttt gtgtgaaccc taaaggttaa accttagccc tgtgggatag gcagttaggg
5821 aggtgtggaa aatctatgca ttaccttaat gtctctgtgt ggttagtatg gtagtgactg
5881 ttcatacatat gacatttact gaagatgaac tgggtccatg atgaagtgtg tgtatgtcca
5941 cgtttgtaat catagaatgg acccattctt tttgttaaat acacaagaga aagctttctg
6001 tgacagttcc aggtcttgaa gctaatacagc atctcaagaa agtatccaga aagaacatct
6061 gctagttggt tataggcggg gggaggaata atatacctaa ttggttatag gtggggggag
6121 catgataagc aaagaaaagg caaacacaag gaaagatcag atgaaacaga agatgatagt
6181 aaaagtgatc ctaagtaaga acataatgta aaattgtcag cagcctcatg gggaggaaaa
6241 aggaagagtc aactcacttg aagaagaggg tcttgagaaa tccttagcat aaagggtac
6301 tggtgagatt gagatctgag caggcaaagc tcaaaagaga gtttgagggt taaaataaat
6361 ttattttttg agtagtgtgc tttgaaatgt gtaaattctt tttctaattg atacaaccac
6421 atttcacata aaaatatgca atttatatgc cagataaaaa taaaacaagt gaatttgcaa
6481 gtgaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaa

```

[0073] SEQ ID NO: 9 is the human wild type amino acid sequence corresponding to the lactase (LCT) enzyme (residues 1-1927) having GenBank Accession No. NP_002290:

```

1  MELSWHVVF  ALLSFSCWGS DWESDRNFIS TAGPLTNDLL HNLGGLLDQ SSNFVAGDKD
61  MYVCHQPLPT FLPEYFSSLH ASQITHYKVF LSWAQLLPAG STQNPDEKTV QCYRRLKAL
121  KTARLQPMVI LHHQTLPAST LRRTEAFADL FADYATFAFH SFGDLVGIWF TFSDLLEVIK
181  ELPHQESRAS QLQTLSDAHR KAYEIIHESY AFQGGKLSVV LRAEDIPELL LEPPISALAQ
241  DTVDFLSLDL SYECQNEASL RQKLSKLQTI EPKVVFIFN LKLPDCPSTM KNPASLLFSL
301  FEANKDQVL TIGFDINEFL SCSSSSKKS SMCSLTGSLAL QPDQQQDHET TDSSPASAYQ
361  RIWEAFANQS RAERDAFLQD TFPEGFLWGA STGAFNVEGG WAEGGRGVS IWDPRRPLNTT
421  EGQATLEVAS DSYHKVASDV ALLCGLRAQV YKFSISWSRI FPMGHGSSPS LPGVAYYNKL
481  IDRLQDAGIE PMATLFHWDL PQALQDHGGW QNESVVD AFLDYAAFCSTF GDRVCLWVTF
541  HEPWVMSYAG YGTGQHPPGI SDPGVASFKV AHLVLKAHAR TWHHYNSHHR PQQQGHVGVIV
601  LNSDWAEP LS PERPEDLRAS ERFLHFM LGWFAHPVFVDGD YPATLRTQIQ QMNRQC SHPV
661  AQLPEFTEAE KQLLKGSADF LGLSHYTSRL ISNAPQNTCI PSYDTIGGFS QHVNHWVPQT
721  SSSWIRVVPW GIRLLQFVS LEYTRGKVPI YLAGNGMPIG ESENLFDDSL RVDYFNQYIN
781  EVLKAIKEDS VDVSRYIARS LIDGFEGPSG YSQRFG LHHVNFSDSSKSRT PRKSAYFFTS
841  IIEKNGLTK GAKRLLPPNT VNLPSKVRAF TFPSEVPSKA KVVWEKFSSQ PKFERDLFYH
901  GTRDDEFLWG VSSSAYQIEG AWDADGKGPS IWDNFTHTPG SNVKDNATGD IACDSYHQLD
961  ADLNMRLALK VKAYRFSISW SRIFPTGRNS SINSHGVDYY NRLINGLVAS NIFPMVTLFH
1021  WDL PQALQDI GGWENPALID LFDSYADFCF QTFGDRVKFW MTFNEP MYLA WLGYGSGEFP
1081  PGVKDPGWAP YRIAHAVIKA HARVYHTYDE KYRQEQKGI SLSLSTHWAE PKSPGVPRDV
1141  EAADRMLQFS LGWFAHP IFRNGDYPDTMKV KVGNRSELQH LATSRLPSFT EEKRFIRAT
1201  ADVFCLNTYY SRIVQHKT PR LNPPSYEDDQ EMAEEDPSW PSTAMNRAAP WGT RLLNWI
1261  KEEYGDPIY ITENGVL TNPTEDTDRI F YHKTYINEAL KAYRLDGIDL RGYVAWSLMD
1321  NFEWLN GYTV KFGLYHVD FNNTNRPR TARA SARYYTEVIT NNGMPLARE D EFLYGRFPEG
1381  FIWSAASAAY QIEGAWRADG KGLSIWDTFS HTPLRVENDA IGDVACDSYH KIAEDLVTLQ
1441  NLGVSHYRFS ISWSRILPDG TTRYINEAGL NYYVRLIDTL LAASIQPQVT IYHWDLPQTL
1501  QDVGGWENET IVQRFKEYAD VLFQRLGDKV KFWITLNEPF VIAYQGYGYG TAAPGVSNRP
1561  GTAPYIVGHN LIKAHA EAWHLYNDVYRASQ GGVISITISS DWAEPRDPSN QEDVEAARRY
1621  VQFMGGWFAH PIFKNGDYNE VMKTRIRD RS LAAGLNKSRL PEFTSEKRR INGTYDFFGF
1681  NHYTTVLAYN LNYATAISSF DADRGV ASIA DRSWPDSGSF WLKMTFPGFR RILNLWKEEY
1741  NDPPIYVTEN GVSQREETDL NDTARIYYLR TYINEALKAV QDKVDLRGYT VWSAMDNFEW
1801  ATGFSERFGL HFVNYS DPSLPRIPKASAKF YASVVR CNGPDPATGP HAC LHQP DAGPTI
1861  SPVRQEEVQF LGLMLGTTEA QTALYVLFSL VLLGVCGLAF LSYKYCKRSK QGKTQRSQQE
1921  LSPVSSF

```

[0074] SEQ ID NO: 10 is the human wild type nucleic acid sequence corresponding to the lactase (LCT) enzyme (bps 1-6274) having GenBank Accession No. NM_002299:

```

1  gttcctagaa aatggagctg tcttgcatg tagtctttat tgccctgcta agtttttcat
61  gctgggggtc agactgggag tctgatagaa atttcatttc caccgtggt cctctaacca
121  atgacttgct gcacaacctg agtgggtctc tgggagacca gagttctaac tttgtagcac
181  gggacaaaga catgtatgtt tgtcaccagc cactgcccac tttcctgcca gaatacttca
241  gcagtctcca tgccagtcag atcaccatt ataaggtatt tctgtcatgg gcacagctcc
301  tcccagcagg aagcaccagc aatccagacg agaaaacagt gcagtgtac cggcgactcc
361  tcaaggccct caagactgca cggcttcagc ccatggtcat cctgcaccac cagaccctcc
421  ctgccagcac cctccggaga accgaagcct ttgctgacct cttcgccgac tatgccacat
481  tcgccttcca ctccttcggg gacctagttg ggatctggtt caccttcagt gacttggagg
541  aagtgatcaa ggagcttccc caccaggaat caagagcgtc acaactccag accctcagtg
601  atgcccacag aaaagcctat gagatttacc acgaaagcta tgcttttcag ggcggaaaac

```

```

661 tctctgttgt cctgcgagct gaagatatcc cggagctcct gctagaacca cccatatctg
721 cgcttgccca ggacacggtc gatttcctct ctcttgattt gtcttatgaa tgccaaaatg
781 aggcaagtct gcggcagaag ctgagtaaat tgcagaccat tgagccaaaa gtgaaagttt
841 tcatcttcaa cctaaaactc ccagactgcc cctccaccat gaagaacca gccagtctgc
901 tcttcagcct ttttgaagcc ataaataaag accaagtgtc caccattggg tttgatatta
961 atgagtttct gagttgttca tcaagttcca agaaaagcat gtcttgttct ctgactggca
1021 gcctggccct tcagcctgac cagcagcagg accacgagac cacggactcc tctcctgcct
1081 ctgcctatca gagaatctgg gaagcatttg ccaatcagtc cagggcggaa agggatgcct
1141 tcctgcagga tactttccct gaaggcttcc tctgggggtgc ctccacagga gcctttaacg
1201 tggaaggagg ctgggccgag ggtgggagag ggggtgagcat ctgggatcca cgcaggcccc
1261 tgaacaccac tgagggccaa gcgacgctgg aggtggccag cgacagttac cacaaggtag
1321 cctctgacgt cgcctgctt tgcggcctcc gggctcaggt gtacaagttc tccatctcct
1381 ggtcccggat ctccccatg gggcacggga gcagccccag cctcccaggc gtgacctact
1441 acaacaagct gattgacagg ctacaggatg cgggcatcga gcccatggcc acgctgttcc
1501 actgggacct gcctcaggcc ctgcaggatc atggtggatg gcagaatgag agcgtggtgg
1561 atgccttcct ggactatgag gccttctgct tctccacatt tggggaccgt gtgaagctgt
1621 ggttgacctt ccatgagccg tgggtgatga gctacgcagg ctatggcacc ggccagcacc
1681 ctcccgccat ctctgaccca ggagtggcct cttttaaggt ggctcacttg gtcctcaagg
1741 ctcatgccag aacttggcac cactacaaca gccatcatcg ccacagcag caggggcacg
1801 tgggcattgt gctgaactca gactgggcag aaccctgtc tccagagagg cctgaggacc
1861 tgagagcctc tgagcgcttc ttgcaattca tgctgggctg gtttgacac cccgtctttg
1921 tggatggaga ctaccagcc accctgagga ccagatcca acagatgaac agacagtgtc
1981 cccatcctgt ggctcaactc cccgagttca cagaggcaga gaagcagctc ctgaaaggct
2041 ctgctgattt tctgggtctg tcgcattaca cctccgcct catcagcaac gcccacaaa
2101 acacctgcat ccctagctat gataccattg gaggtttctc ccaacacgtg aaccatgtgt
2161 ggccccagac ctcatcctct tggattcgtg tgggtgccctg ggggataagg aggctgttgc
2221 agtttgtatc cctggaatac acaagaggaa aagttccaat ataccttgc gggaatggca
2281 tgcccatagg ggaaagtga aatctctttg atgattcctt aagagtagac tacttcaatc
2341 aatatatcaa tgaggtgctc aaggctatca aggaagactc tgtgatgtt cgttcctaca
2401 ttgctcgctt cctcattgat ggcttcgaag gcccttctgg ttacagccag cgttttggcc
2461 tgcaccacgt caacttcagc gacagcagca agtcaaggac tcccaggaaa tctgcctact
2521 ttttactag catcatagaa aagaacggtt tcctcaccaa gggggcaaaa agactgctac
2581 cacctaatac agtaaaccct cctccaaag tcagagcctt cacttttcca tctgaggtgc
2641 cctccaaggc taaagtcgtt tgggaaaagt tctccagcca acccaagttc gaaagagatt
2701 tgttctacca cgggacgttt cgggatgact ttctgtgggg cgtgtcctct tccgcttacc
2761 agattgaagg cgcgtgggat gccgatggca aaggccccag catctgggat aactttaccc
2821 acacaccagg gagcaatgtg aaagacaatg ccactggaga catcgctgtg gacagctatc
2881 accagctgga tgccgatctg aatatgctcc gagctttgaa ggtgaaggcc taccgcttct
2941 ctatctcctg gtctcggatt ttcccaactg ggagaaacag ctctatcaac agtcatgggg
3001 ttgattatta caacaggctg atcaatggct tgggtggcaag caacatctt cccatggtga
3061 cattgttcca ttgggacctg cccagggccc tccaggatat cggaggctgg gagaatcctg
3121 ccttgattga cttgtttgac agctacgcag acttttgttt ccagaccttt ggtgatagag
3181 tcaagttttg gatgactttt aatgagccca tgtacctggc atggctaggt tatggctcag
3241 gggaaatttc cccaggggtg aaggaccag gctgggcacc atataggata gccacgcccg
3301 tcatcaaagc ccatgccaga gtctatcaca cgtacgatga gaaatacagg caggagcaga
3361 agggggtcat ctcgctgagc ctacgtacac actgggcaga gcccagtcac ccaggggctc
3421 ccagagatgt ggaagccgct gaccgaatgc tgcagttctc cctgggctgg tttgctcacc
3481 ccatttttag aaacggagac tatectgaca ccatgaagtg gaaagtgggg aacaggagtg
3541 aactgcagca cttagccacc tcccgcctgc caagcttcac tgaggaagag aagaggttca
3601 tcagggcgac ggccgacgtc ttctgctca acacgtacta ctccagaatc gtgcagcaca
3661 aaacacccag gctaaaccca cctcctacg aagacgacca ggagatggct gaggaggagg
3721 acccttcgtg gccttccacg gcaatgaaca gagctgcgcc ctgggggacg cgaaggctgc
3781 tgaactggat caaggaagag tatggtgaca tccccattta catcaccgaa aacggagtgg
3841 ggctgaccaa tccgaacacg gaggatactg ataggatatt ttaccacaaa acctacatca
3901 atgaggtttt gaaagcctac aggtcgtatg gtatagacct tcgaggtat gtcgcctggg
3961 ctctgatgga caactttgag tggctaaatg gctacacggt caagtttggg ctgtaccatg
4021 ttgatttcaa caacacgaac aggctcgca cagcaagagc ctccgccagg tactacacag

```

```

4081 aggtcattac caacaacggc atgccactgg ccagggagga tgagtttctg tacggacggt
4141 ttcttgaggg cttcatctgg agtgcagctt ctgctgcata tcagattgaa ggtgcgtgga
4201 gagcagatgg caaaggactc agcatttggg acacgttttc tcacacacca ctgaggggtg
4261 agaacgatgc cattggagac gtggcctgtg acagttatca caagattgct gaggatctgg
4321 tcaccctgca gaacctgggc gtgtcccact accgtttttc catctcctgg tctcgcaccc
4381 tccctgatgg aaccaccagg tacatcaatg aagcgggcct gaactactac gtgaggctca
4441 tcgatacact gctggcggcc agcatccagc cccaggtgac catttaccac tgggacctac
4501 cacagacgct ccaagatgta ggaggctggg agaatgagac catcgtgcag cggtttaagg
4561 agtatgcaga tgtgctcttc cagaggctgg gagacaaggt gaagtttttg atcacgctga
4621 atgagccctt tgtcattgct taccagggtt atggctacgg aacagcagct ccaggagtct
4681 ccaataggcc tggcactgcc ccctacattg ttggccacaa tctaataaag gctcatgctg
4741 aggcctggca tctgtacaac gatgtgtacc ggcagagtc aagtggtgctg atttccatca
4801 ccatacagcag tgactgggct gaaccagag atccctctaa ccaggaggat gtggaggcag
4861 ccaggagata tgttcagttc atgggaggct ggtttgaca tcctattttc aagaatggag
4921 attacaatga ggtgatgaag acgcggtacc gtgacaggag cttggctgca ggcctcaaca
4981 agtctcggct gccagaattt acagagagtg agaagaggag gatcaacggc acctatgact
5041 tttttgggtt caatcactac accactgtcc tcgcctacaa cctcaacta gccactgcca
5101 tctcttcttt tgatgcagac agaggagtgt cttccatcgc agatcgctcg tggccagact
5161 ctggctcctt ctggctgaag atgacgcctt ttggcttcag gaggatcctg aactggttaa
5221 aggaaggaata caatgaccct ccaatttatg tcacagagaa tggagtgtcc cagcgggaag
5281 aaacagacct caatgacact gcaaggatct actaccttcg gacttacatc aatgaggccc
5341 tcaaagctgt gcaggacaag gtggaccttc gaggatacac agtttgaggt gcgatggaca
5401 attttgagtg ggccacaggc ttttcagaga gatttggtct gcattttgtg aactacagtg
5461 acccttctct gccaaaggatc cccaaagcat cagcgaagtt ctacgcctct gtggtccgat
5521 gcaatggctt ccctgacccc gctacagggc ctcacgcttg tctccaccag ccagatgctg
5581 gacccaccat cagccccgtg agacaggagg aggtgcagtt cctggggcta atgctcggca
5641 ccacagaagc acagacagct ttgtacgttc tcttttctct tgtgcttctt ggagtctgtg
5701 gcttggcatt tctgtcatat aagtactgca agcgtctctaa gcaagggaaa acacaacgaa
5761 gccaacagga attgagcccg gtgtcttcat tctgatgagt taccacctca agttctatga
5821 agcaggccta gtttcttcat ctatgtttac cggccaccaa acaccttagg gtcttagact
5881 ctgctgatac tggacttctc cataaagtc tgctgcaccg ttagagatga ctttaactct
5941 gaatgatttc gacttgctga gtaaaatgga aatatctcca tcttgctcca gtatcagagt
6001 tcatttgggc atttgagaag caagtagctc ttgcggaaac gtgtagatac tggctctagt
6061 ggtctgtgaa ccacttaatt gaacttaaca gggctgtttt aagtttcaga gttgttaagg
6121 gttgttaagg gagcaaaaac cgtaaaaatc cttcctataa gaagaaatca actccattgc
6181 atagactgca atatcatctc ctgcccttct gcaagctctc cctagcttca catcttgtgt
6241 tttccagaaa ataaaaacag cagactgtcc tttc

```

[0075] As used herein, a “carbohydrate transporter molecule” means a nucleic acid which encodes a polypeptide that exhibits carbohydrate transporter activity, or a polypeptide or peptidomimetic that exhibits carbohydrate transporter activity. For example, a carbohydrate transporter molecule can include the human GLUT2 protein (e.g., having the amino acid sequence shown in SEQ ID NO: 1), or a variant thereof, such as a fragment thereof, that exhibits carbohydrate transporter activity. For example, a carbohydrate transporter molecule can include the human SGLT1 protein (e.g., having the amino acid sequence shown in SEQ ID NO: 3), or a variant thereof, such as a fragment thereof, that exhibits carbohydrate transporter activity. The nucleic acid can be any type of nucleic acid, including genomic DNA, complementary DNA (cDNA), synthetic or semi-synthetic DNA, as well as any form of corresponding RNA. For

example, a carbohydrate transporter molecule can comprise a recombinant nucleic acid encoding human GLUT2 protein or human SGLT1 protein. In one embodiment, a carbohydrate transporter molecule can comprise a non-naturally occurring nucleic acid created artificially (such as by assembling, cutting, ligating or amplifying sequences). A carbohydrate transporter molecule can be double-stranded. A carbohydrate transporter molecule can be single-stranded. The carbohydrate transporter molecules of the invention can be obtained from various sources and can be produced according to various techniques known in the art. For example, a nucleic acid that is a carbohydrate transporter molecule can be obtained by screening DNA libraries, or by amplification from a natural source. The carbohydrate transporter molecules of the invention can be produced via recombinant DNA technology and such recombinant nucleic acids can be prepared by conventional techniques, including chemical synthesis, genetic engineering, enzymatic techniques, or a combination thereof. Non-limiting examples of a carbohydrate transporter molecule, that is a nucleic acid, is the nucleic acid having the nucleotide sequence shown in SEQ ID NO: 2 or SEQ ID NO: 4. Another example of a carbohydrate transporter molecule is a fragment of a nucleic acid having the sequence shown in SEQ ID NO: 2 or SEQ ID NO: 4, wherein the fragment exhibits carbohydrate transporter activity.

[0076] As used herein, a “carbohydrate metabolic enzyme molecule” means a nucleic acid which encodes a polypeptide that exhibits carbohydrate metabolic enzyme activity, or a polypeptide or peptidomimetic that exhibits carbohydrate metabolic enzyme activity. For example, a carbohydrate metabolic enzyme molecule can include the human sucrase-isomaltase (SI) protein (e.g., having the amino acid sequence shown in SEQ ID NO: 5), or a variant thereof, such as a fragment thereof, that exhibits carbohydrate metabolic enzyme activity. For example, a carbohydrate metabolic enzyme molecule can include the human maltase-glucoamylase protein (e.g., having the amino acid sequence shown in SEQ ID NO: 7), or a variant thereof, such as a fragment thereof, that exhibits carbohydrate metabolic enzyme activity. For example, a carbohydrate metabolic enzyme molecule can include the human lactase protein (e.g., having the amino acid sequence shown in SEQ ID NO: 9), or a variant thereof, such as a fragment thereof, that exhibits carbohydrate metabolic enzyme activity. The nucleic acid can be any type of nucleic acid, including genomic DNA, complementary DNA (cDNA), synthetic or semi-synthetic DNA, as well as any form of corresponding RNA. For example, a carbohydrate metabolic enzyme molecule can comprise a recombinant nucleic acid encoding human sucrase-

isomaltase (SI) protein, human maltase-glucoamylase protein, or human lactase protein. In one embodiment, a carbohydrate metabolic enzyme molecule can comprise a non-naturally occurring nucleic acid created artificially (such as by assembling, cutting, ligating or amplifying sequences). A carbohydrate metabolic enzyme molecule can be double-stranded. A carbohydrate metabolic enzyme molecule can be single-stranded. The carbohydrate metabolic enzyme molecules of the invention can be obtained from various sources and can be produced according to various techniques known in the art. For example, a nucleic acid that is a carbohydrate metabolic enzyme molecule can be obtained by screening DNA libraries, or by amplification from a natural source. The carbohydrate metabolic enzyme molecules of the invention can be produced via recombinant DNA technology and such recombinant nucleic acids can be prepared by conventional techniques, including chemical synthesis, genetic engineering, enzymatic techniques, or a combination thereof. A non-limiting example of a carbohydrate metabolic enzyme, that is a nucleic acid, is the nucleic acid having the nucleotide sequence shown in SEQ ID NO: 6, 8, or 10. Another example of a carbohydrate metabolic enzyme molecule is a fragment of a nucleic acid having the sequence shown in SEQ ID NO: 6, 8, or 10, wherein the fragment exhibits carbohydrate metabolic enzyme activity.

[0077] According to this invention, a carbohydrate transporter molecule encompasses orthologs of human GLUT2 and SGLT1. According to this invention, a carbohydrate metabolic enzyme molecule encompasses orthologs of human sucrase-isomaltase (SI), human maltase-glucoamylase, and human lactase. For example, a carbohydrate transporter molecule or a carbohydrate metabolic enzyme molecule encompasses the orthologs in mouse, rat, non-human primates, canines, goat, rabbit, porcine, feline, and horses. In other words, a carbohydrate transporter molecule or a carbohydrate metabolic enzyme molecule can comprise a nucleic acid sequence homologous to the human nucleic acid that encodes a human GLUT2 and SGLT1 protein, or human sucrase-isomaltase (SI), human maltase-glucoamylase, and human lactase protein, respectively, wherein the nucleic acid is found in a different species and wherein that homolog encodes a protein with a glucose transporter function similar to a carbohydrate transporter molecule or an enzymatic function similar to a carbohydrate metabolic enzyme molecule.

[0078] A carbohydrate transporter molecule of this invention also encompasses variants of the human nucleic acid encoding the GLUT2 or SGLT1 proteins that exhibit carbohydrate transporter activity, or variants of the human GLUT2 or SGLT1 proteins that exhibit carbohydrate transporter activity. A carbohydrate transporter molecule of this invention also includes a fragment of the human GLUT2 or SGLT1 nucleic acid which encodes a polypeptide that exhibits carbohydrate transporter activity. A carbohydrate transporter molecule of this invention encompasses a fragment of the human GLUT2 or SGLT1 protein that exhibits carbohydrate transporter activity.

[0079] A carbohydrate metabolic enzyme molecule of this invention also encompasses variants of the human nucleic acid encoding the sucrase-isomaltase (SI), human maltase-glucoamylase, and human lactase proteins that exhibit carbohydrate metabolic enzyme activity, or variants of the human sucrase-isomaltase (SI), human maltase-glucoamylase, and human lactase proteins that exhibit carbohydrate metabolic enzyme activity. A carbohydrate metabolic enzyme molecule of this invention also includes a fragment of the human sucrase-isomaltase (SI), human maltase-glucoamylase, and human lactase nucleic acid which encodes a polypeptide that exhibits carbohydrate metabolic enzyme activity. A carbohydrate metabolic enzyme molecule of this invention encompasses a fragment of the human sucrase-isomaltase (SI), human maltase-glucoamylase, and human lactase protein that exhibits carbohydrate metabolic enzyme activity.

[0080] The variants can comprise, for instance, naturally-occurring variants due to allelic variations between individuals (e.g., polymorphisms), mutated alleles related to autism, or alternative splicing forms. In one embodiment, a carbohydrate transporter molecule is a nucleic acid variant of the nucleic acid having the sequence shown in SEQ ID NO: 2, wherein the variant has a nucleotide sequence identity to SEQ ID NO: 2 of at least about 65%, at least about 75%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% with SEQ ID NO: 2. In another embodiment, a carbohydrate transporter molecule is a nucleic acid variant of the nucleic acid having the sequence shown in SEQ ID NO: 4, wherein the variant has a nucleotide sequence identity to SEQ ID NO: 4 of at least about 65%, at least about 75%, at least about 85%, at least about 90%, at least about 91%,

at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% with SEQ ID NO: 4. In one embodiment, a carbohydrate metabolic enzyme molecule is a nucleic acid variant of the nucleic acid having the sequence shown in SEQ ID NO: 6, wherein the variant has a nucleotide sequence identity to SEQ ID NO: 6 of at least about 65%, at least about 75%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% with SEQ ID NO: 6. In another embodiment, a carbohydrate metabolic enzyme molecule is a nucleic acid variant of the nucleic acid having the sequence shown in SEQ ID NO: 8, wherein the variant has a nucleotide sequence identity to SEQ ID NO: 8 of at least about 65%, at least about 75%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% with SEQ ID NO: 8. In a further embodiment, a carbohydrate metabolic enzyme molecule is a nucleic acid variant of the nucleic acid having the sequence shown in SEQ ID NO: 10, wherein the variant has a nucleotide sequence identity to SEQ ID NO: 10 of at least about 65%, at least about 75%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% with SEQ ID NO: 10.

[0081] In one embodiment, a carbohydrate transporter molecule encompasses any portion of at least about 8 consecutive nucleotides of SEQ ID NO: 2 or 4. In one embodiment, the fragment can comprise at least about 15 nucleotides, at least about 20 nucleotides, or at least about 30 nucleotides of SEQ ID NO: 2 or 4. Fragments include all possible nucleotide lengths between about 8 and 100 nucleotides, for example, lengths between about 15 and 100, or between about 20 and 100. In one embodiment, a carbohydrate metabolic enzyme molecule encompasses any portion of at least about 8 consecutive nucleotides of SEQ ID NO: 6, 8, or 10. In one embodiment, the fragment can comprise at least about 15 nucleotides, at least about 20 nucleotides, or at least about 30 nucleotides of SEQ ID NO: 6, 8, or 10. Fragments include all possible nucleotide lengths between about 8 and 100 nucleotides, for example, lengths between about 15 and 100, or between about 20 and 100.

[0082] The invention further provides for nucleic acids that are complementary to a nucleic acid encoding GLUT2, SGLT1, sucrase-isomaltase (SI), human maltase-glucoamylase, or human lactase proteins. Such complementary nucleic acids can comprise nucleic acid sequences, which hybridize to a nucleic acid sequence encoding a GLUT2, SGLT1, sucrase-isomaltase (SI), maltase-glucoamylase, or lactase protein under stringent hybridization conditions. Non-limiting examples of stringent hybridization conditions include temperatures above 30°C, above 35°C, in excess of 42°C, and/or salinity of less than about 500 mM, or less than 200 mM. Hybridization conditions can be adjusted by the skilled artisan via modifying the temperature, salinity and/or the concentration of other reagents such as SDS or SSC.

[0083] In one embodiment, a carbohydrate transporter molecule comprises a protein or polypeptide encoded by a carbohydrate transporter nucleic acid sequence, such as the sequence shown in SEQ ID NO: 2 or 4. In another embodiment, the polypeptide can be modified, such as by glycosylations and/or acetylations and/or chemical reaction or coupling, and can contain one or several non-natural or synthetic amino acids. An example of a carbohydrate transporter molecule is the polypeptide having the amino acid sequence shown in SEQ ID NO: 1 or 3. In one embodiment, a carbohydrate metabolic enzyme molecule comprises a protein or polypeptide encoded by a carbohydrate metabolic enzyme nucleic acid sequence, such as the sequence shown in SEQ ID NO: 6, 8, or 10. In another embodiment, the polypeptide can be modified, such as by glycosylations and/or acetylations and/or chemical reaction or coupling, and can contain one or several non-natural or synthetic amino acids. An example of a carbohydrate transporter molecule is the polypeptide having the amino acid sequence shown in SEQ ID NO: 5, 7, or 9.

[0084] In another embodiment, a carbohydrate transporter molecule can be a fragment of a carbohydrate transporter protein, such as GLUT2 or SGLT1. For example, the carbohydrate transporter molecule can encompass any portion of at least about 8 consecutive amino acids of SEQ ID NO: 1 or 3. The fragment can comprise at least about 10 amino acids, at least about 20 amino acids, at least about 30 amino acids, at least about 40 amino acids, at least about 50 amino acids, at least about 60 amino acids, or at least about 75 amino acids of SEQ ID NO: 1 or 3. In another embodiment, a carbohydrate metabolic enzyme molecule can be a fragment of a carbohydrate metabolic enzyme protein, such as sucrase-isomaltase (SI), maltase-glucoamylase, or lactase. For example, the carbohydrate metabolic enzyme molecule can encompass any

portion of at least about 8 consecutive amino acids of SEQ ID NO: 5, 7, or 9. The fragment can comprise at least about 10 amino acids, at least about 20 amino acids, at least about 30 amino acids, at least about 40 amino acids, at least about 50 amino acids, at least about 60 amino acids, or at least about 75 amino acids of SEQ ID NO: 5, 7, or 9. Fragments include all possible amino acid lengths between about 8 and 100 amino acids, for example, lengths between about 10 and 100 amino acids, between about 15 and 100 amino acids, between about 20 and 100 amino acids, between about 35 and 100 amino acids, between about 40 and 100 amino acids, between about 50 and 100 amino acids, between about 70 and 100 amino acids, between about 75 and 100 amino acids, or between about 80 and 100 amino acids.

[0085] In certain embodiments, the carbohydrate transporter molecule of the invention includes variants of the human GLUT2 or SGLT1 protein (having the amino acid sequence shown in SEQ ID NO: 1 and 3, respectively). Such variants can include those having at least from about 46% to about 50% identity to SEQ ID NO: 1 or 3, or having at least from about 50.1% to about 55% identity to SEQ ID NO: 1 or 3, or having at least from about 55.1% to about 60% identity to SEQ ID NO: 1 or 3, or having from at least about 60.1% to about 65% identity to SEQ ID NO: 1 or 3, or having from about 65.1% to about 70% identity to SEQ ID NO: 1 or 3, or having at least from about 70.1% to about 75% identity to SEQ ID NO: 1 or 3, or having at least from about 75.1% to about 80% identity to SEQ ID NO: 1 or 3, or having at least from about 80.1% to about 85% identity to SEQ ID NO: 1 or 3, or having at least from about 85.1% to about 90% identity to SEQ ID NO: 1 or 3, or having at least from about 90.1% to about 95% identity to SEQ ID NO: 1 or 3, or having at least from about 95.1% to about 97% identity to SEQ ID NO: 1 or 3, or having at least from about 97.1% to about 99% identity to SEQ ID NO: 1 or 3.

[0086] In certain embodiments, the carbohydrate metabolic enzyme molecule of the invention includes variants of the human sucrase-isomaltase (SI), maltase-glucoamylase, or lactase protein (having the amino acid sequence shown in SEQ ID NO: 5, 7, and 9, respectively). Such variants can include those having at least from about 46% to about 50% identity to SEQ ID NO: 5, 7, or 9, or having at least from about 50.1% to about 55% identity to SEQ ID NO: 5, 7, or 9, or having at least from about 55.1% to about 60% identity to SEQ ID NO: 5, 7, or 9, or having from at least about 60.1% to about 65% identity to SEQ ID NO: 5, 7, or 9, or having from about 65.1% to about 70% identity to SEQ ID NO: 5, 7, or 9, or having at least from about 70.1% to

about 75% identity to SEQ ID NO: 5, 7, or 9, or having at least from about 75.1% to about 80% identity to SEQ ID NO: 5, 7, or 9, or having at least from about 80.1% to about 85% identity to SEQ ID NO: 5, 7, or 9, or having at least from about 85.1% to about 90% identity to SEQ ID NO: 5, 7, or 9, or having at least from about 90.1% to about 95% identity to SEQ ID NO: 5, 7, or 9, or having at least from about 95.1% to about 97% identity to SEQ ID NO: 5, 7, or 9, or having at least from about 97.1% to about 99% identity to SEQ ID NO: 5, 7, or 9.

[0087] In another embodiment, the carbohydrate transporter molecule of the invention encompasses a peptidomimetic which exhibits carbohydrate transporter activity. In another embodiment, the carbohydrate transporter molecule of the invention encompasses a peptidomimetic which exhibits carbohydrate transporter activity. In another embodiment, the carbohydrate metabolic enzyme molecule of the invention encompasses a peptidomimetic which exhibits carbohydrate metabolic enzyme activity. In another embodiment, the carbohydrate metabolic enzyme molecule of the invention encompasses a peptidomimetic which exhibits carbohydrate metabolic enzyme activity. A peptidomimetic is a small protein-like chain designed to mimic a peptide that can arise from modification of an existing peptide in order to protect that molecule from enzyme degradation and increase its stability, and/or alter the molecule's properties (for example modifications that change the molecule's stability or biological activity). These modifications involve changes to the peptide that can not occur naturally (such as altered backbones and the incorporation of non-natural amino acids). Drug-like compounds may be able to be developed from existing peptides. A peptidomimetic can be a peptide, partial peptide or non-peptide molecule that mimics the tertiary binding structure or activity of a selected native peptide or protein functional domain (e.g., binding motif or active site). These peptide mimetics include recombinantly or chemically modified peptides.

[0088] In one embodiment, a carbohydrate transporter molecule comprising SEQ ID NO: 1, SEQ ID NO: 3, variants of each, or fragments thereof, can be modified to produce peptide mimetics by replacement of one or more naturally occurring side chains of the 20 genetically encoded amino acids (or D amino acids) with other side chains. In one embodiment, a carbohydrate metabolic enzyme molecule comprising SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, variants of , or fragments thereof, can be modified to produce peptide mimetics by replacement of one or more naturally occurring side chains of the 20 genetically encoded amino

acids (or D amino acids) with other side chains. This can occur, for instance, with groups such as alkyl, lower alkyl, cyclic 4-, 5-, 6-, to 7-membered alkyl, amide, amide lower alkyl, amide di(lower alkyl), lower alkoxy, hydroxy, carboxy and the lower ester derivatives thereof, and with 4, 5-, 6-, to 7-membered heterocyclics. For example, proline analogs can be made in which the ring size of the proline residue is changed from 5 members to 4, 6, or 7 members. Cyclic groups can be saturated or unsaturated, and if unsaturated, can be aromatic or non-aromatic.

Heterocyclic groups can contain one or more nitrogen, oxygen, and/or sulphur heteroatoms. Examples of such groups include the furazanyl, ifuryl, imidazolidinyl, imidazolyl, imidazolinyl, isothiazolyl, isoxazolyl, morpholinyl (e.g. morpholino), oxazolyl, piperazinyl (e.g. 1-piperazinyl), piperidyl (e.g. 1-piperidyl, piperidino), pyranyl, pyrazinyl, pyrazolidinyl, pyrazolinyl, pyrazolyl, pyridazinyl, pyridyl, pyrimidinyl, pyrrolidinyl (e.g. 1-pyrrolidinyl), pyrrolinyl, pyrrolyl, thiadiazolyl, thiazolyl, thienyl, thiomorpholinyl (e.g. thiomorpholino), and triazolyl. These heterocyclic groups can be substituted or unsubstituted. Where a group is substituted, the substituent can be alkyl, alkoxy, halogen, oxygen, or substituted or unsubstituted phenyl. Peptidomimetics may also have amino acid residues that have been chemically modified by phosphorylation, sulfonation, biotinylation, or the addition or removal of other moieties. For example, peptidomimetics can be designed and directed to amino acid sequences encoded by a carbohydrate transporter molecule comprising SEQ ID NO: 1 or 3. For example, peptidomimetics can be designed and directed to amino acid sequences encoded by a carbohydrate metabolic enzyme molecule comprising SEQ ID NO: 5, 7, or 9.

[0089] A variety of techniques are available for constructing peptide mimetics with the same or similar desired biological activity as the corresponding native but with more favorable activity than the peptide with respect to solubility, stability, and/or susceptibility to hydrolysis or proteolysis (see, e.g., Morgan & Gainor, Ann. Rep. Med. Chem. 24,243-252, 1989). Certain peptidomimetic compounds are based upon the amino acid sequence of the peptides of the invention. Peptidomimetic compounds can be synthetic compounds having a three-dimensional structure (i.e. a peptide motif) based upon the three-dimensional structure of a selected peptide. The peptide motif provides the peptidomimetic compound with the desired biological activity, wherein the binding activity of the mimetic compound is not substantially reduced, and is often the same as or greater than the activity of the native peptide on which the mimetic is modeled. Peptidomimetic compounds can have additional characteristics that enhance their therapeutic

application, such as increased cell permeability, greater affinity and/or avidity and prolonged biological half-life. Peptidomimetic design strategies are readily available in the art (see, e.g., Ripka & Rich, Curr. Op. Chem. Biol. 2, 441452, 1998; Hruby et al., Curr. Op. Chem. Biol. 1, 114119, 1997; Hruby & Balse, Curr. Med. Chem. 9, 945-970, -2000).

[0090] Diagnosis

[0091] The invention provides diagnosis methods based on monitoring a gene encoding a carbohydrate metabolic enzyme molecule (such as sucrase isomaltase, maltase glucoamylase, or lactase) or a carbohydrate transporter molecule (such as GLUT2 or SGLT1). As used herein, the term “diagnosis” includes the detection, typing, monitoring, dosing, comparison, at various stages, including early, pre-symptomatic stages, and late stages, in adults, children, and unborn human children. Diagnosis can include the assessment of a predisposition or risk of development, the prognosis, or the characterization of a subject to define most appropriate treatment (pharmacogenetics).

[0092] The invention provides diagnostic methods to determine whether an individual is at risk of developing autism or an autism spectrum disorder (ASD), or suffers from autism or an ASD, wherein the disease reflects an alteration in the expression of a gene encoding a carbohydrate metabolic enzyme molecule (such as sucrase isomaltase, maltase glucoamylase, or lactase) or a carbohydrate transporter molecule (such as GLUT2 or SGLT1). Subjects diagnosed with autism, as well as ASD, can display some core symptoms in the areas of a) social interactions and relationships, b) verbal and non-verbal communication, and c) physical activity, play, physical behavior. For example, symptoms related to social interactions and relationships can include but are not limited to the inability to establish friendships with children the same age, lack of empathy, and the inability to develop nonverbal communicative skills (for example, eye-to-eye gazing, facial expressions, and body posture). For example, symptoms related to verbal and nonverbal communication comprises delay in learning to talk, inability to learn to talk, failure to initiate or maintain a conversation, failure to interpret or understand implied meaning of words, and repetitive use of language. For example, symptoms related to physical activity, play, physical behavior include, but are not limited to unusual focus on pieces or parts of an

object, such as a toy, a preoccupation with certain topics, a need for routines and rituals, and stereotyped behaviors (for example, body rocking and hand flapping).

[0093] In one embodiment, a method of detecting the presence of or a predisposition to autism or an autism spectrum disorder in a subject is provided. The subject can be a human or a child thereof. The subject can also be a human embryo, a human fetus, or an unborn human child. The method can comprise detecting in a sample from the subject the presence of an alteration in the expression of a gene of a carbohydrate metabolic enzyme molecule (such as sucrase isomaltase, maltase glucoamylase, or lactase) or a carbohydrate transporter molecule (such as GLUT2 or SGLT1). In one embodiment, the detecting comprises detecting whether there is an alteration in the gene locus encoding a carbohydrate metabolic enzyme molecule (such as sucrase isomaltase, maltase glucoamylase, or lactase) or a carbohydrate transporter molecule (such as GLUT2 or SGLT1). In a further embodiment, the detecting comprises detecting whether expression of a carbohydrate metabolic enzyme molecule (such as sucrase isomaltase, maltase glucoamylase, or lactase) or a carbohydrate transporter molecule (such as GLUT2 or SGLT1) is reduced. In some embodiments, the detecting comprises detecting in the sample whether there is a reduction in an mRNA encoding a carbohydrate metabolic enzyme molecule or a carbohydrate transporter molecule, or a reduction in either the carbohydrate metabolic enzyme protein or a carbohydrate transporter protein, or a combination thereof. The presence of such an alteration is indicative of the presence or predisposition to autism or an autism spectrum disorder. The presence of an alteration in a gene encoding a carbohydrate metabolic enzyme molecule or a carbohydrate transporter molecule in the sample is detected through the genotyping of a sample, for example via gene sequencing, selective hybridization, amplification, gene expression analysis, or a combination thereof. In one embodiment, the sample can comprise blood, serum, sputum, lacrimal secretions, semen, vaginal secretions, fetal tissue, skin tissue, ileum tissue, cecum tissue, muscle tissue, amniotic fluid, or a combination thereof.

[0094] The invention also provides a method for treating or preventing autism or an autism spectrum disorder in a subject. In one embodiment, the method comprises (1) detecting the presence of an alteration in a carbohydrate transporter gene or a carbohydrate metabolic enzyme in a sample from the subject, where the presence of the alteration is indicative of autism or an

ASD, or the predisposition to autism or ASD, and, (2) administering to the subject in need a therapeutic treatment against autism or an autism spectrum disorder. In one embodiment, the carbohydrate transporter gene can be a GLUT2 gene or a SGLT1 gene. In another embodiment, the carbohydrate metabolic enzyme gene can be a sucrase isomaltase gene, a maltase glucoamylase gene, or a lactase gene. The therapeutic treatment can be a drug administration (for example, a pharmaceutical composition comprising a functional carbohydrate transporter molecule or a functional carbohydrate metabolic enzyme molecule). In one embodiment, the molecule comprises a carbohydrate transporter polypeptide comprising at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or 100% of the amino acid sequence of SEQ ID NO: 1 or SEQ ID NO: 3, and exhibits the function of restoring functional carbohydrate transporter expression in deficient individuals, thus restoring the capacity for carbohydrate transport. In another embodiment, the molecule comprises a carbohydrate metabolic enzyme polypeptide comprising at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or 100% of the amino acid sequence of SEQ ID NO: 5, 7, or 9, and exhibits the function of restoring functional carbohydrate metabolic enzyme expression in deficient individuals, thus restoring the capacity for carbohydrate metabolism.

[0095] In some embodiments, the molecule comprises a nucleic acid encoding a carbohydrate transporter polypeptide comprising at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or 100% of the nucleic acid sequence of SEQ ID NO: 2 or 4 and encodes a polypeptide with the function of restoring functional carbohydrate transporter expression in deficient individuals, thus restoring the capacity for carbohydrate transport. In further embodiments, the molecule comprises a nucleic acid encoding a carbohydrate metabolic enzyme polypeptide comprising at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or 100% of the nucleic acid sequence of SEQ ID NO: 6, 8, or 10, and encodes a polypeptide with the function of restoring functional carbohydrate metabolic enzyme expression in deficient individuals, thus restoring the capacity for carbohydrate metabolism.

[0096] The alteration can be determined at the DNA, RNA or polypeptide level of the carbohydrate transporter or carbohydrate metabolic enzyme. The detection can also be determined by performing an oligonucleotide ligation assay, a confirmation based assay, a hybridization assay, a sequencing assay, an allele-specific amplification assay, a microsequencing assay, a melting curve analysis, a denaturing high performance liquid chromatography (DHPLC) assay (for example, see Jones et al, (2000) Hum Genet., 106(6):663-8), or a combination thereof. In some embodiments, the detection is performed by sequencing all or part of a carbohydrate transporter or carbohydrate metabolic enzyme gene or by selective hybridization or amplification of all or part of a carbohydrate transporter or carbohydrate metabolic enzyme gene. A carbohydrate transporter or carbohydrate metabolic enzyme gene specific amplification can be carried out before the alteration identification step.

[0097] An alteration in a carbohydrate transporter gene locus (e.g., where GLUT2 or SGLT1 are located) or a carbohydrate metabolic enzyme gene locus (e.g., where SI, MGAM, or LCT are located) can be any form of mutation(s), deletion(s), rearrangement(s) and/or insertions in the coding and/or non-coding region of the locus, alone or in various combination(s). Mutations can include point mutations. Insertions can encompass the addition of one or several residues in a coding or non-coding portion of the gene locus. Insertions can comprise an addition of between 1 and 50 base pairs in the gene locus. Deletions can encompass any region of one, two or more residues in a coding or non-coding portion of the gene locus, such as from two residues up to the entire gene or locus. Deletions can affect smaller regions, such as domains (introns) or repeated sequences or fragments of less than about 50 consecutive base pairs, although larger deletions can occur as well. Rearrangement includes inversion of sequences. The carbohydrate transporter gene locus alteration or carbohydrate metabolic enzyme gene locus alteration can result in amino acid substitutions, RNA splicing or processing, product instability, the creation of stop codons, frame-shift mutations, and/or truncated polypeptide production. The alteration can result in the production of a carbohydrate transporter polypeptide or a carbohydrate metabolic enzyme with altered function, stability, targeting or structure. The alteration can also cause a reduction in protein expression. In one embodiment, the alteration in a carbohydrate transporter gene locus can comprise a point mutation, a deletion, or an insertion in the carbohydrate transporter gene or corresponding expression product. In another embodiment, the alteration in a carbohydrate metabolic enzyme gene locus can comprise

a point mutation, a deletion, or an insertion in the carbohydrate metabolic enzyme gene or corresponding expression product. In one embodiment, the alteration can be a deletion or partial deletion of a carbohydrate transporter gene or a carbohydrate metabolic enzyme gene. The alteration can be determined at the level of the DNA, RNA, or polypeptide of a carbohydrate transporter or a carbohydrate metabolic enzyme.

[0098] In another embodiment, the method can comprise detecting the presence of an altered RNA expression of a carbohydrate transporter or a carbohydrate metabolic enzyme. Altered RNA expression includes the presence of an altered RNA sequence, the presence of an altered RNA splicing or processing, or the presence of an altered quantity of RNA. These can be detected by various techniques known in the art, including by sequencing all or part of the RNA of a carbohydrate transporter or a carbohydrate metabolic enzyme, or by selective hybridization or selective amplification of all or part of the RNA. In a further embodiment, the method can comprise detecting the presence of an altered polypeptide expression of a carbohydrate transporter or a carbohydrate metabolic enzyme. Altered polypeptide expression includes the presence of an altered polypeptide sequence, the presence of an altered quantity of carbohydrate transporter polypeptide or carbohydrate metabolic enzyme polypeptide, or the presence of an altered tissue distribution. These can be detected by various techniques known in the art, including by sequencing and/or binding to specific ligands (such as antibodies).

[0099] Various techniques known in the art can be used to detect or quantify altered gene expression, RNA expression, or sequence, which include, but are not limited to, hybridization, sequencing, amplification, and/or binding to specific ligands (such as antibodies). Other suitable methods include allele-specific oligonucleotide (ASO), oligonucleotide ligation, allele-specific amplification, Southern blot (for DNAs), Northern blot (for RNAs), single-stranded conformation analysis (SSCA), PFGE, fluorescent in situ hybridization (FISH), gel migration, clamped denaturing gel electrophoresis, denaturing HPLC, melting curve analysis, heteroduplex analysis, RNase protection, chemical or enzymatic mismatch cleavage, ELISA, radio-immunoassays (RIA) and immuno-enzymatic assays (IEMA). Some of these approaches (such as SSCA and CGGE) are based on a change in electrophoretic mobility of the nucleic acids, as a result of the presence of an altered sequence. According to these techniques, the altered sequence is visualized by a shift in mobility on gels. The fragments can then be sequenced to

confirm the alteration. Some other approaches are based on specific hybridization between nucleic acids from the subject and a probe specific for wild type or altered gene or RNA. The probe can be in suspension or immobilized on a substrate. The probe can be labeled to facilitate detection of hybrids. Some of these approaches are suited for assessing a polypeptide sequence or expression level, such as Northern blot, ELISA and RIA. These latter require the use of a ligand specific for the polypeptide, for example, the use of a specific antibody.

[00100] *Sequencing.* Sequencing can be carried out using techniques well known in the art, using automatic sequencers. The sequencing can be performed on the complete gene or on specific domains thereof, such as those known or suspected to carry deleterious mutations or other alterations.

[00101] *Amplification.* Amplification is based on the formation of specific hybrids between complementary nucleic acid sequences that serve to initiate nucleic acid reproduction. Amplification can be performed according to various techniques known in the art, such as by polymerase chain reaction (PCR), ligase chain reaction (LCR), strand displacement amplification (SDA) and nucleic acid sequence based amplification (NASBA). These techniques can be performed using commercially available reagents and protocols. Useful techniques in the art encompass real-time PCR, allele-specific PCR, or PCR-SSCP. Amplification usually requires the use of specific nucleic acid primers, to initiate the reaction. For example, nucleic acid primers useful for amplifying sequences from the gene or locus of a carbohydrate transporter (such as GLUT2 or SGLT1) or a carbohydrate metabolic enzyme (such as SI, MGAM, or LCT) are able to specifically hybridize with a portion of the gene locus that flanks a target region of the locus, wherein the target region is altered in certain subjects having autism or an autism spectrum disorder. In one embodiment, amplification comprises using forward and reverse RT-PCR primers comprising nucleotide sequences of SEQ ID NOS: 26, 27, 29, 30, 32, 33, 35, 36, 38, or 39.

[00102] The invention provides for a nucleic acid primer, wherein the primer can be complementary to and hybridize specifically to a portion of a coding sequence (e.g., gene or RNA) of a carbohydrate transporter (such as GLUT2 or SGLT1) or a carbohydrate metabolic enzyme (such as SI, MGAM, or LCT) that is altered in certain subjects having autism or an

autism spectrum disorder. Primers of the invention can thus be specific for altered sequences in a gene or RNA of a carbohydrate transporter or a carbohydrate metabolic enzyme. By using such primers, the detection of an amplification product indicates the presence of an alteration in the gene or the absence of such gene. Examples of primers of this invention can be single-stranded nucleic acid molecules of about 5 to 60 nucleotides in length, or about 8 to about 25 nucleotides in length. The sequence can be derived directly from the sequence of the carbohydrate transporter or the carbohydrate metabolic enzyme gene (e.g., GLUT2 or SGLT1, and SI, MGAM, or LCT, respectively). Perfect complementarity is useful, to ensure high specificity. However, certain mismatch can be tolerated. In one embodiment, the primer can be an isolated nucleic acid comprising a nucleotide sequence of SEQ ID NOS: 26, 27, 29, 30, 32, 33, 35, 36, 38, or 39. For example, a nucleic acid primer or a pair of nucleic acid primers as described above can be used in a method for detecting the presence of or a predisposition to autism or an autism spectrum disorder in a subject.

[00103] Amplification methods include, e.g., polymerase chain reaction, PCR (PCR PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS, ed. Innis, Academic Press, N.Y., 1990 and PCR STRATEGIES, 1995, ed. Innis, Academic Press, Inc., N.Y., ligase chain reaction (LCR) (see, e.g., Wu, Genomics 4:560, 1989; Landegren, Science 241:1077, 1988; Barringer, Gene 89:117, 1990); transcription amplification (see, e.g., Kwok, Proc. Natl. Acad. Sci. USA 86:1173, 1989); and, self-sustained sequence replication (see, e.g., Guatelli, Proc. Natl. Acad. Sci. USA 87:1874, 1990); Q Beta replicase amplification (see, e.g., Smith, J. Clin. Microbiol. 35:1477-1491, 1997), automated Q-beta replicase amplification assay (see, e.g., Burg, Mol. Cell. Probes 10:257-271, 1996) and other RNA polymerase mediated techniques (e.g., NASBA, Cangene, Mississauga, Ontario); see also Berger, Methods Enzymol. 152:307-316, 1987; Sambrook; Ausubel; U.S. Pat. Nos. 4,683,195 and 4,683,202; Sooknanan, Biotechnology 13:563-564, 1995. All the references stated above are incorporated by reference in their entireties.

[00104] *Selective Hybridization.* Hybridization detection methods are based on the formation of specific hybrids between complementary nucleic acid sequences that serve to detect nucleic acid sequence alteration(s). A detection technique involves the use of a nucleic acid probe specific for wild type or altered gene or RNA, followed by the detection of the presence of

a hybrid. The probe can be in suspension or immobilized on a substrate or support (for example, as in nucleic acid array or chips technologies). The probe can be labeled to facilitate detection of hybrids. In one embodiment, the probe according to the invention can comprise a nucleic acid having SEQ ID NOS: 28, 31, 34, 37, or 40. For example, a sample from the subject can be contacted with a nucleic acid probe specific for a wild type carbohydrate transporter or carbohydrate metabolic enzyme gene or an altered carbohydrate transporter or carbohydrate metabolic enzyme gene, and the formation of a hybrid can be subsequently assessed. In one embodiment, the method comprises contacting simultaneously the sample with a set of probes that are specific, respectively, for the wild type carbohydrate transporter or carbohydrate metabolic enzyme gene and for various altered forms thereof. Thus, it is possible to detect directly the presence of various forms of alterations in the carbohydrate transporter gene (e.g., GLUT2 or SGLT1) or carbohydrate metabolic enzyme gene (e.g., SI, MGAM, or LCT) in the sample. Also, various samples from various subjects can be treated in parallel.

[00105] According to the invention, a probe can be a polynucleotide sequence which is complementary to and specifically hybridizes with a, or a target portion of a, carbohydrate transporter or carbohydrate metabolic enzyme gene or RNA, and that is suitable for detecting polynucleotide polymorphisms associated with alleles of such, which predispose to or are associated with autism or an autism spectrum disorder. Useful probes are those that are complementary to the carbohydrate transporter or carbohydrate metabolic enzyme gene, RNA, or target portion thereof. Probes can comprise single-stranded nucleic acids of between 8 to 1000 nucleotides in length, for instance between 10 and 800, between 15 and 700, or between 20 and 500. Longer probes can be used as well. A useful probe of the invention is a single stranded nucleic acid molecule of between 8 to 500 nucleotides in length, which can specifically hybridize to a region of a UGT2B17 gene or RNA that carries an alteration.

[00106] The sequence of the probes can be derived from the sequences of the carbohydrate transporter or carbohydrate metabolic enzyme genes provided herein. Nucleotide substitutions can be performed, as well as chemical modifications of the probe. Such chemical modifications can be accomplished to increase the stability of hybrids (e.g., intercalating groups) or to label the probe. Some examples of labels include, without limitation, radioactivity, fluorescence, luminescence, and enzymatic labeling.

[00107] A guide to the hybridization of nucleic acids is found in e.g., Sambrook, ed., MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, 1989; CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley & Sons, Inc., New York, 1997; LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID PROBES, PART I. Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y., 1993.

[00108] *Specific Ligand Binding.* As indicated herein, alteration in a carbohydrate transporter or carbohydrate metabolic enzyme gene locus or in carbohydrate transporter or carbohydrate metabolic enzyme expression can also be detected by screening for alteration(s) in corresponding polypeptide sequence or expression levels. Different types of ligands can be used, such as specific antibodies. In one embodiment, the sample is contacted with an antibody specific for a carbohydrate transporter or carbohydrate metabolic enzyme polypeptide and the formation of an immune complex is subsequently determined. Various methods for detecting an immune complex can be used, such as ELISA, radioimmunoassays (RIA) and immuno-enzymatic assays (IEMA).

[00109] For example, an antibody can be a polyclonal antibody, a monoclonal antibody, as well as fragments or derivatives thereof having substantially the same antigen specificity. Fragments include Fab, Fab'2, or CDR regions. Derivatives include single-chain antibodies, humanized antibodies, or poly-functional antibodies. An antibody specific for a carbohydrate transporter or a carbohydrate metabolic enzyme polypeptide can be an antibody that selectively binds a carbohydrate transporter or carbohydrate metabolic enzyme polypeptide, respectively, namely, an antibody raised against a carbohydrate transporter or carbohydrate metabolic enzyme polypeptide or an epitope-containing fragment thereof. Although non-specific binding towards other antigens can occur, binding to the target polypeptide occurs with a higher affinity and can be reliably discriminated from non-specific binding. In one embodiment, the method comprises contacting a sample from the subject with an antibody specific for a wild type or an altered form of a carbohydrate transporter or carbohydrate metabolic enzyme polypeptide, and determining the presence of an immune complex. Optionally, the sample can be contacted to a support coated with antibody specific for the wild type or altered form of a carbohydrate transporter or

carbohydrate metabolic enzyme polypeptide. In one embodiment, the sample can be contacted simultaneously, or in parallel, or sequentially, with various antibodies specific for different forms of a carbohydrate transporter or carbohydrate metabolic enzyme polypeptide, such as a wild type and various altered forms thereof.

[00110] The invention also provides for a diagnostic kit comprising products and reagents for detecting in a sample from a subject the presence of an alteration in a carbohydrate transporter gene (e.g., GLUT2 or SGLT1) or a carbohydrate metabolic enzyme gene (e.g., SI, MGAM, or LCT), or a carbohydrate transporter polypeptide or carbohydrate metabolic enzyme polypeptide; alteration in the expression of a carbohydrate transporter gene (e.g., GLUT2 or SGLT1) or carbohydrate metabolic enzyme gene (e.g., SI, MGAM, or LCT), or a carbohydrate transporter or carbohydrate metabolic enzyme polypeptide; and/or an alteration in carbohydrate transporter or carbohydrate metabolic enzyme activity. The kit can be useful for determining whether a sample from a subject exhibits reduced carbohydrate transporter or carbohydrate metabolic enzyme expression or exhibits a gene deletion of a carbohydrate transporter (e.g., GLUT2 or SGLT1) or carbohydrate metabolic enzyme (e.g., SI, MGAM, or LCT). For example, the diagnostic kit according to the present invention comprises any primer, any pair of primers, any nucleic acid probe and/or any ligand, (for example, an antibody directed to a carbohydrate transporter or carbohydrate metabolic enzyme). The diagnostic kit according to the present invention can further comprise reagents and/or protocols for performing a hybridization, amplification or antigen-antibody immune reaction. In one embodiment, the kit can comprise nucleic acid primers that specifically hybridize to and can prime a polymerase reaction from a carbohydrate transporter (e.g., GLUT2 or SGLT1) or carbohydrate metabolic enzyme (e.g., SI, MGAM, or LCT). In another embodiment, the primer can comprise a nucleotide sequence of SEQ ID NOS: 26, 27, 29, 30, 32, 33, 35, 36, 38, or 39.

[00111] The diagnosis methods can be performed *in vitro*, *ex vivo*, or *in vivo*. These methods utilize a sample from the subject in order to assess the status of a carbohydrate transporter gene locus or a carbohydrate metabolic enzyme gene locus. The sample can be any biological sample derived from a subject, which contains nucleic acids or polypeptides. Examples of such samples include, but are not limited to, fluids, tissues, cell samples, organs, or tissue biopsies. Non-limiting examples of samples include blood, plasma, saliva, urine, or

seminal fluid. Pre-natal diagnosis can also be performed by testing fetal cells or placental cells, for instance. Screening of parental samples can also be used to determine risk/likelihood of offspring possessing the germline mutation. The sample can be collected according to conventional techniques and used directly for diagnosis or stored. The sample can be treated prior to performing the method, in order to render or improve availability of nucleic acids or polypeptides for testing. Treatments include, for instance, lysis (e.g., mechanical, physical, or chemical), centrifugation. Also, the nucleic acids and/or polypeptides can be pre-purified or enriched by conventional techniques, and/or reduced in complexity. Nucleic acids and polypeptides can also be treated with enzymes or other chemical or physical treatments to produce fragments thereof. In one embodiment, the sample is contacted with reagents, such as probes, primers, or ligands, in order to assess the presence of an altered carbohydrate transporter gene locus or carbohydrate metabolic enzyme gene locus. Contacting can be performed in any suitable device, such as a plate, tube, well, or glass. In specific embodiments, the contacting is performed on a substrate coated with the reagent, such as a nucleic acid array or a specific ligand array. The substrate can be a solid or semi-solid substrate such as any support comprising glass, plastic, nylon, paper, metal, or polymers. The substrate can be of various forms and sizes, such as a slide, a membrane, a bead, a column, or a gel. The contacting can be made under any condition suitable for a complex to be formed between the reagent and the nucleic acids or polypeptides of the sample.

[00112] Identifying an altered polypeptide, RNA or DNA of a carbohydrate transporter (e.g., GLUT2 or SGLT1) or a carbohydrate metabolic enzyme (e.g., SI, MGAM, or LCT) in the sample is indicative of the presence of an altered carbohydrate transporter or carbohydrate metabolic enzyme gene in the subject, which can be correlated to the presence, predisposition or stage of progression of autism or an autism spectrum disorder. For example, an individual having a germ line mutation in a carbohydrate transporter gene (e.g., GLUT 2 or SGLT1) or a carbohydrate metabolic enzyme gene (e.g., SI, MGAM, or LCT) has an increased risk of developing autism or an autism spectrum disorder. The determination of the presence of an altered gene locus in a subject also allows the design of appropriate therapeutic intervention, which is more effective and customized. Also, this determination at the pre-symptomatic level allows a preventive regimen to be applied.

[00113] GI Bacterial Colonization in ASD subjects

[00114] An aspect of the invention provides for a new PCR strategy for the identification, quantitation, and taxonomic classification of *Sutterella* bacterial colonization from biological samples. As shown in Example 2 herein, intestinal biopsies of children with autism accompanied by gastrointestinal (GI) complaints showed highly significant elevation of intestinal levels of *Sutterella* bacteria. These findings may provide insights into pathogenesis of autism associated with GI disorder, enabling new strategies for therapeutic intervention.

[00115] Bacterial members of the genus *Sutterella*, a class of Beta-proteobacteria in the order Burkholderiales and the family *Alcaligenaceae* have been associated with human infections below the diaphragm (A1). Furthermore, *Sutterella sp.* sequences have been identified in intestinal biopsies and fecal samples from individuals with Crohn's disease and ulcerative colitis (A2, A3). *Sutterella sp.* have also been found in canine faeces and the cecal microbiota of domestic and wild turkeys (A4, A5). However, little is known about the pathogenic potential of *Sutterella sp.* According to the *Sutterella sp.*-specific PCR methods described herein, *Sutterella* detection can be achieved in a mammal, such as a dog, a cat, a cow, a horse, a rabbit, a monkey, a pig, a sheep, a goat, a turkey, or a human.

[00116] *Sutterella* bacterial infections have been associated with ASD in addition to Crohn's disease and ulcerative colitis. Bacterial infections are also associated with various intestinal diseases, such as irritable bowel syndrome (IBS). Over 40 million people in the U.S. suffer from irritable bowel syndrome (IBS), a type of inflammatory bowel disease. IBS, though not fatal, has a huge impact on quality-of-life. After the common cold, IBS is the second most common reason for missed work and is estimated to generate \$30B in healthcare costs. Few simple molecular diagnostic tests for IBS/IBD are presently available. Diagnosis usually relies upon symptom analysis and/or invasive colonoscopy procedures. The IBD/IBS diagnostics market has significant growth potential.

[00117] Little is known of the epidemiology and pathogenesis of *Sutterella* infection and their role in Crohn's disease, ASD, and ulcerative colitis. Current methods for *Sutterella* biopsies are costly, laborious and non-specific. There are no known rapid, specific, or cost-effective technologies to identify *Sutterella sp.* in biological samples.

[00118] An aspect of the invention provides for a PCR assay that allows for rapid identification, quantification, classification, and diagnosis of *Sutterella sp.* in biological or industrial samples. This would allow for specific therapies to be implemented in subjects in need (e.g., ASD patients, IB patients, intestinal disease patients, etc.) following identification of *Sutterella* in infections. Directed administration of antimicrobial agents (e.g., antibiotics) that limit the growth of *Sutterella* could be facilitated rapidly following identification of *Sutterella* species. An antibiotic refers to any compound known to one of ordinary skill in the art that will inhibit the growth of, or kill, bacteria. Useful, non-limiting examples of an antibiotic include lincosamides (clindomycin); chloramphenicols; tetracyclines (such as Tetracycline, Chlortetracycline, Demeclocycline, Methacycline, Doxycycline, Minocycline); aminoglycosides (such as Gentamicin, Tobramycin, Netilmicin, Amikacin, Kanamycin, Streptomycin, Neomycin); beta-lactams (such as penicillins, cephalosporins, Imipenem, Aztreonam); vancomycins; bacitracins; macrolides (erythromycins), amphotericins; sulfonamides (such as Sulfanilamide, Sulfamethoxazole, Sulfacetamide, Sulfadiazine, Sulfisoxazole, Sulfacytine, Sulfadoxine, Mafenide, p-Aminobenzoic Acid, Trimethoprim-Sulfamethoxazole); Methenamin; Nitrofurantoin; Phenazopyridine; trimethoprim; rifampicins; metronidazoles; cefazolins; Lincomycin; Spectinomycin; mupirocins; quinolones (such as Nalidixic Acid, Cinoxacin, Norfloxacin, Ciprofloxacin, Perfloxacin, Ofloxacin, Enoxacin, Fleroxacin, Levofloxacin); novobiocins; polymixins; gramicidins; and antipseudomonals (such as Carbenicillin, Carbenicillin Indanyl, Ticarcillin, Azlocillin, Mezlocillin, Piperacillin) or any salts or variants thereof. Such antibiotics can be obtained commercially, e.g., from Daiichi Sankyo, Inc. (Parsippany, NJ), Merck (Whitehouse Station, NJ), Pfizer (New York, NY), Glaxo Smith Kline (Research Triangle Park, NC), Johnson & Johnson (New Brunswick, NJ), AstraZeneca (Wilmington, DE), Novartis (East Hanover, NJ), and Sanofi-Aventis (Bridgewater, NJ). The antibiotic used will depend on the type of bacterial infection.

[00119] In one embodiment, the invention provides for a method of detecting *Sutterella sp.* DNA from biological or industrial sources, e.g. intestinal tissue, feces, blood, or skin. In another embodiment, the invention provides for *Sutterella* diagnostics to detect *Sutterella sp.* in samples from children with autism as well as patients with intestinal disease, e.g. irritable bowel syndrome (IBS). In some embodiments, the invention provides for PCR-based methods of

assessing a subject's response to exposure to therapeutic treatments directed at bacterial infections, for example, *Sutterella* sp. infections, or exposure to other pathogens.

[00120] For example, primers having SEQ ID NOS: 11, 12, 15, or 16 can be used for detecting *Sutterella* sp. DNA. SEQ ID NOS: 17 and 18 can also be used for detecting *Sutterella* sp. DNA.

[00121] Sutt For Primer (SEQ ID NO: 17) -

5'-CGGTGGATGATGTGGATTAATTCGAYGCAACGCGAAAAACCTTAC**S**YIGCC
TTGACATG**Y**C**RR**GA**BB**Y**B**B**V**D**K**RR-3'

[00122] Sutt Rev Primer (SEQ ID NO: 18) -

5'-CCCTCTGTTCCGACCATTGTATGACGTGTGAR**R**G**C**CC**Y**AG**C**C**R**TAAGGGCCA
TGAGGACTT-3'

[00123] Sutt Probe 3 (SEQ ID NO: 19) -

5'-GT**R**CCCG**H**AAG**R**GA**R**Y**Y**Y**G****R**RCACAGGTGCTGCATGGCTGTCGT**C**AGCTCG
TGTCGTG**R**GAT**R**T**Y****R**RGTTA**R**GTCCCGCA**R**C**R**AGCGCAACCCT**Y**G**W**CA**Y**-3'

[00124] In addition to the primers having SEQ ID NOS: 11, 12, and 15-18, additional primers containing any part of SEQ ID NOS: 17, 18, or 19 and containing any portion of the italicized DNA sequence regions may be used to assess the presence or absence of *Sutterella* species. Further, inclusion of degenerate bases (bolded and underlined) may be used to increase coverage of *Sutterella* species (for example, where **S** can be a G nucleotide and/or a C nucleotide; where **Y** can be a C nucleotide and/or T nucleotide; where **R** can be an A nucleotide and/or G nucleotide; where **W** can be an A nucleotide and/or T nucleotide; where **H** can be an A nucleotide and/or T nucleotide and/or C nucleotide; where **B** can be a T nucleotide, C nucleotide, or G nucleotide; where **V** can be an A nucleotide, G nucleotide, or C nucleotide; where **D** can be an A nucleotide, G nucleotide, or T nucleotide; where **K** can be a G nucleotide or T nucleotide).

[00125] In addition to the highlighted probe sequence of SEQ ID NO:19 as well as SEQ ID NOS: 13 and 14, any portion of SEQ ID NO: 19 shown above may be used for *Sutterella* species

detection and quantitation. The reverse complement of SEQ ID NOS: 11, 12, or 15-19 can also be used to detect the opposite DNA strand of *Sutterella* species 16S rRNA genes.

[00126] The invention can be used to detect *Sutterella sp.* 16S rRNA sequences in small amounts of DNA from any biological or industrial source. These sources include, but are not limited to human or animal intestinal tissue, feces, blood, or skin (swabs or tissue). Based on these findings, the invention can be used to detect, quantitate, and classify *Sutterella sp.* in biological samples from children with Autism. In one embodiment, the invention can be used to detect *Sutterella sp.* in biological samples from individuals with various forms of intestinal disease. Intestinal diseases include, but are not limited to, Crohn's disease and Ulcerative colitis. In one embodiment, detection of *Sutterella sp.* can occur in biological samples from any undiagnosed infection below or above the diaphragm. The invention will allow for large cohort investigations of *Sutterella sp.* in the aforementioned, and as yet to be determined, diseases in order to establish an association between *Sutterella sp.* and disease manifestation. In one embodiment, the presence and quantity of *Sutterella sp.* in intestinal tissues can be investigated following any number of experimental manipulations. Experimental manipulations include, but are not limited to, responses to chemicals (i.e. antibiotics), changes in diet, pathogen exposure (i.e. pathogenic viruses, bacteria, fungi), or probiotic usage. The rapid identification of *Sutterella sp.* in human and animal samples facilitated by this invention may lead to rapid diagnosis and directed antimicrobial treatment of infections caused by *Sutterella sp.*

[00127] Gene, Vectors, Recombinant Cells, and Polypeptides

[00128] The invention encompasses an altered or mutated genes of a carbohydrate transporter or carbohydrate metabolic enzyme, or a fragment thereof. The invention also encompasses nucleic acid molecules encoding an altered or mutated polypeptide of s carbohydrate transporter or carbohydrate metabolic enzyme, or a fragment thereof. The alteration or mutation of the nucleotide or amino acid sequence modifies the carbohydrate transporter or carbohydrate metabolic enzyme activity, respectively. The invention provides for a vector that comprises a nucleic acid encoding a carbohydrate transporter or carbohydrate metabolic enzyme polypeptide (for example, a nucleic acid comprising SEQ ID NO: 2 or 4, and a nucleic acid comprising SEQ ID NO: 6, 8, or 10, respectively) or mutant thereof. The vector

can be a cloning vector or an expression vector, i.e., a vector comprising regulatory sequences causing resulting in the expression of carbohydrate transporter or carbohydrate metabolic enzyme polypeptides from the vector in a competent host cell. These vectors can be used to express polypeptides, or mutants thereof, of carbohydrate transporters or carbohydrate metabolic enzymes *in vitro*, *ex vivo*, or *in vivo*, to create transgenic or Knock-Out non-human animals, to amplify the nucleic acids, or to express antisense RNAs.

[00129] The nucleic acids used to practice the invention, whether RNA, RNAi, antisense nucleic acid, cDNA, genomic DNA, vectors, viruses or hybrids thereof, can be produced or isolated from a variety of sources, genetically engineered, amplified, and/or expressed/generated recombinantly. Recombinant polypeptides generated from these nucleic acids can be individually isolated or cloned and tested for a desired activity. Any recombinant expression system can be used, including bacterial, mammalian, yeast, insect or plant cell expression systems. Alternatively, these nucleic acids can be synthesized *in vitro* by well-known chemical synthesis techniques, as described in, e.g., Adams, J. Am. Chem. Soc. 105:661, 1983; Belousov, Nucleic Acids Res. 25:3440-3444, 1997; Frenkel, Free Radic. Biol. Med. 19:373-380, 1995; Blommers, Biochemistry 33:7886-7896, 1994; Narang, Meth. Enzymol. 68:90, 1979; Brown Meth. Enzymol. 68:109, 1979; Beaucage, Tetra. Lett. 22:1859, 1981; U.S. Pat. No. 4,458,066, all of which are incorporated by reference in their entireties.

[00130] The invention provides oligonucleotides comprising sequences of the invention, e.g., subsequences of the exemplary sequences of the invention. Oligonucleotides can include, e.g., single stranded poly-deoxynucleotides or two complementary polydeoxynucleotide strands which can be chemically synthesized.

[00131] Techniques for the manipulation of nucleic acids, such as, subcloning, labeling probes (for example, random-primer labeling using Klenow polymerase, nick translation, amplification), sequencing, and hybridization are well described in the scientific and patent literature, see, e.g., Sambrook, ed., MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, 1989; CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley & Sons, Inc., New York, 1997; LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY:

HYBRIDIZATION WITH NUCLEIC ACID PROBES, Part I. Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y., 1993.

[00132] Nucleic acids, vectors, or polypeptides can be analyzed and quantified by any of a number of general means well known to those of skill in the art. These include, for example, analytical biochemical methods such as radiography, electrophoresis, NMR, spectrophotometry, capillary electrophoresis, thin layer chromatography (TLC), high performance liquid chromatography (HPLC), and hyperdiffusion chromatography; various immunological methods, such as immuno-electrophoresis, Southern analysis, Northern analysis, dot-blot analysis, fluid or gel precipitation reactions, immunodiffusion, quadrature radioimmunoassay (RIAs), enzyme-linked immunosorbent assays (ELISAs), immuno-fluorescent assays, gel electrophoresis (e.g., SDS-PAGE), nucleic acid or target or signal amplification methods, radiolabeling, scintillation counting, and affinity chromatography.

[00133] Obtaining and manipulating nucleic acids used to practice the methods of the invention can be done by cloning from genomic samples, and, if desired, screening and re-cloning inserts isolated or amplified from, e.g., genomic clones or cDNA clones. Sources of nucleic acid used in the methods of the invention include genomic or cDNA libraries contained in, e.g., mammalian artificial chromosomes (MACs), see, e.g., U.S. Pat. Nos. 5,721,118; 6,025,155; human artificial chromosomes, see, e.g., Rosenfeld, Nat. Genet. 15:333-335, 1997; yeast artificial chromosomes (YAC); bacterial artificial chromosomes (BAC); P1 artificial chromosomes, see, e.g., Woon, Genomics 50:306-316, 1998; P1-derived vectors (PACs), see, e.g., Kern, Biotechniques 23:120-124, 1997; cosmids, recombinant viruses, phages or plasmids

[00134] The vectors of this invention can comprise a coding sequence for a carbohydrate transporter molecule or a carbohydrate metabolic enzyme molecule that is operably linked to regulatory sequences, e.g., a promoter, or a polyA tail. Operably linked indicates that the coding and regulatory sequences are functionally associated so that the regulatory sequences cause expression (e.g., transcription) of the coding sequences. The vectors can further comprise one or several origins of replication and/or selectable markers. The promoter region can be homologous or heterologous with respect to the coding sequence, and can provide for ubiquitous, constitutive, regulated and/or tissue specific expression, in any appropriate host cell, including for in vivo use.

Examples of promoters include bacterial promoters (T7, pTAC, Trp promoter), viral promoters (LTR, TK, CMV-IE), mammalian gene promoters (albumin, PGK), etc.

[00135] The vector can be a plasmid, a virus, a cosmid, a phage, a BAC, a YAC. Plasmid vectors can be prepared from commercially available vectors such as pBluescript, pUC, or pBR,. Viral vectors can be produced from baculoviruses, retroviruses, adenoviruses, or AAVs, according to recombinant DNA techniques known in the art. In one embodiment, a recombinant virus can encode a polypeptide of a carbohydrate transporter or carbohydrate metabolic enzyme of the invention. The recombinant virus is useful if replication-defective, for example, if selected from E1- and/or E4-defective adenoviruses, Gag-, pol- and/or env-defective retroviruses and Rep- and/or Cap-defective AAVs. Such recombinant viruses can be produced by techniques known in the art, such as by transfecting packaging cells or by transient transfection with helper plasmids or viruses. Examples of virus packaging cells include PA317 cells, PsiCRIP cells, GPenv+ cells, or 293 cells. Detailed protocols for producing such replication-defective recombinant viruses can be found for instance in WO95/14785, WO96/22378, U.S. Pat. No. 5,882,877, U.S. Pat. No. 6,013,516, U.S. Pat. No. 4,861,719, U.S. Pat. No. 5,278,056 and WO94/19478, which are all hereby incorporated by reference.

[00136] In another embodiment, the invention provides for a recombinant host cell comprising a recombinant carbohydrate transporter gene (e.g., GLUT2 or SGLT1) or a carbohydrate metabolic enzyme gene (e.g., SI, MGAM, or LCT), or a recombinant vector as described herein. Suitable host cells include, without limitation, prokaryotic cells (such as bacteria) and eukaryotic cells (such as yeast cells, mammalian cells, insect cells, or plant cells). Specific examples include E. coli, the yeasts Kluyveromyces or Saccharomyces, mammalian cell lines (e.g., Vero cells, CHO cells, 3T3 cells, or COS cells) as well as primary or established mammalian cell cultures (e.g., produced from fibroblasts, embryonic cells, epithelial cells, nervous cells, or adipocytes). In a further embodiment, the invention provides a method for producing a recombinant host cell expressing a polypeptide of a carbohydrate transporter or carbohydrate metabolic enzyme. The method can entail (a) introducing *in vitro* or *ex vivo* into a competent host cell a recombinant nucleic acid or a vector as described herein, (b) culturing *in vitro* or *ex vivo* the recombinant host cells obtained, and (c) optionally, selecting the cells which express the polypeptide of a carbohydrate transporter or carbohydrate metabolic enzyme. Such

recombinant host cells can be used for the production of carbohydrate transporter or carbohydrate metabolic enzyme polypeptides, as well as for screening of active molecules, as described below. Such cells can also be used as a model system to study autism. These cells can be maintained in suitable culture media, such as HAM, DMEM, or RPMI, in any appropriate culture device (plate, flask, dish, tube, or pouch).

[00137] The practice of aspects of the present invention can employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); DNA Cloning, Volumes I and II (D. N. Glover ed., 1985); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Pat. No: 4,683,195; Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Methods In Enzymology, Vols. 154 and 155 (Wu et al. eds.), Immunochemical Methods In Cell And Molecular Biology (Caner and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). All patents, patent applications and references cited herein are incorporated in their entirety by reference.

[00138] Administration and Dosing

[00139] A carbohydrate transporter molecule (e.g., GLUT2 or SGLT1) or carbohydrate metabolic enzyme molecule (e.g., SI, MGAM, or LCT) can be administered to the subject once (e.g., as a single injection or deposition). Alternatively, a carbohydrate transporter or carbohydrate metabolic enzyme molecule of the invention can be administered once or twice daily to a subject in need thereof for a period of from about two to about twenty-eight days, or

from about seven to about ten days. It can also be administered once or twice daily to a subject for a period of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 times per year, or a combination thereof.

Furthermore, the carbohydrate transporter or carbohydrate metabolic enzyme molecule of the invention can be co-administrated with another therapeutic, such as an anti-depressant, an anti-psychotic, a benzodiazepine drug, or a combination thereof. Where a dosage regimen comprises multiple administrations, the effective amount of the carbohydrate transporter or carbohydrate metabolic enzyme molecule administered to the subject can comprise the total amount of gene product administered over the entire dosage regimen.

[00140] The carbohydrate transporter or carbohydrate metabolic enzyme molecules of the invention can be administered to a subject by any means suitable for delivering the carbohydrate transporter or carbohydrate metabolic enzyme molecules to cells of the subject, such as ileum cell or cecum cells. For example, carbohydrate transporter or carbohydrate metabolic enzyme molecules can be administered by methods suitable to transfect cells. Transfection methods for eukaryotic cells are well known in the art, and include direct injection of the nucleic acid into the nucleus or pronucleus of a cell; electroporation; liposome transfer or transfer mediated by lipophilic materials; receptor mediated nucleic acid delivery, bioballistic or particle acceleration; calcium phosphate precipitation, and transfection mediated by viral vectors.

[00141] The compositions of this invention can be formulated and administered to reduce the symptoms associated with autism or an ASD by any means that produces contact of the active ingredient with the agent's site of action in the body of an animal. They can be administered by any conventional means available for use in conjunction with pharmaceuticals, either as individual therapeutic active ingredients or in a combination of therapeutic active ingredients. They can be administered alone, but are generally administered with a pharmaceutical carrier selected on the basis of the chosen route of administration and standard pharmaceutical practice.

[00142] Pharmaceutical compositions for use in accordance with the invention can be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. The therapeutic compositions of the invention can be formulated for a variety of routes of administration, including systemic and topical or localized administration. Techniques

and formulations generally can be found in Remington's Pharmaceutical Sciences, Meade Publishing Co., Easton, Pa (1985), the entire disclosure of which is herein incorporated by reference. For systemic administration, an injection is useful, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For injection, the therapeutic compositions of the invention can be formulated in liquid solutions, for example in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the therapeutic compositions can be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included. Pharmaceutical compositions of the present invention are characterized as being at least sterile and pyrogen-free. These pharmaceutical formulations include formulations for human and veterinary use.

[00143] Pharmaceutical formulations of the invention can comprise a carbohydrate transporter or carbohydrate metabolic enzyme molecule (e.g., 0.1 to 90% by weight), or a physiologically acceptable salt thereof, mixed with a pharmaceutically-acceptable carrier. The pharmaceutical formulations of the invention can also comprise the carbohydrate transporter or carbohydrate metabolic enzyme molecules of the invention which are encapsulated by liposomes and a pharmaceutically-acceptable carrier. Useful pharmaceutically-acceptable carriers are water, buffered water, normal saline, 0.4% saline, 0.3% glycine, or hyaluronic acid.

[00144] Pharmaceutical compositions of the invention can also comprise conventional pharmaceutical excipients and/or additives. Suitable pharmaceutical excipients include stabilizers, antioxidants, osmolality adjusting agents, buffers, and pH adjusting agents. Suitable additives include physiologically biocompatible buffers (e.g., tromethamine hydrochloride), additions of chelants (such as, for example, DTPA or DTPA-bisamide) or calcium chelate complexes (as for example calcium DTPA, CaNaDTPA-bisamide), or, optionally, additions of calcium or sodium salts (for example, calcium chloride, calcium ascorbate, calcium gluconate or calcium lactate). Pharmaceutical compositions of the invention can be packaged for use in liquid form, or can be lyophilized.

[00145] For solid pharmaceutical compositions of the invention, conventional nontoxic solid pharmaceutically-acceptable carriers can be used; for example, pharmaceutical grades of

mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, or magnesium carbonate.

[00146] Solid formulations can be used for enteral (oral) administration. They can be formulated as, e.g., pills, tablets, powders or capsules. For solid compositions, conventional nontoxic solid carriers can be used which include, e.g., pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, or magnesium carbonate. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 10% to 95% of active ingredient (e.g., peptide). A non-solid formulation can also be used for enteral administration. The carrier can be selected from various oils including those of petroleum, animal, vegetable or synthetic origin, e.g., peanut oil, soybean oil, mineral oil, or sesame oil. Suitable pharmaceutical excipients include e.g., starch, cellulose, talc, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, magnesium stearate, sodium stearate, glycerol monostearate, sodium chloride, dried skim milk, glycerol, propylene glycol, water, ethanol.

[00147] Nucleic acids, peptides, or polypeptides of the invention, when administered orally, can be protected from digestion. This can be accomplished either by complexing the nucleic acid, peptide or polypeptide with a composition to render it resistant to acidic and enzymatic hydrolysis or by packaging the nucleic acid, peptide or polypeptide in an appropriately resistant carrier such as a liposome. Means of protecting compounds from digestion are well known in the art, see, e.g., Fix, Pharm Res. 13: 1760-1764, 1996; Samanen, J. Pharm. Pharmacol. 48: 119-135, 1996; U.S. Pat. No. 5,391,377, describing lipid compositions for oral delivery of therapeutic agents (for example, liposomal delivery). In one embodiment, the carbohydrate transporter molecule (e.g., GLUT2 or SGLT1) or carbohydrate metabolic enzyme molecule (e.g., SI, MGAM, or LCT) can be delivered to the alimentary canal or intestine of the subject via oral administration that is can withstand digestion and degradation.

[00148] For oral administration, the therapeutic compositions can take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinised maize starch, polyvinylpyrrolidone or

hydroxypropyl methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulphate). The tablets can be coated by methods well known in the art. Liquid preparations for oral administration can take the form of, for example, solutions, syrups or suspensions, or they can be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations can be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g., lecithin or acacia); non-aqueous vehicles (e.g., ationd oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations can also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

[00149] Preparations for oral administration can be suitably formulated to give controlled release of the active agent. For buccal administration the therapeutic compositions can take the form of tablets or lozenges formulated in a conventional manner. For administration by inhalation, the compositions for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit can be determined by providing a valve to deliver a metered amount. Capsules and cartridges of e.g., gelatin for use in an inhaler or insufflate or can be formulated containing a powder mix of the therapeutic agents and a suitable powder base such as lactose or starch.

[00150] The therapeutic compositions can be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection can be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions can take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and can contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient can be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

[00151] Suitable enteral administration routes for the present methods include oral, rectal, or intranasal delivery. Suitable parenteral administration routes include intravascular administration (e.g. intravenous bolus injection, intravenous infusion, intra-arterial bolus injection, intra-arterial infusion and catheter instillation into the vasculature); peri- and intra-tissue injection (e.g., peri-tumoral and intra-tumoral injection, intra-retinal injection, or subretinal injection); subcutaneous injection or deposition including subcutaneous infusion (such as by osmotic pumps); direct application to the tissue of interest, for example by a catheter or other placement device (e.g., a retinal pellet or a suppository or an implant comprising a porous, non-porous, or gelatinous material); and inhalation. For example, the carbohydrate transporter or carbohydrate metabolic enzyme molecules of the invention can be administered by injection, infusion, or oral delivery.

[00152] In addition to the formulations described previously, the therapeutic compositions can also be formulated as a depot preparation. Such long acting formulations can be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. For example, the therapeutic compositions can be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

[00153] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents can be used to facilitate permeation. Transmucosal administration can be through nasal sprays or using suppositories. For topical administration, the compositions of the invention are formulated into ointments, salves, gels, or creams as generally known in the art. A wash solution can be used locally to treat an injury or inflammation to accelerate healing. For oral administration, the therapeutic compositions are formulated into conventional oral administration forms such as capsules, tablets, and tonics.

[00154] A composition of the present invention can also be formulated as a sustained and/or timed release formulation. Such sustained and/or timed release formulations can be made by

sustained release means or delivery devices that are well known to those of ordinary skill in the art, such as those described in U.S. Pat. Nos.: 3,845,770; 3,916,899; 3,536,809; 3,598,123; 4,008,719; 4,710,384; 5,674,533; 5,059,595; 5,591,767; 5,120,548; 5,073,543; 5,639,476; 5,354,556; and 5,733,566, the disclosures of which are each incorporated herein by reference.

The pharmaceutical compositions of the present invention can be used to provide slow or sustained release of one or more of the active ingredients using, for example, hydropropylmethyl cellulose, other polymer matrices, gels, permeable membranes, osmotic systems, multilayer coatings, microparticles, liposomes, microspheres, or the like, or a combination thereof to provide the desired release profile in varying proportions. Suitable sustained release formulations known to those of ordinary skill in the art, including those described herein, can be readily selected for use with the pharmaceutical compositions of the invention. Single unit dosage forms suitable for oral administration, such as, but not limited to, tablets, capsules, gel-caps, caplets, or powders, that are adapted for sustained release are encompassed by the present invention.

[00155] In the present methods, the carbohydrate transporter or carbohydrate metabolic enzyme molecules can be administered to the subject either as RNA, in conjunction with a delivery reagent, or as a nucleic acid (e.g., a recombinant plasmid or viral vector) comprising sequences which expresses the gene product. Suitable delivery reagents for administration of the carbohydrate transporter or carbohydrate metabolic enzyme molecules include the Mirus Transit TKO lipophilic reagent; lipofectin; lipofectamine; cellfectin; or polycations (e.g., polylysine), or liposomes.

[00156] The dosage administered can be a therapeutically effective amount of the composition sufficient to result in amelioration of symptoms of autism or an autism spectrum disorder in a subject, and can vary depending upon known factors such as the pharmacodynamic characteristics of the active ingredient and its mode and route of administration; age, sex, health and weight of the recipient; nature and extent of symptoms; kind of concurrent treatment, frequency of treatment and the effect desired. For example, an effective enzyme unit of amount of SI, MGAM, and/or LCT can be administered to a subject in need thereof. The enzyme unit (U) is a unit for the amount of a particular enzyme. One U is defined as the amount of the enzyme that catalyzes the conversion of 1 micro mole of substrate per minute. In one

embodiment, the therapeutically effective amount of the administered carbohydrate enzyme (e.g., SI, MGAM, or LCT) is at least about 1 U, at least about 10 U, at least about 20 U, at least about 25 U, at least about 50 U, at least about 100 U, at least about 150 U, at least about 200 U, at least about 250 U, at least about 300 U, at least about 350 U, at least about 400 U, at least about 450 U, at least about 500 U, at least about 550 U, at least about 600 U, at least about 650 U, at least about 700 U, at least about 750 U, at least about 800 U, at least about 850 U, at least about 900 U, at least about 950 U, at least about 1000 U, at least about 1250 U, at least about 1500 U, at least about 1750 U, at least about 2000 U, at least about 2250 U, at least about 2500 U, at least about 2750 U, at least about 3000 U, at least about 3250 U, at least about 3500 U, at least about 4000 U, at least about 4500 U, at least about 5000 U, at least about 5500 U, at least about 6000 U, at least about 6500 U, at least about 7000 U, at least about 7500 U, at least about 8000 U, at least about 8500 U, at least about 9000 U, at least about 9250 U, at least about 9500 U, or at least about 10,000 U.

[00157] In some embodiments, the effective amount of the administered carbohydrate transporter molecule (e.g., GLUT2 or SGLT1) is at least about 0.01 µg/kg body weight, at least about 0.025 µg/kg body weight, at least about 0.05 µg/kg body weight, at least about 0.075 µg/kg body weight, at least about 0.1 µg/kg body weight, at least about 0.25 µg/kg body weight, at least about 0.5 µg/kg body weight, at least about 0.75 µg/kg body weight, at least about 1 µg/kg body weight, at least about 5 µg/kg body weight, at least about 10 µg/kg body weight, at least about 25 µg/kg body weight, at least about 50 µg/kg body weight, at least about 75 µg/kg body weight, at least about 100 µg/kg body weight, at least about 150 µg/kg body weight, at least about 200 µg/kg body weight, at least about 250 µg/kg body weight, at least about 300 µg/kg body weight, at least about 350 µg/kg body weight, at least about 400 µg/kg body weight, at least about 450 µg/kg body weight, at least about 500 µg/kg body weight, at least about 550 µg/kg body weight, at least about 600 µg/kg body weight, at least about 650 µg/kg body weight, at least about 700 µg/kg body weight, at least about 750 µg/kg body weight, at least about 800 µg/kg body weight, at least about 850 µg/kg body weight, at least about 900 µg/kg body weight, at least about 950 µg/kg body weight, or at least about 1000 µg/kg body weight.

[00158] Toxicity and therapeutic efficacy of therapeutic compositions of the present invention can be determined by standard pharmaceutical procedures in cell cultures or

experimental animals, e.g., for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Therapeutic agents that exhibit large therapeutic indices are useful. Therapeutic compositions that exhibit some toxic side effects can be used.

[00159] A therapeutically effective dose of carbohydrate transporter or carbohydrate metabolic enzyme molecules can depend upon a number of factors known to those of ordinary skill in the art. The dose(s) of the carbohydrate transporter or carbohydrate metabolic enzyme molecules can vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the carbohydrate transporter or carbohydrate metabolic enzyme molecules to have upon the nucleic acid or polypeptide of the invention. These amounts can be readily determined by a skilled artisan.

[00160] Pharmaceutical Composition and Therapy

[00161] The invention provides methods for treating or preventing autism or an autism spectrum disorder in a subject. In one embodiment, the method can comprise administering to the subject a functional (e.g., wild-type) carbohydrate transporter molecule (e.g., GLUT2 or SGLT1) or carbohydrate metabolic enzyme molecule (e.g., SI, MGAM, or LCT), which can be a polypeptide or a nucleic acid.

[00162] Various approaches can be carried out to restore the carbohydrate transporter or carbohydrate metabolic enzyme activity or function in a subject, such as those carrying an altered gene locus comprising a carbohydrate transporter gene (e.g., GLUT2 or SGLT1) or a carbohydrate metabolic enzyme gene (e.g., SI, MGAM, or LCT). Supplying wild-type function of the carbohydrate transporter or carbohydrate metabolic enzyme to such subjects can suppress phenotypic expression of autism or an autism spectrum disorders in a pathological cell or organism. Increasing carbohydrate transporter or carbohydrate metabolic enzyme activity can be accomplished through gene or protein therapy as discussed later herein.

[00163] A nucleic acid encoding a carbohydrate transporter or carbohydrate metabolic enzyme or a functional part thereof can be introduced into the cells of a subject in one embodiment of the invention. The wild-type carbohydrate transporter gene or carbohydrate metabolic enzyme gene (or a functional part thereof) can also be introduced into the cells of the subject in need thereof using a vector as described herein. The vector can be a viral vector or a plasmid. The gene can also be introduced as naked DNA. The gene can be provided so as to integrate into the genome of the recipient host cells, or to remain extra-chromosomal. Integration can occur randomly or at precisely defined sites, such as through homologous recombination. For example, a functional copy of the carbohydrate transporter gene or a carbohydrate metabolic enzyme gene can be inserted in replacement of an altered version in a cell, through homologous recombination. Further techniques include gene gun, liposome-mediated transfection, or cationic lipid-mediated transfection. Gene therapy can be accomplished by direct gene injection, or by administering *ex vivo* prepared genetically modified cells expressing a functional polypeptide.

[00164] Gene Therapy and Protein Replacement Methods

[00165] Delivery of nucleic acids into viable cells can be effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., lentivirus, adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). Non-limiting techniques suitable for the transfer of nucleic acid into mammalian cells *in vitro* include the use of liposomes, electroporation, microinjection, cell fusion, DEAE-dextran, and the calcium phosphate precipitation method (see, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp. 25-20 (1998)). Introduction of a nucleic acid or a gene encoding a polypeptide of the invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of therapeutic compositions of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes.

[00166] Nucleic acids can be inserted into vectors and used as gene therapy vectors. A number of viruses have been used as gene transfer vectors, including papovaviruses, e.g., SV40

(Madzak et al., 1992), adenovirus (Berkner, 1992; Berkner et al., 1988; Gorziglia and Kapikian, 1992; Quantin et al., 1992; Rosenfeld et al., 1992; Wilkinson et al., 1992; Stratford-Perricaudet et al., 1990), vaccinia virus (Moss, 1992), adeno-associated virus (Muzyczka, 1992; Ohi et al., 1990), herpesviruses including HSV and EBV (Margolskee, 1992; Johnson et al., 1992; Fink et al., 1992; Breakfield and Geller, 1987; Freese et al., 1990), and retroviruses of avian (Biandyopadhyay and Temin, 1984; Petropoulos et al., 1992), murine (Miller, 1992; Miller et al., 1985; Sorge et al., 1984; Mann and Baltimore, 1985; Miller et al., 1988), and human origin (Shimada et al., 1991; Helseth et al., 1990; Page et al., 1990; Buchschacher and Panganiban, 1992). Non-limiting examples of *in vivo* gene transfer techniques include transfection with viral (typically retroviral) vectors (see U.S. Pat. No. 5,252,479, which is incorporated by reference in its entirety) and viral coat protein-liposome mediated transfection (Dzau et al., Trends in Biotechnology 11:205-210 (1993), incorporated entirely by reference). For example, naked DNA vaccines are generally known in the art; see Brower, Nature Biotechnology, 16:1304-1305 (1998), which is incorporated by reference in its entirety. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see, e.g., U.S. Pat. No. 5,328,470) or by stereotactic injection (see, e.g., Chen, et al., 1994. Proc. Natl. Acad. Sci. USA 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

[00167] For reviews of gene therapy protocols and methods see Anderson et al., Science 256:808-813 (1992); U.S. Pat. Nos. 5,252,479, 5,747,469, 6,017,524, 6,143,290, 6,410,010 6,511,847; and U.S. Application Publication Nos. 2002/0077313 and 2002/00069, which are all hereby incorporated by reference in their entireties. For additional reviews of gene therapy technology, see Friedmann, Science, 244:1275-1281 (1989); Verma, Scientific American: 68-84 (1990); Miller, Nature, 357: 455-460 (1992); Kikuchi et al., J Dermatol Sci. 2008 May;50(2):87-98; Isaka et al., Expert Opin Drug Deliv. 2007 Sep;4(5):561-71; Jager et al., Curr Gene Ther. 2007 Aug;7(4):272-83; Waehler et al., Nat Rev Genet. 2007 Aug;8(8):573-87; Jensen et al., Ann Med. 2007;39(2):108-15; Herweijer et al., Gene Ther. 2007 Jan;14(2):99-107; Eliyahu et al.,

Molecules, 2005 Jan 31;10(1):34-64; and Altaras et al., Adv Biochem Eng Biotechnol. 2005;99:193-260, all of which are hereby incorporated by reference in their entireties.

[00168] Protein replacement therapy can increase the amount of protein by exogenously introducing wild-type or biologically functional protein by way of infusion. A replacement polypeptide can be synthesized according to known chemical techniques or may be produced and purified via known molecular biological techniques. Protein replacement therapy has been developed for various disorders. For example, a wild-type protein can be purified from a recombinant cellular expression system (e.g., mammalian cells or insect cells-see U.S. Pat. No. 5,580,757 to Desnick et al.; U.S. Pat. Nos. 6,395,884 and 6,458,574 to Selden et al.; U.S. Pat. No. 6,461,609 to Calhoun et al.; U.S. Pat. No. 6,210,666 to Miyamura et al.; U.S. Pat. No. 6,083,725 to Selden et al.; U.S. Pat. No. 6,451,600 to Rasmussen et al.; U.S. Pat. No. 5,236,838 to Rasmussen et al. and U.S. Pat. No. 5,879,680 to Ginns et al.), human placenta, or animal milk (see U.S. Pat. No. 6,188,045 to Reuser et al.), or other sources known in the art. After the infusion, the exogenous protein can be taken up by tissues through non-specific or receptor-mediated mechanism.

[00169] A polypeptide encoded by a carbohydrate transporter gene (e.g., GLUT2 or SGLT1) or a carbohydrate metabolic enzyme gene (for example, SI, MGAM, or LCT) can also be delivered in a controlled release system. For example, the polypeptide may be administered using intravenous infusion, an implantable osmotic pump, a transdermal patch, liposomes, or other modes of administration. In one embodiment, a pump may be used (see is Langer, supra; Sefton, CRC Crit. Ref. Biomed. Eng. 14:201 (1987); Buchwald et al., Surgery 88:507 (1980); Saudek et al., N. Engl. J. Med. 321:574 (1989)). In another embodiment, polymeric materials can be used (see Medical Applications of Controlled Release, Langer and Wise (eds.), CRC Pres., Boca Raton, Fla. (1974); Controlled Drug Bioavailability, Drug Product Design and Performance, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, J. Macromol. Sci. Rev. Macromol. Chem. 23:61 (1983); see also Levy et al., Science 228:190 (1985); During et al., Ann. Neurol. 25:351 (1989); Howard et al., J. Neurosurg. 71:105 (1989)). In yet another embodiment, a controlled release system can be placed in proximity of the therapeutic target thus requiring only a fraction of the systemic dose (see, e.g., Goodson, in

Medical Applications of Controlled Release, supra, vol. 2, pp. 115-138 (1984)). Other controlled release systems are discussed in the review by Langer (Science 249:1527-1533 (1990)).

[00170] These methods described herein are by no means all-inclusive, and further methods to suit the specific application is understood by the ordinary skilled artisan. Moreover, the effective amount of the compositions can be further approximated through analogy to compounds known to exert the desired effect.

[00171] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention.

[00172] All publications and other references mentioned herein are incorporated by reference in their entirety, as if each individual publication or reference were specifically and individually indicated to be incorporated by reference. Publications and references cited herein are not admitted to be prior art.

EXAMPLES

[00173] Examples are provided below to facilitate a more complete understanding of the invention. The following examples illustrate the exemplary modes of making and practicing the invention. However, the scope of the invention is not limited to specific embodiments disclosed in these Examples, which are for purposes of illustration only, since alternative methods can be utilized to obtain similar results.

Example 1: *Identification of carbohydrate transporters and carbohydrate metabolic enzymes as biomarkers in a subset of Autism Spectrum Disorders (ASD)*

[00174] Gastrointestinal disturbances complicate clinical management in some children with autism. Reports of ileo-colonic lymphoid nodular hyperplasia and deficiencies in disaccharidase enzymatic activity led us to survey intestinal gene expression and microflora in children with autism and gastrointestinal disease (AUT-GI) or gastrointestinal disease alone (Control-GI). In AUT-GI subjects, ileal transcripts for the disaccharidases sucrase isomaltase, maltase glucoamylase, and lactase, and the monosaccharide transporters, sodium-dependent glucose co-transporter, and glucose transporter 2 were significantly decreased. Alterations in intestinal carbohydrates as a result of these deficiencies would have a distinct impact on the composition of AUT-GI intestinal microbiota. Bacterial 16S rRNA gene pyrosequencing analysis of biopsy material from ileum and cecum revealed decreased Bacteroidetes, increased Firmicute/Bacteroidete ratios, higher cumulative levels of Firmicutes and Proteobacteria, and increased Betaproteobacteria in AUT-GI as compared with Control-GI biopsies. These results suggest a complex dependence between intestinal gene expression and bacterial community structure that contributes to gastrointestinal dysfunction in AUT-GI children.

[00175] Deficiencies in intestinal disaccharidase and/or glucoamylase activity are reported in over half of autistic children with gastrointestinal disturbances (AUT-GI) (Horvath et al., 1999). To determine whether functional deficits reflect decreased levels of mRNA encoding these enzymes, we examined transcript levels for three primary brush border disaccharidases (sucrase isomaltase [SI], maltase glucoamylase [MGAM], and lactase [LCT]) in ileal biopsies of AUT-GI and Control-GI children by real time PCR. Levels of mRNA for all three enzymes were decreased in AUT-GI: SI (**FIG. 16A**: Mann-Whitney, $p=0.001$), MGAM (**FIG. 16B**: Mann-

Whitney, $p=0.003$) and LCT (**FIG. 16C**: Mann-Whitney, $p=0.032$). Deficiencies in LCT mRNA in AUT-GI children were not attributable to disproportionate adult-type hypolactasia genotypes in the AUT-GI group relative to the Control-GI group (**FIGS. 21A-21E** and **Methods**). Within the ASD-GI group, 86.7% (SI), 80% (MGAM), and 80% (LCT) of children had transcript levels below the 25th percentile of Control-GI children (**Table 5A**). Nearly all (14/15, or 93.3%) AUT-GI children had deficiencies in at least one disaccharidase enzyme; 80% had deficiencies in 2 or more enzymes; and 73.3% had deficiencies in all three enzymes (**Table 5A**). **Tables 5A-C** are summary tables for gene expression and bacterial assays. Increases or decreases in AUT-GI children in both gene expression and bacterial parameters were determined for each individual based on the levels of each parameter in the Control-GI group. The values for a given parameter in the AUT-GI children that exceeded the 75th (arrow pointing up) percentile or were below the 25th percentile (arrow pointing down) for the corresponding parameter in the Control-GI children were scored as an increase or decrease, respectively. Values that were also above the 90th or below the 10th percentiles of Control-GI children are indicated by double arrows.

[00176] Table 5A. Summary tables for gene expression and bacterial assays.

ASD Patient #	SI	MGAM	LCT	SGLT1	GLUT2	CDX2	Villin	# Disaccharidases	# Transporters	Total
1	↕↕	↕↕	↕↕	↕↕	↕↕	↕	n.c.	3/3	2/2	5/5
2	↕↕	↕↕	↕↕	↕↕	↕↕	↕↕	↕↕	3/3	2/2	5/5
3	↕↕	↕↕	↕↕	↕↕	↕↕	↕↕	↕↕	3/3	2/2	5/5
4	↕↕	↕↕	↕↕	↕↕	↕↕	↕↕	n.c.	3/3	2/2	5/5
5	↕↕	↕↕	↕↕	↕↕	↕↕	↕↕	↕↕	3/3	2/2	5/5
6	↕↕	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	1/3	0/2	1/5
7	n.c.	n.c.	↕↕	n.c.	n.c.	↕	↕↕	0/3	0/2	0/5
8	↕↕	↕↕	↕↕	↕↕	↕↕	n.c.	↕	3/3	2/2	5/5
9	↕↕	↕↕	↕↕	↕↕	↕↕	n.c.	↕	3/3	1/2	4/5
10	↕↕	↕↕	↕↕	↕↕	↕↕	n.c.	↕	3/3	2/2	5/5
11	↕↕	↕↕	↕↕	↕↕	↕↕	n.c.	↕	3/3	2/2	5/5
12	↕↕	↕↕	↕↕	↕↕	↕↕	n.c.	n.c.	3/3	2/2	5/5
13	↕↕	↕↕	↕↕	↕↕	↕↕	↕↕	↕↕	3/3	2/2	5/5
14	↕↕	↕↕	n.c.	n.c.	↕↕	n.c.	↕↕	2/3	1/2	3/5
15	n.c.	n.c.	↕↕	n.c.	n.c.	n.c.	n.c.	1/3	0/2	1/5
% below controls	86.7%	80.0%	80.0%	73.3%	73.3%	33.3%	26.7%	Summary All 3 = 73.3% At least 2 = 80% At least 1 = 93.3%	Summary Both = 66.7% At least 1 = 80%	Summary All 5 = 66.7% At least 4 = 73.3% At least 3 = 80% At least 1 = 93.3%

[00177] Table 5B. Summary tables for gene expression and bacterial assays.

ASD Patient #	Bacteroidetes				Firmicutes								Proteobacteria				
	Bacteroidetes RT-Ileum	Bacteroidetes RT-Cecum	Bacteroidetes 454-Ileum	Bacteroidetes 454-Cecum	Firmicutes RT-Ileum	Firmicutes RT-Cecum	Firmicutes 454-Ileum	Firmicutes 454-Cecum	Clostridia 454-Ileum	Clostridia 454-Cecum	Lach. + Rumino. 454-Ileum	Lach. + Rumino. 454-Cecum	Proteobacteria 454-Ileum	Proteobacteria 454-Cecum	Beta-Proteobacteria 454-Ileum	Beta-Proteobacteria 454-Cecum	Sutrellia-RT & 454 (Ileum & Cecum)
1	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞
2	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	☞☞☞☞	☞☞☞☞	n.c.
3	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	☞☞☞☞	☞☞☞☞	n.c.
4	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	☞☞☞☞	☞☞☞☞	n.c.
5	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	n.c.	☞☞☞☞	n.c.	☞☞☞☞	n.c.	☞☞☞☞	☞☞☞☞	n.c.	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞
6	☞☞☞☞	n.c.	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	☞☞☞☞	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.
7	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞
8	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.
9	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	☞☞☞☞	☞☞☞☞	n.c.
10	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞
11	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	☞☞☞☞	n.c.	n.c.	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞
12	☞☞☞☞	☞☞☞☞	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	☞☞☞☞	☞☞☞☞
13	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	n.c.	☞☞☞☞	n.c.
14	☞☞☞☞	n.c.	☞☞☞☞	n.c.	n.c.	n.c.	☞☞☞☞	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	n.c.	n.c.	n.c.	n.c.
15	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	n.c.	☞☞☞☞	n.c.
% below or above controls	100%	86.7%	80.0%	73.3%	46.7%	53.3%	60.0%	46.7%	66.7%	53.3%	60%	53.3%	26.7%	40.0%	53.3%	66.7%	46.7%

[00178] Table 5C. Summary tables for gene expression and bacterial assays.

ASD Patient #	Firm./Bacteroid. Ratio-RT Ileum	Firm./Bacteroid. Ratio-RT Cecum	Firm./Bacteroid. Ratio-454 Ileum	Firm./Bacteroid. Ratio-454 Cecum	Clostridiales/Bacteroidales Ratio-454 Ileum	Clostridiales/Bacteroidales Ratio-454 Cecum	Firm. + Proteobac. Ratio-454 Ileum	Firm. + Proteobac. Ratio-454 Cecum
1	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞
2	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞
3	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞
4	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞
5	☞☞☞☞	n.c.	☞☞☞☞	n.c.	☞☞☞☞	n.c.	☞☞☞☞	n.c.
6	☞☞☞☞	☞☞☞☞	n.c.	n.c.	☞☞☞☞	n.c.	n.c.	☞☞☞☞
7	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞
8	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞
9	☞☞☞☞	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞
10	☞☞☞☞	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞
11	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞
12	☞☞☞☞	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
13	☞☞☞☞	n.c.	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞
14	☞☞☞☞	n.c.	☞☞☞☞	n.c.	☞☞☞☞	n.c.	☞☞☞☞	n.c.
15	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞	☞☞☞☞
% above controls	100%	60%	73.3%	66.7%	80.0%	66.7%	80%	73.3%

[00179] Two hexose transporters, SGLT1 and GLUT2, mediate transport of monosaccharides in the intestine. SGLT1, located on the luminal membrane of enterocytes, is responsible for the active transport of glucose and galactose from the intestinal lumen into enterocytes. GLUT2 transports glucose, galactose and fructose across the basolateral membrane into the circulation and may also translocate to the apical membrane (Kellett et al., 2008). Real-time PCR revealed a decrease in SGLT1 mRNA (**FIG. 16D**: Mann-Whitney, $p=0.008$) and GLUT2 mRNA (**FIG. 16E**: Mann-Whitney, $p=0.010$) in AUT-GI children. For SGLT1, 73.3% of AUT-GI children had transcript levels below the 25th percentile of Control-GI children; 73.3% of AUT-GI children had GLUT2 transcript levels below the 25th percentile of Control-GI children (**Table 5A**). Deficiencies were found in at least one hexose transporter in 80% of AUT-GI children; 66.7% had deficiencies in both transporters. In total, 66.7% of AUT-GI children had mRNA deficiencies in all 5 molecules involved in carbohydrate digestion and transport (**Table 5A**). Expression levels were correlated (Bonferroni-adjusted Spearman rank order correlations) in the AUT-GI group for all gene combinations except LCT and GLUT2, for which only a trend was observed. In the Control-GI group, significance was limited to correlations of SI- MGAM, MGAM-SGLT1, and LCT-SGLT1 (**Table 2**).

[00180] **Table 2.** Spearman correlations between ileal gene expression and bacterial abundance variables. Spearman correlations are shown for the AUT-GI group alone (AUT) and the Control-GI group alone (Control).

	Group	SI	MGAM	LCT	SGLT1	GLUT2	Villin	CDX2
SI	AUT	1	0.89***	0.59*	0.88**	0.76**	0.24	0.59*
	Control	1	0.93*	0.54	0.68†	0.75†	0.57	0.68†
MGAM	AUT	-	1	0.56*	0.86**	0.75**	0.31	0.63*
	Control	-	1	0.75†	0.82*	0.64	0.71†	0.82*
LCT	AUT	-	-	1	0.62*	0.52†	0.58*	0.65*
	Control	-	-	1	0.86*	0.57	0.82*	0.86*
SGLT1	AUT	-	-	-	1	0.71**	0.34	0.54*
	Control	-	-	-	1	0.64	0.96*	1.00*
GLUT2	AUT	-	-	-	-	1	0.51†	0.69**
	Control	-	-	-	-	1	0.54	0.64
Villin	AUT	-	-	-	-	-	1	0.60*
	Control	-	-	-	-	-	1	0.96*
CDX2	AUT	-	-	-	-	-	-	1
	Control	-	-	-	-	-	-	1
Bacteroidetes Ileum	AUT	0.33	0.10	0.31	0.52† ^a	0.07	0.02	-0.01
	Control	-0.29	-0.29	-0.32	-0.18	-0.75† ^a	0.00	-0.18
Bacteroidetes Cecum	AUT	0.18	0.06	0.23	0.33	0.05	0.12	0.10
	Control	-0.93*	-1.00*	-0.75†	-0.82*	-0.64	-0.71†	-0.82*
Firmicutes Ileum	AUT	-0.61* ^a	-0.55* ^a	-0.00	0.12	0.23	0.64*	0.48† ^a
	Control	0.43	0.36	0.18	0.32	0.61	0.14	0.32
Firmicutes Cecum	AUT	-0.06	0.05	-0.05	0.04	0.15	0.58*	0.14
	Control	0.86*	0.86*	0.68†	0.89*	0.86*	0.79†	0.89*
Firm./Bacteroid. Ileum	AUT	-0.72** ^a	-0.65* ^a	-0.61* ^a	-0.65* ^a	-0.55* ^a	0.36	-0.58* ^a
	Control	0.43	0.36	0.18	0.32	0.61	0.14	0.32
Firm./Bacteroid. Cecum	AUT	-0.51† ^a	-0.08	-0.11	-0.23	0.00	0.42	0.06
	Control	0.86*	0.86*	0.68†	0.89*	0.86*	0.79†	0.89*
Betaproteo. Ileum	AUT	-0.63*	-0.60*	-0.56*	-0.44†	-0.60*	-0.45†	-0.70**
	Control	-0.75†	-0.82*	-0.54	-0.61	-0.57	-0.39	-0.61
Betaproteo. Cecum	AUT	-0.56*	-0.59*	-0.64*	-0.51†	-0.61*	-0.61*	-0.85**
	Control	-0.43	-0.43	0.14	-0.00	0.14	0.14	-0.00

* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$, † = $p < 0.1$ (trend)

a = values obtained from bacteria-specific real-time PCR

[00181] To determine whether reductions in disaccharidase and transporter transcript levels reflected loss of or damage to intestinal epithelial cells, we measured mRNA levels associated with a tissue-specific marker restricted to these cells, villin (Khurana and George, 2008). Ileal villin mRNA levels were not decreased in AUT-GI children (Mann-Whitney, $p=0.307$) (**FIG. 16F**). Normalization of SI, MGAM, LCT, SGLT1 and GLUT2 to villin mRNA did not correct AUT-GI deficits in gene expression for these transcripts (**FIGS. 22A-22E**).

[00182] CDX2, a member of the caudal-related homeobox transcription factor family, regulates expression of SI, LCT, GLUT2, SGLT1 and villin (Suh and Traber, 1996; Troelsen et al., 1997; Uesaka et al., 2004; Balakrishnan et al., 2008; and Yamamichi et al., 2009). Real-time PCR experiments demonstrated lower levels of CDX2 mRNA in some AUT-GI subjects as compared with controls, but group differences were not significant (**FIG. 16G**: Mann-Whitney, $p=0.192$). Only 33.3% of AUT-GI patients had CDX2 mRNA levels below the 25th percentile of

the Control-GI group (**FIG. 23A**). However, 86.7% of AUT-GI children had CDX2 levels below the 50th percentile of Control-GI children. Only one AUT-GI child (patient 7) had CDX2 levels above the 75th percentile of Control-GI children. This child was the only subject who did not show signs of deficiencies in any disaccharidases or transporters (**Table 5A**). In the AUT-GI group, expression of CDX2 was correlated with that of SI, MGAM, LCT, SGLT1, GLUT2, and villin (Bonferroni-adjusted Spearman rank order correlations; **Table 2**). Among Control-GI subjects, the expression of CDX2 was correlated only with that of MGAM, LCT, SGLT1, and villin (**Table 2**).

[00183] To determine whether deficient carbohydrate digestion and absorption influenced the composition of intestinal microflora, ileal and cecal biopsies from AUT-GI and Control-GI children were analyzed by bacterial 16S rRNA gene pyrosequencing (See also **Methods** and **FIGS. 23A-23D**). Bacteroidetes and Firmicutes were the most prevalent taxa present in the ileal and cecal tissues of AUT-GI children, with the exception of the ileal samples of patients 2, 15, and 19 and cecal samples of patient 15, wherein levels of Proteobacteria exceeded those of Firmicutes and/or Bacteroidetes (**FIGS. 17A-B** and **FIGS. 24A-B**). Other phyla identified at lower levels included Verrucomicrobia, Actinobacteria, Fusobacteria, Lentisphaerae, TM7, and Cyanobacteria, as well as unclassified bacterial sequences (**FIGS. 17A-B** and **FIGS. 24A-24D**). The abundance of Bacteroidetes was lower in AUT-GI ileal (**FIG. 17C**: Mann-Whitney, $p=0.012$) and cecal samples (**FIG. 17D**: Mann-Whitney, $p=0.008$) as compared with the abundance of Bacteroidetes in Control-GI samples. Real-time PCR using Bacteroidete-specific primers confirmed decreases in Bacteroidetes in AUT-GI ilea (**FIG. 17E**: Mann-Whitney, $p=0.003$) and ceca (**FIG. 17F**: Mann-Whitney, $p=0.022$), with levels below the 25th percentile of Control-GI children in 100% of AUT-GI ilea and 86.7% of AUT-GI ceca (**Table 5B**). Family-level analysis of Bacteroidete diversity from pyrosequencing reads suggested that losses among members of the family *Bacteroidaceae* in AUT-GI patient samples contributed substantially to overall decreases in Bacteroidete levels in ilea (**FIG. 17G**) and ceca (**FIG. 17H**). OTU (Operational Taxonomic Unit) analysis of Bacteroidete sequences suggested that deficiencies in Bacteroidete sequences in AUT-GI subjects were attributable to cumulative losses of 12 predominant phylotypes of Bacteroidetes, rather than loss of any one specific phylotype (**FIGS. 25A-25E** and **Methods**).

[00184] Analysis of pyrosequencing reads revealed an increase in Firmicute/Bacteroidete ratios in AUT-GI ilea (**FIG. 18A**: Mann-Whitney, $p=0.026$) and ceca (**FIG. 18B**: Mann-Whitney, $p=0.032$). An increase was also observed at the order level for Clostridiales/Bacteroidales ratios in ilea (**FIG. 26A**: Mann-Whitney, $p=0.012$) and ceca (**FIG. 26B**: Mann-Whitney, $p=0.032$). Real-time PCR using Firmicute- and Bacteroidete-specific primers confirmed these increases in Firmicute/Bacteroidete ratios in AUT-GI ilea (**FIG. 26C**: Mann-Whitney, $p=0.0006$) and ceca (**FIG. 26D**: Mann-Whitney, $p=0.022$). Firmicute/Bacteroidete ratios were above the 75th percentile of Control-GI values in 100% of AUT-GI ilea and 60% of AUT-GI ceca (**Table 5C**). Order-level analysis of pyrosequencing reads indicated trends toward increased Clostridiales in AUT-GI ilea (**FIG. 27E**: Mann-Whitney, $p=0.072$) and ceca (**FIG. 27F**: Mann-Whitney, $p=0.098$). Family-level analysis revealed that increased Clostridiales levels in AUT-GI patient samples were largely attributable to increases in members of the families *Lachnospiraceae* and *Ruminococcaceae* (**FIGS. 18C-18F**). Cumulative levels of *Lachnospiraceae* and *Ruminococcaceae* above the 75th percentile of the corresponding levels in Control-GI samples were found in 60% of AUT-GI ileal and 53.3% of AUT-GI cecal samples (**FIGS. 18E-18F** and **Table 5B**). Genus-level analysis indicated that members of the genus *Faecalibacterium* within the family *Ruminococcaceae* contributed to the overall trend toward increased Clostridia levels (**FIGS. 28A-B**). Within *Lachnospiraceae*, members of the genus *Lachnospiraceae Incertae Sedis*, *Unclassified Lachnospiraceae*, and to a lesser extent *Bryantella* (cecum only) contributed to the overall trend toward increased Clostridia in ASD-GI patients (**FIGS. 28A-B**).

[00185] The cumulative level of Firmicutes and Proteobacteria was higher in AUT-GI group in both ileal (**FIG. 18G**: Mann-Whitney, $p = 0.015$) and cecal samples (**FIG. 18H**: Mann-Whitney, $p = 0.007$) (**FIGS. 18I-J**); however, neither Firmicute nor Proteobacteria levels showed significant differences on their own (**FIGS. 19A-19B** and **FIGS. 27A-27D**). Levels of Betaproteobacteria tended to be higher in the ilea of AUT-GI patients (**FIG. 19C**: Mann-Whitney, $p = 0.072$); significantly higher levels of Betaproteobacteria were found in AUT-GI ceca (**FIG. 19D**: Mann-Whitney, $p = 0.038$). Levels of Betaproteobacteria were above the 75th percentile of Control-GI children in 53.3% of AUT-GI ilea and 66.7% of AUT-GI ceca (**Table 5B**). Family-level analysis revealed that members of the families *Alcaligenaceae* and *Incetiae Sedis 5* (patient 2 only) contributed substantively to the observed increases in Beta-

Proteobacteria in ilea (**FIG. 19E**) and ceca (**FIG. 19F**). *Alcaligenaceae* sequences were detected in 46.7% of AUT-GI children and none of the Control-GI children. Overly elevated levels of Proteobacteria in AUT-GI ilea and ceca reflected increased Alpha- (families *Methylobacteriaceae* and *Unclassified Rhizobiales*) and Betaproteobacteria (family *Incertae Sedis 5*) for patient #2 and increased Gammaproteobacteria (family *Enterobacteriaceae*) for patients #8 and #15 (**FIGS. 19E-19F**). Levels of Alpha-, Delta-, Gamma-, and Epsilonproteobacteria were not significantly different between AUT-GI and Control-GI samples.

[00186] The relationships between ileal and cecal microflora and levels of disaccharidases, transporters, villin, and CDX2 were assessed (**Table 2**). In the AUT-GI group, significant inverse Spearman correlations were found for ileal Firmicutes vs. SI and MGAM; the ileal Firmicute/Bacteroidete ratio vs. SI, MGAM, LCT, SGLT1, GLUT2, and CDX2; and ileal and cecal Betaproteobacteria vs. SI, MGAM, LCT, GLUT2, and CDX2. In the Control-GI group significant inverse Spearman correlations were found for cecal Bacteroidetes vs. SI, MGAM, SGLT1, and CDX2; as well as ileal Betaproteobacteria vs. MGAM. Positive Spearman correlations were also found in the Control-GI group: cecal Firmicutes vs. SI, MGAM, SGLT1, GLUT2, and CDX2; and cecal Firmicute/Bacteroidete ratio vs. SI, MGAM, SGLT1, GLUT2, and CDX2 (**Table 2**). These results suggest a complex dependence between carbohydrate metabolizing and transporting genes and the composition of the intestinal microbiome (See **FIG. 20A-20C**).

[00187] **Discussion**

[00188] ASD are brain disorders defined using behavioral criteria; however, many affected individuals also have substantial GI morbidity. A previous report on GI disturbances in ASD found low activities of at least one disaccharidase or glucoamylase in duodenum in 58% of children examined (21 of 36) (Horvath et al., 1999). In our study, 93.3% of AUT-GI children had decreased mRNA levels for at least one of the three disaccharidases (SI, MGAM, or LCT). In addition, we found decreased levels of mRNA for two important hexose transporters, SGLT1 and GLUT2. Transcripts for the enterocyte marker, villin, were not deficient in AUT-GI ilea; thus these deficiencies are unlikely to be due to a general loss of enterocytes. However, defects in enterocyte maturational or migration along the crypt-villus axis may compromise

transcriptional regulation of ileal enzymes and transporters (Hodin et al., 1995). The expression of CDX2, a master transcriptional regulator in the intestine, was correlated with expression of disaccharidases and transporters in AUT-GI children. Therefore, CDX2 could play a role in the observed expression deficits for these genes. Whatever the mechanism, reduced capacity for digestion and transport of carbohydrates may have profound effects. Within the intestine malabsorbed monosaccharides can lead to osmotic diarrhea; non-absorbed sugars may also serve as substrates for intestinal microflora that produce fatty acids and gases (methane, hydrogen, and carbon dioxide), promoting additional GI symptoms such as bloating and flatulence. The deficiency of even a single gene in this important pathway can result in severe GI disease, as occurs with Glucose-galactose malabsorption syndrome caused by SGLT1 deficiency, Fanconi-Bickel syndrome resulting from GLUT2 mutations, sucrase-isomaltase deficiency, and congenital lactase deficiency. A potential link between neurological dysfunction and malabsorption in childhood autism has been previously hypothesized (Goodwin et al., 1971). Extra-intestinal manifestations of GI disease, including neurologic presentation, are described in patients with inflammatory bowel disease and celiac disease (Bushara 2005; Lossos et al., 1995; Gupta et al., 2005). An association between language regression and GI symptoms has been reported in ASD, supporting a link between GI disease and behavioral outcomes (Valicenti-McDermott et al., 2008). Outside the intestine, the major role of dietary carbohydrates is to serve as the primary source of cellular energy throughout the body. Following digestion, nearly all ingested carbohydrates are converted to glucose, which serves a central role in metabolism and cellular homeostasis. The brain, of all organs, is quantitatively the most energy-demanding, accounting for 50% of total body glucose utilization (Owen et al., 1967). Abnormalities in glucose metabolism and homeostasis have been documented in ASD: recovery of blood glucose levels was delayed in ASD children following insulin-induced hypoglycemia (Maher et al., 1975). Brain glucose metabolism is decreased in ASD by positron emission tomography (Toal et al., 2005; Haznedar et al., 2000; Haznedar et al., 2006). A reduced capacity to digest carbohydrates and absorb glucose due to deficient expression of disaccharidases and hexose transporters could provide a mechanistic explanation for these previous observations in ASD.

[00189] Changes in diet can influence composition of intestinal microflora; thus, we reasoned that carbohydrate malabsorption might have similar effects in AUT-GI subjects. 16S rRNA pyrosequencing revealed multicomponent dysbiosis in AUT-GI children including

decreased levels of Bacteroidetes, an increase in the Firmicute/Bacteroidete ratio, increased cumulative levels of Firmicutes and Proteobacteria, and an increase in the class Betaproteobacteria. Bacteroidetes are implicated in mediating maturational and functional processes in the intestine as well as immune modulation. Monocolonization of mice with the prototypic gut symbiont, *Bacteroides thetaiotaomicron*, reverses the maturational defect in ileal epithelial glycan fucosylation that occurs in germ-free mice and regulates the expression of host genes, including SGLT-1 and LCT, that participate in key intestinal functions (i.e., nutrient absorption, metabolism, epithelial barrier function, and intestinal maturation) (Hooper et al., 2001).

[00190] A direct role for Bacteroidetes in carbohydrate metabolism is also evident. *B. thetaiotaomicron* encodes in its genome an expansive number of genes dedicated to polysaccharide acquisition and processing, including 236 glycoside hydrolases and 15 polysaccharide lyases (Flint et al., 2008). Thus, deficient digestion and absorption of di- and monosaccharides in the small intestine may alter the milieu of growth substrates in the ileum and cecum. As such, the growth advantages that Bacteroidetes enjoy in the healthy intestine as a result of their expansive capacity to thrive on polysaccharides may be compromised in AUT-GI children as bacterial species better suited for growth on undigested and unabsorbed carbohydrates flourish. Furthermore, polysaccharide A (PSA), a single molecule from another Bacteroidete member, *Bacteroides fragilis*, protects germ-free mice from *Helicobacter hepaticus*- and chemically-induced colitis by correcting defects in T-cell development, suppressing production of IL-17 and TNF-alpha, and inducing IL-10 (Mazmanian et al., 2008). These reports highlight the multiple roles Bacteroidete members play in the maintenance of intestinal homeostasis, including maturation of epithelium; regulation of intestinal gene expression, including carbohydrate metabolizing genes and transporters; metabolism of polysaccharides in the colon; and development of a competent immune system. Thus, deficient levels of Bacteroidetes in the muco-epithelium of AUT-GI children may directly compromise carbohydrate metabolism and trigger inflammatory pathways.

[00191] Mice that are genetically obese (ob/ob) have 50% fewer Bacteroidetes. A lower abundance of Bacteroidetes is reported in stool samples from obese individuals (Ley et al., 2005; Ley et al., 2006). Using Bacteroidete-specific real-time PCR, we found dramatic decreases in

the ilea (~50% lower abundance) as well as significantly lower levels in the ceca (~25% lower abundance) of AUT-GI compared to Control-GI children. In ob/ob mice, diet-induced obese mice, and in obese humans, the decrease in Bacteroidetes is accompanied by an increase in Firmicutes (Turnbaugh et al., 2008; Ley et al., 2005; Ley et al., 2006). The increased Firmicute/Bacteroidete ratio in obesity is hypothesized to increase the capacity to harvest energy from the diet (Turnbaugh et al., 2006). In our study, the trend toward increased Firmicutes and the significant decrease in Bacteroidetes led to a significant increase in the Firmicute/Bacteroidete ratio in ilea and ceca of AUT-GI compared to Control-GI children. The trend toward increased Firmicutes was largely attributable to Clostridia members; based on pyrosequencing result, members of *Ruminococcaceae* and *Lachnospiraceae* were the major contributors.

[00192] Several members of *Ruminococcaceae* and *Lachnospiraceae* are known butyrate producers and may thus influence short-chain fatty acid (SCFA) levels (Louis et al., 2010). SCFA influence colonic pH and *Bacteroides sp.* are relatively sensitive to acidic pH (Duncan et al., 2009). Three previous reports indicated differences in Clostridia species in stool samples from ASD-GI as compared to control children, including greater abundance of *Clostridium* clusters I, II, XI and *C. bolteae* (Finegold et al, 2002; Song et al., 2004; Parracho et al., 2005). Although only a trend was observed for increased Firmicutes in AUT-GI children, the cumulative levels of Firmicutes and Proteobacteria were significantly higher. Three AUT-GI patients had extremely high levels of Alpha- and Beta-, or Gammaproteobacteria. In addition, the AUT-GI group had elevated levels of Betaproteobacteria compared to the Control-GI group, reflecting the presence of *Alcaligenaceae* members in the ilea and ceca of 46.7% of AUT-GI children. *Alcaligenaceae* sequences were not detected in tissues from Control-GI children.

[00193] **Conclusions:**

[00194] Metabolic interactions between intestinal symbionts and the human host are only beginning to be understood. Increasing evidence shows that gastrointestinal disease and dysbiosis exert system-wide effects on normal host physiology. In this study, we have shown that GI disease in autism has a molecular profile distinct from GI disease in normally-developing children. AUT-GI children have deficiencies in disaccharidase and hexose transporter gene

expression that likely promote malabsorption and multicomponent, compositional dysbiosis. Although the extra-intestinal effects these changes may elicit remain speculative, the identification of specific molecular and microbial signatures that define gastrointestinal pathophysiology in AUT-GI children sets the stage for further research aimed at defining the epidemiology, diagnosis and informed treatment of GI symptoms in autism.

[00195] Materials and Methods:

[00196] *Patient samples.* Patient biopsies were collected as part of a study to assess the frequency of measles virus transcripts in ilea and ceca of children with autistic disorder and gastrointestinal complaints (AUT-GI, n=15) and children with gastrointestinal complaints without brain disorder (Control-GI, n=7). This cohort has been previously described in detail (Hornig et al., 2008). The present study restricted to male, Caucasian children from the original cohort between 3 and 5 years of age to control for confounding effects of gender, race and age on intestinal gene expression and bacterial microbiota. The age at biopsy was similar for AUT-GI and Control-GI subjects (median, in years [interquartile range, IQR]: AUT-GI, 4.5 (1.2); Control-GI, 3.98 (0.9); Mann-Whitney, $p=0.504$] (See Table 3).

[00197] Table 3. Patient information Table.

Patient #	Group	Age (yrs.)	LCT (13910:22018)
215	1	ASD	4.35
478	2	ASD	5.94
513	3	ASD	4.66
530	4	ASD	5.46
554	5	ASD	4.01
562	6	ASD	3.80
566	7	ASD	3.49
581	8	ASD	4.29
589	9	ASD	5.62
648	10	ASD	4.71
678	11	ASD	5.28
686	12	ASD	5.03
688	13	ASD	4.00
733	14	ASD	4.53
800	15	ASD	3.51
667	16	Control	3.98
755	17	Control	5.06
760	18	Control	3.89
796	19	Control	5.48
797	20	Control	3.98
814	21	Control	3.95
842	22	Control	4.12

[00198] *RNA and DNA extraction.* RNA and DNA were extracted sequentially from individual ileal and cecal biopsies (total of 176 biopsies: 88 ileal and 88 cecal biopsies; 4 biopsies per patient per region; 15 AUT-GI patients and 7 Control-GI patients) in TRIzol using standard protocols. RNA and DNA concentrations and integrity were determined using a Nanodrop ND-1000 Spectrophotometer (Nanodrop Technologies, Wilmington, DE) and Bioanalyzer (Agilent Technologies, Foster City, CA) and stored at -80°C .

[00199] *Quantitative Real-Time PCR of human mRNA.* Intron/exon spanning, gene-specific PCR primers and probes for sucrase isomaltase, maltase glucoamylase, lactase, SGLTI, GLUT2, Villin, and CDX2, with GAPDH and Beta-actin as dual housekeeping gene controls were designed for real-time PCR using Primer Express 1.0 software (Applied Biosystems, Foster City, CA). Taqman probes were labeled with the reporter FAM (6-carboxyfluorescein) and the quencher BBQ (Blackberry) (TIB MolBiol). PCR standards for determining copy numbers of target transcripts were generated from amplicons cloned into the vector pGEM-T easy (Promega Corporation, Madison, WI). Linearized plasmids were quantitated by UV spectroscopy and 10-fold serial dilutions (ranging from 5×10^5 to 5×10^0 copies) were created in water containing

yeast tRNA (1 ng/ μ l). Unpooled RNA from individual ileal biopsies were used for real time PCR assays; each individual biopsy was assayed in duplicate. cDNA was synthesized using Taqman reverse transcription reagents (Applied Biosystems) from 2 μ g unpooled RNA per 100 μ l reaction. Each 25- μ l amplification reaction contained 10 μ l template cDNA, 12.5 μ l Taqman Universal PCR Master Mix (Applied Biosystems), 300 nM gene-specific primers and 200 nM gene-specific probe (**Table 2**). The thermal cycling profile using a ABI StepOnePlus Real-time PCR System (Applied Biosystems) consisted of: Stage 1, one cycle at 50° C for 2 min; Stage 2, 1 cycle at 95° C for 10 min; Stage 3, 45 cycles at 95° C for 15 s and 60° C for 1 min (1 min 30 s for LCT). GAPDH and B-actin mRNA were amplified in duplicate reactions by real-time PCR from the same reverse transcription reaction as was performed for the gene of interest. The mean concentration of GAPDH or Beta-actin in each sample was used to control for integrity of input RNA and to normalize values of target gene expression to those of the housekeeping gene expression. The final results shown were expressed as the mean copy number from replicate biopsies per patient, relative to values obtained for GAPDH mRNA. Beta-actin normalization gave similar results to GAPDH normalization for all assays. Due to insufficient or poor quality RNA, only 3 of the 4 biopsies were included for 3 patients (Patient #s 4, 7, 10) and only 2 of the 4 biopsies were included for 1 patient (Patient # 2). Thus, 83 of the original 88 ileal biopsies were used in real-time PCR experiments.

[00200] *Lactase genotyping.* Genomic DNA from AUT-GI (n=15) and Control-GI (n=7) patients was subjected to previously-described PCR-restriction fragment length polymorphism (PCR-RFLP) analysis for the C/T-13910 and G/A-22018 polymorphisms associated with Adult-type Hypolactasia with minor modifications (Buning et al., 2003). Genotyping primers for C/T-13910 and G/A-22018 polymorphisms are as follows: C/T-13910For (5'-GGATGCACTGC TGTGATGAG-3'[SEQ ID NO: 20]), C/T-13910Rev (5'-CCCACTGACCTATCCTCGTG-3'[SEQ ID NO: 21]), G/A-22018For (5'-AACAGGCACGTGGAGGAGTT-3'[SEQ ID NO: 22]), and G/A-22018Rev (5'-CCCACCTCAGCCTCTTGAGT-3'[SEQ ID NO: 23]). Each 50- μ l amplification reaction contained 500 ng genomic DNA, 400 nM forward and reverse primers, and 25 μ l High Fidelity PCR master mix. Thermal cycling consisted of 1 cycle at 94° C for 4 min followed by 40 cycles at 94° C for 1 min, 60° C for 1 min, and 72° C for 1 min. PCR reactions for C/T-13910 were directly digested with the restriction enzyme BsmFI at 65° C for 5 hrs. PCR

reactions for G/A-22018 were resolved on 1% agarose gels followed by gel extraction of the prominent 448bp amplicon. Gel extracted G/A-22018 amplicons were then digested with the restriction enzyme HhaI at 37°C for 5 hrs. Restriction digests of C/T-13910 and G/A-22018 were resolved on 1.5% ethidium-stained agarose gels for genotyping analysis. BsmFI digestion of the C/T-13910 amplicons generates two fragments (351bp and 97bp) for the hypolactasia genotype (C/C), four fragments (351bp, 253bp, 98bp, and 97bp) for the heterozygous genotype (C/T), and three fragments (253bp, 98bp, and 97bp) for the normal homozygous allele (T/T). HhaI digestion of the G/A-22018 amplicons generates two fragments (284bp and 184bp) for the hypolactasia genotype (G/G), three fragments (448bp, 284bp, and 184bp) for the heterozygous genotype (G/A), and a single fragment (448bp) for the normal homozygous allele (A/A).

[00201] *PCR amplification of bacterial 16S rRNA gene and barcoded 454 pyrosequencing of intestinal microbiota.* For DNA samples from 88 ileal biopsies (4 biopsies per patient; 15 AUT-GI patients, 7 Control-GI patients) and 88 cecal biopsies from the same patients, PCR was carried out using bacterial 16S rRNA gene-specific (V2-region), barcoded primers as previously described (Hamady et al., 2008). Composite primers were as follows: (For) 5'-GCCTTGCCAGCCCGCTCAGTCAGAGTTTGATCCTGGCTCAG-3' [SEQ ID NO: 24], (Rev) 5'-GCCTCCCTCGCGCCATCAGNNNNNNNNCATGCTGCCTCCCGTAGGAGT-3' [SEQ ID NO: 25]. Underlined sequences in the Forward and Reverse primers represent the 454 Life Sciences@ primer B and primer A, respectively. Bold sequences in the forward and reverse primers represent the broadly-conserved bacterial primer 27F and 338R, respectively. NNNNNNNN represents the eight-base barcode, which was unique for each patient. PCR reactions consisted of 8 µl 2.5X 5 PRIME HotMaster Mix (5 PRIME Inc., Gaithersburg, MD), 6 µl of 4 µM forward and reverse primer mix, and 200 ng DNA in a 20-µl reaction volume. Thermal cycling consisted of one cycle at 95° C for 2 min; and 30 cycles at 95° C for 20 seconds, 52° C for 20 seconds, and 65° C for 1 min. Each of 4 biopsies per patient was amplified in triplicate, with a single, distinct barcode applied per patient. Ileal and cecal biopsies were assayed separately. Triplicate reactions of individual biopsies were combined, and PCR products were purified using Ampure magnetic purification beads (Beckman Coulter Genomics, Danvers, MA) and quantified with the Quanti-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA) and Nanodrop ND-1000 Spectrophotometer (Nanodrop Technologies, Wilmington, DE).

Equimolar ratios were combined to create two master DNA pools, one for ileum and one for cecum, with a final concentration of 25 ng/μl. Master pools were sent for unidirectional pyrosequencing with primer A at 454 Life Sciences (Branford, CT) on a GS FLX sequencer.

[00202] *Real-time PCR of Bacteroidete and Firmicute 16S rRNA genes.* Primer sequences used for real-time PCR are listed in Table 4.

[00203] **Table 4.** Real-time PCR primers and probes used for gene expression and bacterial quantitative analysis.

Name	SEQ ID NO.	Primers and Probe	Amplicon size (bp)
SI	26	For: 5'-TCTTCATGAGTTTTATGAGGATACGAAC-3'	150
	27	Rev: 5'-TTTGCACCAGATTTCATAATCATACC-3'	
	28	Probe: 5'-CAGATACTGTGAGTGCCTACATCCCTGATGCTATT-3'	
MGAM	29	For: 5'-TACCTTGATGCATAAGGCCCA-3'	150
	30	Rev: 5'-GGCATTACGCTCCAGGACA-3'	
	31	Probe: 5'-CGTCACTGTTGTGCGGCTCTGC-3'	
LCT	32	For: 5'-CAGGAATCAAGAGCGTCACAACT-3'	180
	33	Rev: 5'-AAATCGACCGTGTCTCTGGG-3'	
	34	Probe: 5'-TCCTGCTAGAACCACCCATATCTGCGCT-3'	
SGLT1	35	For: 5'-GCTCATGCCCAATGGACTG-3'	125
	36	Rev: 5'-CGGACCTTGGCGTAGATGTC-3'	
	37	Probe: 5'-ACAGCGCCAGCACCCCTCTTCACC-3'	
Glut2	38	For: 5'-AGTTAGATGAGGAAGTCAAAGCAA-3'	164
	39	Rev: 5'-TAGGCTGTCCGTAGCTGG-3'	
	40	Probe: 5'-ACAAAGCTTGAAAAGACTCAGAGGATATGATGATGTC-3'	
Villin	41	For: 5'-CATGCGCTGAACCTTCATCAA-3'	120
	42	Rev: 5'-GGTTGGACGCTGTCCACTTC-3'	
	43	Probe: 5'-CGGCCGCTCTTTCAGCAGCTCTTCC-3'	
CDX2	44	For: 5'-GGCAGCCAAAGTGAAAACCCAG-3'	112
	45	Rev: 5'-TCCGGATGGTGATGTAGCG-3'	
	46	Probe: 5'-ACCACCAGCGGCTGGAGCTGG-3'	
β-Actin	47	For: 5'-AGCCTCGCCTTTGCCGA-3'	175
	48	Rev: 5'-CTGGTGCCTGGGGCG-3'	
	49	Probe: 5'-CCGCCGCCCGTCCACACCCGCC	
GAPDH	50	For: 5'-CCTGTTGACAGTCAGCCG-3'	100
	51	Rev: 5'-CGACCAATCCGTTGACTCC-3'	
	52	Probe: 5'-CGTCGCCAGCCGAGCCACA-3'	
Bacteroidetes	53	For: 5'-AACGCTAGCTACAGGCTT-3'	~293 (Frank et al.)
	54	Rev: 5'-CCAATGTGGGGGACCTTC-3'	
Firmicutes	55	For: 5'-GGAGYATGTGGTTTAATTCGAAGCA-3'	~126 (Guo et al.)
	56	Rev: 5'-AGCTGACGACAACCATGCAC-3'	
Total Bacteria	57	For: 5'-GTGCCAGCMGCCGCGGTAA-3'	~295 (Frank et al.)
	58	Rev: 5'-GACTACCAGGTATCTAAT-3'	

[00204] PCR standards for determining copy numbers of bacterial 16S rDNA were prepared from representative amplicons of the partial 16S rRNA genes of Bacteroidetes and Firmicutes and total Bacteria cloned into the vector PGEM-T easy (Promega). A representative amplicon with high homology to *Bacteroides Vulgatus* (Accession #: NC_009614) was used with Bacteroidete-specific primers. A representative amplicon with high homology to *Faecalibacterium prausnitzii* (Accession #: NZ_ABED02000023) was used with Firmicute-specific primers. A representative amplicon with high homology to *Bacteroides intestinalis* (Accession #: NZ_ABJL02000007) 16S rRNA gene was used with total Bacteria primers. Cloned sequences were classified using the RDP Seqmatch tool and confirmed by the Microbes BLAST database. Plasmids were linearized with the SphI restriction enzyme and ten-fold serial dilutions of plasmid standards were created ranging from 5×10^7 to 5×10^0 copies for Bacteroidetes, Firmicutes and total Bacteria. Amplification and detection of DNA by real-time PCR were performed with the ABI StepOnePlus Real-time PCR System (Applied Biosystems). Cycling parameters for Bacteroidetes and total Bacteria were as previously described (Frank et al., 2007), as were cycling parameters for Firmicutes (Guo et al., 2008). Each 25- μ l amplification reaction mixture contained 50 ng DNA, 12.5 μ l SYBR Green Master Mix (Applied Biosystems), and 300 nM bacteria-specific (Bacteroidete, Firmicute or total Bacteria) primers. DNA from each of 88 ileal biopsies (4 biopsies per patient) and 88 cecal biopsies (4 biopsies per patient) was assayed in duplicate. The final results were expressed as the mean number of Bacteroidete or Firmicute 16S rRNA gene copies normalized to 16S rRNA gene copies obtained using total Bacterial primers. Eight water/reagent controls were included for all amplifications. The average copy number for water/reagent controls (background) was subtracted from each ileal and cecal amplification prior to normalization. For the Bacteroidete assay all water controls contained undetectable levels of amplification. For the Firmicute assay average amplification signal from water samples were minimal, 12.03 +/- 15.0 copies.

[00205] *Bioinformatic analysis of pyrosequencing reads.* Pyrosequencing reads ranging from 235 to 300 base pairs in length (encompassing all sequences within the major peak obtained from pyrosequencing) were filtered for analysis. Low-quality sequences - i.e., those with average quality scores below 25 - were removed based on previously described criteria (Huse et al., 2007; Hamady et al., 2008). Additionally, reads with any ambiguous characters were omitted

from analysis. Sequences were then binned according to barcode, followed by removal of primer and barcode sequences. Taxonomic classifications of bacterial 16S rRNA sequences were obtained using the RDP Classifier with a minimum 80% bootstrap confidence estimate. To normalize data for differences in total sequences obtained per patient, phylotype abundance was expressed as a percentage of total bacterial sequence reads per patient at all taxonomic levels.

[00206] *Statistical analysis.* Data were not normally distributed, based on Kolmogorov-Smirnov test and evaluation of skewness and kurtosis; thus, the non-parametric Mann-Whitney U test was performed using StatView (Windows version 5.0.1; SAS Institute, Cary, NC). The comparative results of gene expression and bacteria levels were visualized as box-and-whisker plots showing: the median and the interquartile (midspread) range (boxes containing 50% of all values), the whiskers (representing the 25th and 75th percentiles) and the extreme data points (open circles). Associations between different variables were assessed by Spearman rank correlation test. Chi-squared test was used to evaluate between-group genotypes for adult-type hypolactasia. Kruskal-Wallis one-way analysis of variance was employed to assess significance of LCT mRNA expression levels split by genotype and group. Significance was accepted at $p < 0.05$.

[00207] *Genetically determined lactase non-persistence is not responsible for deficient lactase mRNA in AUT-GI.* Although it is beyond the scope of this study to evaluate all possible mutations in carbohydrate genes that may affect expression, we have confirmed that deficient LCT mRNA is not a result of the common adult-type hypolactasia genotype. LCT mRNA levels can be affected by two single nucleotide polymorphisms that determine adult-type hypolactasia; therefore, we genotyped these children using PCR-RFLP analysis (**FIG. 21A**). The homozygous, hypolactasia variant alleles were found in 20% (3 out of 15) of AUT-GI children and 14.3% (1 out of 7) of Control-GI children (chi-squared test, $p=0.896$) (**FIG. 21B**). LCT mRNA expression was significantly lower in individuals with the homozygous hypolactasia genotype compared to all other genotypes (**FIG. 21C**: Mann-Whitney, $p=0.033$). Comparison of LCT mRNA expression across genotype and group failed to reach significance (**FIG. 21D**: Kruskal-Wallis, $p=0.097$). Comparison of mRNA expression in subjects carrying at least one copy of the normal allele confirmed a significant decrease in LCT mRNA in AUT-GI relative to Control-GI subjects, independent of the individuals with the homozygous hypolactasia genotype (**FIG. 21E**:

Mann-Whitney, $p=0.025$). In summary, although our data support the notion that LCT genotype affects gene expression, deficient LCT mRNA in AUT-GI was not attributable to disproportionate hypolactasia genotypes between the AUT-GI and Control-GI groups.

[00208] *Barcoded 16S rRNA gene pyrosequencing.* A total of 525,519 sequencing reads (representing 85% of the initial number of sequencing reads) remained after filtering based on read length, removing low-quality sequences and combining duplicate pyrosequencing runs (271,043 reads for ilea; 254,476 reads for ceca). Binning of sequences by barcode revealed similar numbers of 16S rRNA gene sequence reads per patient (average # sequences per patient +/- STD for ilea = 12,320 +/-1220; average # sequences per patient +/- STD for ceca = 11,567 +/- 1589). There was not a significant difference between the AUT-GI and Control-GI groups in terms of the number of reads per patient. In order to assess whether sufficient sampling was achieved in the total pyrosequencing data set for all AUT-GI and Control-GI subjects, OTUs (Operational Taxonomic Units) were defined at a threshold of 97% identity, split by data for ileum and cecum, and rarefaction analysis was carried out (**FIGS. 23A-23B**). Rarefaction curves showed a tendency toward reaching plateau for all subjects; however failure to reach plateau means that additional sampling would be required to achieve complete coverage of all OTUs present in ileal and cecal biopsies. Investigation of diversity in AUT-GI and Control-GI patients was carried out using the Shannon Diversity Index calculated from OTU data for each subject. Rarefaction analysis revealed that all Shannon Diversity estimates had reached stable values (**FIGS. 23C-23D**). While Shannon Diversity estimates varied widely between individuals, there was not an apparent overall difference (loss or gain of diversity) between the AUT-GI and Control-GI groups in ileal (**FIG. 23C**) or cecal (**FIG. 23D**) biopsies.

[00209] *OTU Analysis of Bacteroidetes.* In order to determine whether the decreased abundance of Bacteroidete members was attributable to the loss of specific Bacteroidete phylotypes, we investigated the distribution of Bacteroidete OTUs (defined using a threshold of 97% identity or greater, 3% distance). The number Bacteroidete OTUs per patient ranged from 23 to 102 for ileal samples and 10 to 130 for cecal samples. Interestingly, no single OTU was significantly over or underrepresented between AUT-GI and Control-GI children and many OTUs contained single sequences. Thus, we sought to determine whether, the decrease in OTUs could be attributed to overall losses of the most prevalent Bacteroidete phylotypes. In both ileal

and cecal samples, 12 OTUs accounted for the majority of Bacteroidete sequences (**FIGS. 25A-25B**). The cumulative levels of these 12 OTUs were significantly lower in AUT-GI compared to Control-GI children in both the ileum (**FIG. 25C**: Mann-Whitney, $p=0.008$) and cecum (**FIG. 25D**: Mann-Whitney, $p=0.008$). Representative sequences from each of these 12 OTUs were classified using Green Genes Blast (greengenes.lbl.gov) and microbial blast alignment (NCBI) (**FIG. 25E**). The majority of sequences were members of the family Bacteroidaceae (OTUs 3, 5, 6, 7, and 19), except in the case of patient 20, where Prevotellaceae were the dominant phylotype. These results suggest that the loss of Bacteroidetes in AUT-GI children is primarily attributable to overall decreases in the dominant phylotypes of Bacteroidetes.

[00210] References

1. Abrams GD, Bauer H, Sprinz H. Influence of the normal flora on mucosal morphology and cellular renewal in the ileum. A comparison of germ-free and conventional mice. *Lab Invest* 1963;12:355-64.
2. Abt MC, Artis D. The intestinal microbiota in health and disease: the influence of microbial products on immune cell homeostasis. *Curr Opin Gastroenterol* 2009;25:496-502.
3. Agarwal S, Mayer L. Pathogenesis and treatment of gastrointestinal disease in antibody deficiency syndromes. *J Allergy Clin Immunol* 2009;124:658-64.
4. Agarwal S, Mayer L. Gastrointestinal manifestations in primary immune disorders. *Inflamm Bowel Dis* 2010;16:703-11.
5. Alberti A, Pirrone P, Elia M, Waring RH, Romano C. Sulphation deficit in "low-functioning" autistic children: a pilot study. *Biol Psychiatry* 1999;46:420-4.
6. Alper CM, Bluestone CD, Buchman C, et al. Recent advances in otitis media. 3. Middle ear physiology and pathophysiology. *Ann Otol Rhinol Laryngol Suppl* 2002;188:26-35.
7. Ashwood P, Wills S, Van de Water J. The immune response in autism: a new frontier for autism research. *J Leukoc Biol* 2006;80:1-15.
8. Backhed F, Ding H, Wang T, et al. The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A* 2004;101:15718-23.

9. Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. *Science* 2005;307:1915-20.
10. Beck PL, Xavier R, Wong J, et al. Paradoxical roles of different nitric oxide synthase isoforms in colonic injury. *Am J Physiol Gastrointest Liver Physiol* 2004;286:G137-47.
11. Born P. Carbohydrate malabsorption in patients with non-specific abdominal complaints. *World J Gastroenterol* 2007;13:5687-91.
12. Bover LC, Cardo-Vila M, Kuniyasu A, et al. A previously unrecognized protein-protein interaction between TWEAK and CD163: potential biological implications. *J Immunol* 2007;178:8183-94.
13. Brockmann K. The expanding phenotype of GLUT1-deficiency syndrome. *Brain Dev* 2009;31:545-52.
14. Brown AM, Ransom BR. Astrocyte glycogen and brain energy metabolism. *Glia* 2007;55:1263-71.
15. Buie T, Campbell DB, Fuchs GJ, 3rd, et al. Evaluation, diagnosis, and treatment of gastrointestinal disorders in individuals with ASDs: a consensus report. *Pediatrics* 2010;125 Suppl 1:S1-18.
16. Buning C, Ockenga J, Kruger S, et al. The C/C(-13910) and G/G(-22018) genotypes for adult-type hypolactasia are not associated with inflammatory bowel disease. *Scand J Gastroenterol* 2003;38:538-42.
17. Burkly LC, Michaelson JS, Hahm K, Jakubowski A, Zheng TS. TWEAKing tissue remodeling by a multifunctional cytokine: role of TWEAK/Fn14 pathway in health and disease. *Cytokine* 2007;40:1-16.
18. Bushara KO. Neurologic presentation of celiac disease. *Gastroenterology* 2005;128:S92-7.
19. Collins SM, Bercik P. The relationship between intestinal microbiota and the central nervous system in normal gastrointestinal function and disease. *Gastroenterology* 2009;136:2003-14.

20. Corbett BA, Kantor AB, Schulman H, et al. A proteomic study of serum from children with autism showing differential expression of apolipoproteins and complement proteins. *Mol Psychiatry* 2007;12:292-306.
21. D'Eufemia P, Celli M, Finocchiaro R, et al. Abnormal intestinal permeability in children with autism. *Acta Paediatr* 1996;85:1076-9.
22. Dawson G. Recent advances in research on early detection, causes, biology, and treatment of autism spectrum disorders. *Curr Opin Neurol* 2010;23:95-6.
23. Dohi T, Borodovsky A, Wu P, et al. TWEAK/Fn14 pathway: a nonredundant role in intestinal damage in mice through a TWEAK/intestinal epithelial cell axis. *Gastroenterology* 2009;136:912-23.
24. Duncan SH, Louis P, Thomson JM, Flint HJ. The role of pH in determining the species composition of the human colonic microbiota. *Environ Microbiol* 2009;11:2112-22.
25. Dyer J, Daly K, Salmon KS, et al. Intestinal glucose sensing and regulation of intestinal glucose absorption. *Biochem Soc Trans* 2007;35:1191-4.
26. Enstrom AM, Onore CE, Van de Water JA, Ashwood P. Differential monocyte responses to TLR ligands in children with autism spectrum disorders. *Brain Behav Immun* 2010;24:64-71.
27. Fabriek BO, van Bruggen R, Deng DM, et al. The macrophage scavenger receptor CD163 functions as an innate immune sensor for bacteria. *Blood* 2009;113:887-92.
28. Fehm HL, Kern W, Peters A. The selfish brain: competition for energy resources. *Prog Brain Res* 2006;153:129-40.
29. Filkova M, Haluzik M, Gay S, Senolt L. The role of resistin as a regulator of inflammation: Implications for various human pathologies. *Clin Immunol* 2009;133:157-70.
30. Finegold SM, Molitoris D, Song Y, et al. Gastrointestinal microflora studies in late-onset autism. *Clin Infect Dis* 2002;35:S6-S16.
31. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol* 2008;6:121-31.

32. Fraser DA, Laust AK, Nelson EL, Tenner AJ. C1q differentially modulates phagocytosis and cytokine responses during ingestion of apoptotic cells by human monocytes, macrophages, and dendritic cells. *J Immunol* 2009;183:6175-85.
33. Fullwood A, Drossman DA. The relationship of psychiatric illness with gastrointestinal disease. *Annu Rev Med* 1995;46:483-96.
34. Furlano RI, Anthony A, Day R, et al. Colonic CD8 and gamma delta T-cell infiltration with epithelial damage in children with autism. *J Pediatr* 2001;138:366-72.
35. Goodwin MS, Cowen MA, Goodwin TC. Malabsorption and cerebral dysfunction: a multivariate and comparative study of autistic children. *J Autism Child Schizophr* 1971;1:48-62.
36. Gupta G, Gelfand JM, Lewis JD. Increased risk for demyelinating diseases in patients with inflammatory bowel disease. *Gastroenterology* 2005;129:819-26.
37. Gupta S, Rimland B, Shilling PD. Pentoxifylline: brief review and rationale for its possible use in the treatment of autism. *J Child Neurol* 1996;11:501-4.
38. Haznedar MM, Buchsbaum MS, Metzger M, Solimando A, Spiegel-Cohen J, Hollander E. Anterior cingulate gyrus volume and glucose metabolism in autistic disorder. *Am J Psychiatry* 1997;154:1047-50.
39. Hodgson S, Ioannides AS. Genetic testing in other GI diseases. *Best Pract Res Clin Gastroenterol* 2009;23:245-56.
40. Hodin RA, Chamberlain SM, Meng S. Pattern of rat intestinal brush-border enzyme gene expression changes with epithelial growth state. *Am J Physiol* 1995;269:C385-91.
41. Hooper LV, Wong MH, Thelin A, Hansson L, Falk PG, Gordon JI. Molecular analysis of commensal host-microbial relationships in the intestine. *Science* 2001;291:881-4.
42. Horvath K, Papadimitriou JC, Rabsztyrn A, Drachenberg C, Tildon JT. Gastrointestinal abnormalities in children with autistic disorder. *J Pediatr* 1999;135:559-63.
43. Iqbal CW, Qandeel HG, Zheng Y, Duenes JA, Sarr MG. Mechanisms of ileal adaptation for glucose absorption after proximal-based small bowel resection. *J Gastrointest Surg* 2008;12:1854-64; discussion 64-5.

44. Ishigame H, Kakuta S, Nagai T, et al. Differential roles of interleukin-17A and -17F in host defense against mucoc epithelial bacterial infection and allergic responses. *Immunity* 2009;30:108-19.
45. Jacobs DM, Gaudier E, van Duynhoven J, Vaughan EE. Non-digestible food ingredients, colonic microbiota and the impact on gut health and immunity: a role for metabolomics. *Curr Drug Metab* 2009;10:41-54.
46. Johansson L, Linner A, Sundén-Cullberg J, et al. Neutrophil-derived hyperresistinemia in severe acute streptococcal infections. *J Immunol* 2009;183:4047-54.
47. Jyonouchi H, Geng L, Ruby A, Zimmerman-Bier B. Dysregulated innate immune responses in young children with autism spectrum disorders: their relationship to gastrointestinal symptoms and dietary intervention. *Neuropsychobiology* 2005;51:77-85.
48. Kalhan SC, Kilic I. Carbohydrate as nutrient in the infant and child: range of acceptable intake. *Eur J Clin Nutr* 1999;53 Suppl 1:S94-100.
49. Kellett GL, Brot-Laroche E, Mace OJ, Leturque A. Sugar absorption in the intestine: the role of GLUT2. *Annu Rev Nutr* 2008;28:35-54.
50. Knivsberg AM, Reichelt KL, Høien T, Nodland M. A randomised, controlled study of dietary intervention in autistic syndromes. *Nutr Neurosci* 2002;5:251-61.
51. Kubes P, McCafferty DM. Nitric oxide and intestinal inflammation. *Am J Med* 2000;109:150-8.
52. Lapointe TK, O'Connor PM, Buret AG. The role of epithelial malfunction in the pathogenesis of enteropathogenic *E. coli*-induced diarrhea. *Lab Invest* 2009;89:964-70.
53. Le Gall M, Tobin V, Stolarczyk E, Dalet V, Leturque A, Brot-Laroche E. Sugar sensing by enterocytes combines polarity, membrane bound detectors and sugar metabolism. *J Cell Physiol* 2007;213:834-43.
54. Leturque A, Brot-Laroche E, Le Gall M. GLUT2 mutations, translocation, and receptor function in diet sugar managing. *Am J Physiol Endocrinol Metab* 2009;296:E985-92.
55. Lossos A, River Y, Eliakim A, Steiner I. Neurologic aspects of inflammatory bowel disease. *Neurology* 1995;45:416-21.

56. Lu JH, Teh BK, Wang L, et al. The classical and regulatory functions of C1q in immunity and autoimmunity. *Cell Mol Immunol* 2008;5:9-21.
57. Lupp C, Robertson ML, Wickham ME, et al. Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. *Cell Host Microbe* 2007;2:204.
58. Lupp C, Robertson ML, Wickham ME, et al. Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. *Cell Host Microbe* 2007;2:119-29.
59. Maher KR, Harper JF, Macleay A, King MG. Peculiarities in the endocrine response to insulin stress in early infantile autism. *J Nerv Ment Dis* 1975;161:180-4.
60. Mariat D, Firmesse O, Levenez F, et al. The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol* 2009;9:123.
61. Mazmanian SK, Round JL, Kasper DL. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* 2008;453:620-5.
62. McNay EC, Gold PE. Food for thought: fluctuations in brain extracellular glucose provide insight into the mechanisms of memory modulation. *Behav Cogn Neurosci Rev* 2002;1:264-80.
63. McNay EC, McCarty RC, Gold PE. Fluctuations in brain glucose concentration during behavioral testing: dissociations between brain areas and between brain and blood. *Neurobiol Learn Mem* 2001;75:325-37.
64. Melis D, Parenti G, Della Casa R, et al. Brain damage in glycogen storage disease type I. *J Pediatr* 2004;144:637-42.
65. Montassir H, Maegaki Y, Ogura K, et al. Associated factors in neonatal hypoglycemic brain injury. *Brain Dev* 2009;31:649-56.
66. Nehlig A. Cerebral energy metabolism, glucose transport and blood flow: changes with maturation and adaptation to hypoglycaemia. *Diabetes Metab* 1997;23:18-29.
67. Nichols BL, Avery SE, Karnsakul W, et al. Congenital maltase-glucoamylase deficiency associated with lactase and sucrase deficiencies. *J Pediatr Gastroenterol Nutr* 2002;35:573-9.

68. Nichols BL, Nichols VN, Putman M, et al. Contribution of villous atrophy to reduced intestinal maltase in infants with malnutrition. *J Pediatr Gastroenterol Nutr* 2000;30:494-502.
69. Nichols BL, Quezada-Calvillo R, Robayo-Torres CC, et al. Mucosal maltase-glucoamylase plays a crucial role in starch digestion and prandial glucose homeostasis of mice. *J Nutr* 2009;139:684-90.
70. Onofre G, Kolackova M, Jankovicova K, Krejsek J. Scavenger receptor CD163 and its biological functions. *Acta Medica (Hradec Kralove)* 2009;52:57-61.
71. Parracho HM, Bingham MO, Gibson GR, McCartney AL. Differences between the gut microflora of children with autistic spectrum disorders and that of healthy children. *J Med Microbiol* 2005;54:987-91.
72. Pascual JM, Wang D, Hinton V, et al. Brain glucose supply and the syndrome of infantile neuroglycopenia. *Arch Neurol* 2007;64:507-13.
73. Pascual JM, Wang D, Lecumberri B, et al. GLUT1 deficiency and other glucose transporter diseases. *Eur J Endocrinol* 2004;150:627-33.
74. Penders J, Stobberingh EE, van den Brandt PA, Thijs C. The role of the intestinal microbiota in the development of atopic disorders. *Allergy* 2007;62:1223-36.
75. Penders J, Thijs C, Vink C, et al. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* 2006;118:511-21.
76. Pfannkuche H, Gabel G. Glucose, epithelium, and enteric nervous system: dialogue in the dark. *J Anim Physiol Anim Nutr (Berl)* 2009;93:277-86.
77. Rautava S, Walker WA. Commensal bacteria and epithelial cross talk in the developing intestine. *Curr Gastroenterol Rep* 2007;9:385-92.
78. Sandler RH, Finegold SM, Bolte ER, et al. Short-term benefit from oral vancomycin treatment of regressive-onset autism. *J Child Neurol* 2000;15:429-35.
79. Scheepers A, Joost HG, Schurmann A. The glucose transporter families SGLT and GLUT: molecular basis of normal and aberrant function. *JPEN J Parenter Enteral Nutr* 2004;28:364-71.

80. Schulzke JD, Troger H, Amasheh M. Disorders of intestinal secretion and absorption. *Best Pract Res Clin Gastroenterol* 2009;23:395-406.
81. Seiderer J, Elben I, Diegelmann J, et al. Role of the novel Th17 cytokine IL-17F in inflammatory bowel disease (IBD): upregulated colonic IL-17F expression in active Crohn's disease and analysis of the IL17F p.His161Arg polymorphism in IBD. *Inflamm Bowel Dis* 2008;14:437-45.
82. Sekirov I, Finlay BB. The role of the intestinal microbiota in enteric infection. *J Physiol* 2009;587:4159-67.
83. Song Y, Liu C, Finegold SM. Real-time PCR quantitation of clostridia in feces of autistic children. *Appl Environ Microbiol* 2004;70:6459-65.
84. Sonnenburg ED, Sonnenburg JL, Manchester JK, Hansen EE, Chiang HC, Gordon JI. A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing in vivo to carbohydrate metabolism. *Proc Natl Acad Sci U S A* 2006;103:8834-9.
85. Stecher B, Hardt WD. The role of microbiota in infectious disease. *Trends Microbiol* 2008;16:107-14.
86. Swallow DM. Genetic influences on carbohydrate digestion. *Nutr Res Rev* 2003;16:37-43.
87. Takahashi T. Pathophysiological significance of neuronal nitric oxide synthase in the gastrointestinal tract. *J Gastroenterol* 2003;38:421-30.
88. Tammali R, Reddy AB, Ramana KV, Petrash JM, Srivastava SK. Aldose reductase deficiency in mice prevents azoxymethane-induced colonic preneoplastic aberrant crypt foci formation. *Carcinogenesis* 2009;30:799-807.
89. Torrente F, Anthony A, Heuschkel RB, Thomson MA, Ashwood P, Murch SH. Focal-enhanced gastritis in regressive autism with features distinct from Crohn's and *Helicobacter pylori* gastritis. *Am J Gastroenterol* 2004;99:598-605.
90. Torrente F, Ashwood P, Day R, et al. Small intestinal enteropathy with epithelial IgG and complement deposition in children with regressive autism. *Mol Psychiatry* 2002;7:375-82, 34.

91. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006;444:1027-31.
92. Ullner PM, Di Nardo A, Goldman JE, et al. Murine Glut-1 transporter haploinsufficiency: postnatal deceleration of brain weight and reactive astrocytosis. *Neurobiol Dis* 2009;36:60-9.
93. Valicenti-McDermott MD, McVicar K, Cohen HJ, Wershil BK, Shinnar S. Gastrointestinal symptoms in children with an autism spectrum disorder and language regression. *Pediatr Neurol* 2008;39:392-8.
94. Van Citters GW, Lin HC. Ileal brake: neuropeptidergic control of intestinal transit. *Curr Gastroenterol Rep* 2006;8:367-73.
95. Wakefield AJ, Ashwood P, Limb K, Anthony A. The significance of ileo-colonic lymphoid nodular hyperplasia in children with autistic spectrum disorder. *Eur J Gastroenterol Hepatol* 2005;17:827-36.
96. Wakefield AJ, Puleston JM, Montgomery SM, Anthony A, O'Leary JJ, Murch SH. Review article: the concept of entero-colonic encephalopathy, autism and opioid receptor ligands. *Aliment Pharmacol Ther* 2002;16:663-74.
97. Warren RP, Odell JD, Warren WL, et al. Brief report: immunoglobulin A deficiency in a subset of autistic subjects. *J Autism Dev Disord* 1997;27:187-92.
98. Wells SM, Buford MC, Migliaccio CT, Holian A. Elevated asymmetric dimethylarginine alters lung function and induces collagen deposition in mice. *Am J Respir Cell Mol Biol* 2009;40:179-88.
99. Wong JM, de Souza R, Kendall CW, Emam A, Jenkins DJ. Colonic health: fermentation and short chain fatty acids. *J Clin Gastroenterol* 2006;40:235-43.
100. Wong JM, Jenkins DJ. Carbohydrate digestibility and metabolic effects. *J Nutr* 2007;137:2539S-46S.
101. Wright EM, Hirayama BA, Loo DF. Active sugar transport in health and disease. *J Intern Med* 2007;261:32-43.

102. Yap IK, Angley M, Veselkov KA, Holmes E, Lindon JC, Nicholson JK. Urinary metabolic phenotyping differentiates children with autism, from their unaffected siblings and age-matched controls. *J Proteome Res* 2010.
103. Yu LC, Flynn AN, Turner JR, Buret AG. SGLT-1-mediated glucose uptake protects intestinal epithelial cells against LPS-induced apoptosis and barrier defects: a novel cellular rescue mechanism? *FASEB J* 2005;19:1822-35.
104. Zhao Y, Fung C, Shin D, et al. Neuronal glucose transporter isoform 3 deficient mice demonstrate features of autism spectrum disorders. *Mol Psychiatry* 2010;15:286-99.
105. Zijlmans WC, van Kempen AA, Serlie MJ, Sauerwein HP. Glucose metabolism in children: influence of age, fasting, and infectious diseases. *Metabolism* 2009;58:1356-65.

Example 2: *Intestinal Inflammation, Impaired Carbohydrate Metabolism and Transport, and Microbial Dysbiosis In Autism*

[00211] The objective of this study was to survey host gene expression and microflora in intestinal biopsies from children with autistic disorder and gastrointestinal complaints (AUT-GI) vs children with gastrointestinal complaints alone (Control-GI).

[00212] This example herein describes a rapid and specific PCR-based assay for diagnostic detection of *Sutterella* species in biological samples. It is a PCR-based detection scheme utilizing new genomic 16S rRNA sequences to allow rapid, sensitive, and specific species identification from gut samples.

[00213] Overview

[00214] **Methods.** Transcription profiling was pursued by cDNA microarray using RNA extracted from ileal biopsies (4 per patient) of 15 male AUT-GI and 7 age-matched, male Control-GI patients. Pathway analysis was performed using Ingenuity Pathway Analysis and GO Ontology. Changes in gene expression were confirmed by quantitative real-time PCR. Intestinal microbiota were investigated in ileal and cecal biopsies from AUT-GI and Control-GI children using amplicon-based, bar-coded pyrosequencing of the V2 region of bacterial 16S rDNA. Taxonomic classification of 525,519 bacterial sequences was accomplished using the Ribosomal

Database Project classifier tool. Differences in microbiota between the two groups were further evaluated and confirmed using Bacteroidete-, Firmicute-, and *Sutterella*-specific real-time PCR.

[00215] Results. Microarray and pathway analysis revealed significant changes in genes involved in carbohydrate metabolism and transport and inflammation in ileal biopsies from AUT-GI as compared to Control-GI subjects. Real-time PCR confirmed significant decreases in the AUT-GI group in the primary brush border disaccharidases, sucrase isomaltase ($p=0.0013$), maltase glucoamylase ($p=0.0027$), and lactase ($p=0.0316$) as well as in two enterocyte hexose transporters, sodium glucose co-transporter 1 ($p=0.0082$) and glucose transporter 2 ($p=0.0101$). In contrast, increases were confirmed for inflammation-related genes in AUT-GI subjects: complement component 1, q subcomponent, A chain ($p=0.0022$), resistin ($p=0.0316$), CD163 ($p=0.0150$), tumor necrosis factor-like weak inducer of apoptosis ($p=0.015$), and interleukin 17F ($p=0.0220$). No significant group differences were observed for the enterocyte-specific marker, villin. In conjunction with changes in intestinal gene expression, bacterial content differed between the AUT-GI and Control-GI groups: pyrosequencing and real-time PCR revealed lower levels of Bacteroidetes (ileum: 50% reduction, $p=0.0027$; cecum: 25% reduction, $p=0.0220$, and higher Firmicute/Bacteroidete ratios in AUT-GI children (ileum: $p=0.0006$; cecum: $p=0.0220$). High levels of *Sutterella* species were found in 47% of AUT-GI biopsies (7/15), whereas *Sutterella* was not detected in any Control-GI biopsies (0/7; ileum: $p = 0.0220$; cecum: $p = 0.0368$).

[00216] Conclusions. We describe a distinctive syndrome in autistic children wherein gastrointestinal dysfunction is associated with altered gene expression reflecting intestinal inflammation, impaired carbohydrate metabolism and transport, and dysbiosis. These findings may provide insights into pathogenesis and enable new strategies for therapeutic intervention.

[00217] In this study, high levels of *Sutterella sp.* were found in ileal and cecal biopsies from children with autism spectrum disorders (ASD) and gastrointestinal disease, while *Sutterella sp.* were undetectable in control children with gastrointestinal disease. Little is known about the epidemiology and pathogenesis of *Sutterella sp.* and their role in infectious diseases of humans and animals. Current methods for detecting *Sutterella sp.* are costly, labor intensive, and non-specific requiring isolation and anaerobic culture of the bacteria or generation, screening,

sequencing, and sequence analysis of hundreds to thousands of bacterial 16S rRNA gene sequences from bacterial libraries or pyrosequencing analysis of hundreds of thousands of sequences. These methods can be costly, lack specificity, ease of execution, and are not strictly quantitative.

[00218] Herein, we describe a rapid and specific PCR-based assay for the diagnostic identification, quantification, and phylogenetic analysis of *Sutterella sp.* in biological samples based on the variable sequence (V6-V8 region) of the 16S rRNA gene of *Sutterella sp.*

[00219] Study Background

[00220] An association between autistic spectrum disorder (ASD) and gastrointestinal (GI) immunopathology is supported by reports of a higher incidence of GI complaints, ileo-colonic lymphoid nodular hyperplasia, and enterocolitis in children with autism. In this study, intestinal bacteria were assessed in ileal (4 biopsies per patient) and cecal (4 biopsies per patient) biopsies from male ASD children (aged 3-5 years) with gastrointestinal symptoms (ASD-GI; n=15) and normally developing age-matched, male controls with gastrointestinal symptoms (Control-GI; n=7) by 454 pyrosequencing of the V2 region of the bacterial 16S rRNA gene. Taxonomic classification of 525,519 bacterial sequences was performed using the Ribosomal Database Project classifier tool. Genus-level analysis of pyrosequencing reads revealed a significant increase in *Sutterella sp.* The average confidence estimate of all genus-level *Sutterella* sequences identified using the RDP Classifier was high (99.1%) with the majority of sequences at 100% confidence.

[00221] Comparison of ASD-GI and Control-GI patients revealed significant increases in *Sutterella sp.* In the ileum (**FIG. 8A**: Mann-Whitney U, p=0.022) and cecum (**FIG. 8B**: Mann-Whitney U, p=0.0368). *Sutterella sp.* sequences were completely absent from all Control-GI samples (% of total bacteria=0). Individual analysis of ASD-GI patients revealed that 7 out of 15 ASD-GI patients (46.7%) had high levels of *Sutterella sp.* sequences in both the ileum and cecum (**FIG. 8C** and **FIG. 8D**). By patient, ileal *Sutterella sp.* sequence abundance ranged from 1.7 to 6.7% of total bacterial reads (**FIG. 8C**). Similarly, in the Cecum *Sutterella sp.* sequence abundance ranged from 1.9 to 7.0% of total bacterial reads for the same patients (**FIG. 8D**). *Sutterella sp.* Sequences represented the majority of sequences present in the class Beta-

proteobacteria in these select ASD-GI patients. In the Ileum of these ASD-GI patients, *Sutterella* *sp.* sequences accounted for 75.6% to 97.8% of all Beta-proteobacteria sequences (**Fig. 8E**). In the cecum, *Sutterella* *sp.* sequences accounted for 92.7% to 98.2% of all Beta-proteobacteria sequences (**Fig. 8F**). The results of this costly, time consuming, non-specific pyrosequencing analysis prompted the design of a *Sutterella* *sp.*-specific PCR assay to confirm, quantitate, and determine taxonomy of *Sutterella* *sp.* in the same samples analyzed by pyrosequencing.

[00222] **Methods**

[00223] Primer and Probe Design: *Sutterella* *sp.*-specific 16S rRNA gene PCR primers and probe were designed against the 16S sequence for *Sutterella wadsworthensis* (Genbank Accession # L37785) and *Sutterella* clone LW53 (Genbank Accession # AY976224) using Primer Express 1.0 software (Applied Biosystems, Foster City, CA). Genus specificity of candidate primers was evaluated using the RDP Probe Match tool. While several potential primer pairs were identified, only one pair showed high specificity for *Sutterella* *sp.* In PCR assays. These primers are designated here as SuttFor and SuttRev (Sequences of primers and probe are shown in **Table 1**).

[00224] **Table 1.** *Sutterella* *sp.*-specific primers and probes for classical and real-time PCR assays and pan-bacterial primers used for normalization.

SEQ ID NO:	Primers and Probe	Amplicon size (bp)
11	SuttFor: 5'-CGCGAAAAACCTTACCTAGCC-3'	~260
12	SuttRev: 5'-GACGTGTGAGGCCCTAGCC-3'	
13	SuttProbe1: 5'-CACAGGTGCTGCATGGCTGTCGT-3'	
14	SuttProbe2: 5'-CCG CAAGGGAATCTGGACACAGGT-3'	
15	515For: 5'-GTGCCAGCMGCCGCGGTAA-3'	~295 (Frank et al.)
16	805Rev: 5'-GACTACCAGGGTATCTAAT-3'	

[00225] Evaluation of good quality sequences that were >1200 bases in the RDP database revealed a total of 248 *Sutterella* sequences at the time of analysis. SuttFor and SuttRev primers

showed high exclusivity for the genus *Sutterella*. Approximately 90% of RDP matches for SuttFor were in the genus *Sutterella* and 100% of matches for the reverse primer were *Sutterella* sequences. The SuttFor primer sequence matched exactly with approximately 91% (225/248 *Sutterella* sequences) of all *Sutterella* sequences, while the SuttRev primer matched exactly with approximately 81% (200/248 *Sutterella* sequences) of all *Sutterella* sequences. The SuttProbe1 used for real-time PCR had low exclusivity but high coverage of *Sutterella* sequences (100%). An additional probe with high exclusivity, but low coverage of *Sutterella* sequences (58.8%) was also designed and can be used when sequence information is available for *Sutterella* sp. in biological samples.

[00226] Classical PCR. The SuttFor and SuttRev primers amplify a 260 bp region between variable regions 6, 7 and 8 (V6-V8) of the 16S rRNA of *Sutterella*. Classical PCR for detection of *Sutterella* was carried out in 25ul reactions consisting of 25ng genomic DNA, 300nm each SuttFor and SuttRev primers, 2ul dNTP mix (10mM; Applied Biosystems), 2.5ul of 10x PCR Buffer (Qiagen), 5U of HotStarTaq DNA polymerase (Qiagen), and 5ul Q-solution (Qiagen). Cycling parameters consisted of an initial denaturation step at 950C for 15 min, followed by 30 cycles of 940C for 1 min, 600C for 1 min, and 720C for 1 min and a final extension at 720C for 5 min. Amplified products were run on a 1.5% agarose gel, extracted from the gel and either sent for direct PCR product sequencing using SuttFor and SuttRev primers or cloned into PGEM-T easy cloning vector for construction of bacterial libraries followed by sequencing using vector primers. Specificity of the assay was confirmed through direct sequence analysis of PCR products and clone sequences using the RDP Seqmatch and Classifier tools. All PCR products and clones were classified as *Sutterella* by RDP. In order to test linearity and sensitivity of the assay, the *Sutterella* clone used for real-time PCR standards was tested by classical PCR using the same conditions as all intestinal DNA. Ten fold dilutions of the *Sutterella* clone ranging from 5x10⁵ to 5x10⁰ were amplified by classical PCR alone as well as spiked into ileal DNA from a *Sutterella* negative patient. Both in the presence and absence of background ileal DNA, the Classical PCR was linear in the range of 5x10⁵ to 5x10² copies and had an end-point detection limit of 5x10¹ copies (**FIG. 9**).

[00227] Quantitative Real-time PCR. PCR standards for determining copy numbers of bacterial 16S rDNA were prepared from representative clones of the partial 16S rDNA of

Sutterella obtained using the Classical PCR assay. Cloned sequences were classified using the RDP Seqmatch tool and confirmed by the Microbes BLAST database. Plasmids were linearized with the *SphI* restriction enzyme and ten fold serial dilutions of plasmid standards were created ranging from 500,000 to 5 copies for *Sutterella* (**FIG. 10A** and **FIG. 10B**). Amplification and detection of DNA by real-time PCR were performed with the ABI StepOnePlus Real-time PCR System (Applied Biosystems). For *Sutterella sp.*-specific real-time PCR, each 25ul reaction contained 50ng DNA, 12.5ul Taqman universal master mix (ABI), 300nm each of SuttFor and SuttRev primers, and 200nm SuttProbe1 (Reporter=FAM, Quencher=BBQ). The standard curve had sensitivity down to 5 copies of plasmid, with a slope of -3.08 , y-intercept of 41.787, and with an R2 value of 0.996 (**FIG. 10A** and **FIG. 10B**). DNA from each of 88 ileal biopsies and 88 cecal biopsies was assayed in duplicate. The final results were expressed as the mean number of copies normalized to 16S rRNA copies obtained using Pan-bacterial primers (**Table 1**: primers 515For and 805Rev) in a SYBR Green Real-time PCR assay (see Ref. 6 for more information). While normalization to total bacteria is not necessary, we have implemented its use in this study to control for variation in input DNA. Eight water/reagent controls were included for all amplifications. The average copy number for water controls (background) was subtracted from each ileal and cecal amplification prior to normalization. Where background copy number values exceeded amplification values in ileal and cecal samples, copy number was set to a value of 0. Average amplification signal from water samples with the *Sutterella* assay were very low (125.8 +/- 40 copies) compared to amplification in *Sutterella* positive samples (all ranging between 50,000 and 1,000,000 copies). Average copy numbers for all ileum and cecum *Sutterella*-negative amplifications was 26.6 +/- 21.0 copies (all were lower than the background controls).

[00228] *Taxonomic Classification of Sutterella sp.* Sequence alignments using sequences obtained by direct sequencing of *Sutterella sp.* from the classical PCR assay and phylogenetic analyses were conducted using MEGA4 software. Primer sequences were trimmed from the sequences obtained by direct sequencing of amplicons. Classification was confirmed using the RDP classifier and seqmatch tools. *Sutterella* sequences obtained from ileal and cecal biopsies were aligned with sequences from the 11 known isolates of *Sutterella sp.* found in the RDP database. Sequences from known *Sutterella sp.* Isolates were trimmed to the length of the sequences obtained from ileal and cecal biopsies. Phylogenetic trees were constructed according

to the neighbour-joining method, rooted to the outgroup *Burkholderia pseudomallei*, and the stability of the groupings was estimated by bootstrap analysis (1000 replications) using MEGA4.

[00229] Results

[00230] Implementation of *Sutterella* sp.-specific Classical PCR for Detection. Classical PCR analysis of *Sutterella* sp. using DNA from all 88 ileal and 88 cecal biopsies showed that the same individuals identified as having high levels of *Sutterella* by V2 pyrosequencing were also positive by the V6-V8 *Sutterella* sp.-specific PCR. Additionally, all 4 biopsies per region in all 7 *Sutterella*-positive patients showed *Sutterella* amplicons, while no amplicons were observed in any Control-GI patients or ASD-GI patients that lacked *Sutterella* sequences in V2 pyrosequencing experiments (**FIG. 11**). All patients amplicons were confirmed to represent *Sutterella* by direct sequencing of PCR products and cloning of individual amplicons to create bacterial libraries followed by sequencing of 50 individual clones.

[00231] Implementation of *Sutterella* sp.-specific Real-time PCR for Quantification. Real-time PCR analysis using the same V6-V8 primers and a high coverage Taqman probe (SuttProbe1), revealed significant increases in *Sutterella* in ASD-GI compared to Control-GI patients for both the ileum (**FIG. 12A**:Mann-Whitney U, $p=0.0368$) and cecum (**FIG. 12B**:Mann-Whitney U, $p=0.0368$). *Sutterella* copy numbers were quite high in both the ileum and cecum (in the range of 10^4 to 10^5 copies) of *Sutterella*-positive patients (**FIG. 12C** and **FIG. 12D**). The distribution of *Sutterella* abundance by patient and the copy number revealed by V2 pyrosequencing and V6-V8 real-time PCR, respectively, were in striking concordance (Compare ileum **FIG. 8C** with **FIG. 12C** and compare cecum **FIG. 8D** with **FIG. 12D**). There was 100% congruence between V2 region 454 pyrosequencing and both classical and real-time PCR using the V6-V8 region *Sutterella* sp.-specific primers.

[00232] Implementation of *Sutterella* sp.-specific Classical PCR for Taxonomic Classification. Sequences obtained from direct cloning and clone libraries of the V6-V8 regions of each patient were aligned following removal of primer sequences. This analysis revealed that the consensus sequence obtained in ileal biopsies matched exactly with sequences in cecal biopsies from the same patient. Furthermore, alignment of sequences revealed that patients 1, 3, 10, 11, and 12 had the exact same sequence for the V6-V8 region, while patients 5 and 7 had a

distinct, but identical sequence (**FIG. 13**). These findings are in agreement with OTU analysis of V2 pyrosequencing reads in which patients 1, 3, 10, 11, and 12 clustered together with OTU 11 containing the majority of *Sutterella* sequences and patient 5 and 7 clustered together with OTU 38 containing the majority *Sutterella* sequences (**FIG. 14**). Treeing analysis of the V6-V8 sequences revealed that *Sutterella sp.* found in patients 1,3, 10, 11, and 12 were phylogenetically most closely associated with the isolates *Sutterella stercoricanis* (supported by a bootstrap resampling value of 70%) and *Parasutterella sp.* (supported by a bootstrap resampling value of 68%). In contrast, treeing analysis revealed that *Sutterella sp.* sequences found in patients 5 and 7 were most closely associated with the isolate *Sutterella wadsworthensis* (supported by a bootstrap resampling value of 94%) (**FIG. 15A**). These findings were consistent with treeing analysis obtained from V2 sequences obtained from pyrosequencing analysis in which V2 *Sutterella* sequences from patients 1,3, 10, 11, and 12 were most closely associated with the isolates *Sutterella stercoricanis* and *Sutterella sanguinus* (supported by a bootstrap resampling value of 67%) while the V2 *Sutterella* sequences from patients 5 and 7 were most closely associated with the isolates of *Sutterella wadsworthensis* (supported by a bootstrap resampling value of 100%) (**FIG. 15B**). Thus, sequences from patients 5 and 7 clustered with *Sutterella wadsworthensis* isolates using both the V2 pyrosequencing reads and the V6-V8 sequences obtained from this assay. In contrast, sequences from patients 1, 3, 10, 11, and 12 clustered with *Sutterella stercoricanis* using both the V2 pyrosequencing reads and the V6-V8 sequence obtained from this assay. However, there was some divergence between the V2 and V6-V8 regions in determining relationships to other isolates (i.e. relatedness to *Sutterella sanguinus* from the V2 sequences and relatedness to *Parasutterella sp.* from the V6-V8 sequences).

[00233] **References**

- A1.) Wexler HM, Reeves D, Summanen PH, Molitoris E, McTeague M, Duncan J, Wilson KH, Finegold SM. 1996. *Sutterella wadsworthensis* gen. nov., sp. nov., bile-resistant microaerophilic *Campylobacter gracilis*-like clinical isolates. *Int J Syst Bacteriol*, 46(1): 252-258.
- A2.) Mangin I, Bonnet R, Seksik P, Rigottier-Gois L, Sutren M, Bouhnik Y, Neut C, Collins MD, Colombel JF, Marteau P, Doré J. 2004. Molecular inventory of faecal microflora in patients with Crohn's disease. *FEMS Microbiol Ecol*, 50(1): 25-36.

- A3.) Gophna U, Sommerfeld K, Gophna S, Doolittle WF, Veldhuyzen van Zanten SJ. 2006. Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *J Clin Microbiol*, 44(11): 4136-4141.
- A4.) Greetham HL, Collins MD, Gibson GR, Giffard C, Falsen E, Lawson PA. 2004. *Sutterella stercoricanis* sp. nov., isolated from canine faeces. *Int J Syst Evol Microbiol*. 54: 1581-1584.
- A5.) J Scupham A, Patton TG, Bent E, Bayles DO. 2008. Comparison of the cecal microbiota of domestic and wild turkeys. *Microb Ecol*. 56: 322-331.
- A6.) Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. 2007. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci USA*. 104: 13780-13785.
- King A, Downes J, Nord CE, Phillips I; European Study Group. 1999. Antimicrobial susceptibility of non-*Bacteroides fragilis* group anaerobic Gram-negative bacilli in Europe. *Clin Microbiol Infect*. 5: 404-416.
- Goldstein EJ, Citron DM. 2009. Activity of a novel carbapenem, doripenem, against anaerobic pathogens. *Diagn Microbiol Infect Dis*. 63: 447-454.
- Wexler HM, Molitoris D, St John S, Vu A, Read EK, Finegold SM. 2002. In vitro activities of faropenem against 579 strains of anaerobic bacteria. *Antimicrob Agents Chemother*. 46: 3669-3675.
- Wexler HM, Molitoris D, Finegold SM. 2000. In vitro activities of MK-826 (L-749,345) against 363 strains of anaerobic bacteria. *Antimicrob Agents Chemother*. 44: 2222-2224.
- Molitoris E, Wexler HM, Finegold SM. 1997. Sources and antimicrobial susceptibilities of *Campylobacter gracilis* and *Sutterella wadsworthensis*. *Clin Infect Dis. Suppl 2*: S264-265.
- Wexler HM, Molitoris E, Molitoris D, Finegold SM. 1996. In vitro activities of trovafloxacin against 557 strains of anaerobic bacteria. *Antimicrob Agents Chemother*. 40: 2232-2235.

What is claimed is:

1. A method for detecting the presence of or a predisposition to autism or an autism spectrum disorder (ASD) in a human subject, the method comprising:
 - (a) obtaining a biological sample from a human subject; and
 - (b) detecting whether or not there is an alteration in the expression of a carbohydrate metabolic enzyme protein in the subject as compared to a non-autistic subject.
2. A method for detecting the presence of or a predisposition to autism or an autism spectrum disorder (ASD) in a human subject, the method comprising:
 - (a) obtaining a biological sample from a human subject; and
 - (b) detecting whether or not there is an alteration in the expression of a carbohydrate transporter protein in the subject as compared to a non-autistic subject.
3. The method of claim 1 or 2, wherein the subject is a child of a human subject.
4. The method of claim 1, wherein the carbohydrate metabolic enzyme comprises sucrase isomaltase, maltase glucoamylase, lactase, or a combination thereof.
5. The method of claim 2, wherein the carbohydrate transporter comprises GLUT2, SGLT1, or a combination thereof.
6. The method of claim 1 or claim 2 further comprising detecting a decrease in Bacteroidetes, an increase in the Firmicute/Bacteroidete ratios, an increase in

cumulative levels of Firmicutes and Proteobacteria, an increase in Beta-proteobacteria, or an increase in *Sutterella sp.* in the small intestine or large intestine of the subject.

7. The method of claim 1 or claim 2 further comprising detecting an increase in *Sutterella sp.* in the small intestine or large intestine of the subject.
8. The method of claim 1, wherein the detecting comprises detecting whether there is an alteration in a gene locus that encodes the carbohydrate metabolic enzyme.
9. The method of claim 2, wherein the detecting comprises detecting whether there is an alteration in a gene locus that encodes the carbohydrate transporter.
10. The method of claim 1 or claim 2, wherein the detecting comprises detecting whether mRNA expression of the protein is reduced.
11. The method of claim 1 or claim 2, wherein the subject is a human embryo, a human fetus, or an unborn human child.
12. The method of claim 1 or claim 2, wherein the sample comprises blood, serum, sputum, lacrimal secretions, semen, vaginal secretions, fetal tissue, small intestine tissue, large intestine tissue, liver tissue, amniotic fluid, or a combination thereof.
13. A method for treating or preventing autism or an ASD in a subject in need thereof, the method comprising administering to the subject a therapeutic amount of a pharmaceutical composition comprising a functional carbohydrate metabolic

enzyme molecule, a functional carbohydrate transporter molecule, or a combination thereof, thereby treating or preventing autism or an ASD.

14. The method of claim 13, wherein the administering comprises a subcutaneous, intra-muscular, intra-peritoneal, or intravenous injection; an infusion; oral, nasal, or topical delivery; or a combination thereof.
15. The method of claim 13, wherein the administering comprises delivery of a functional carbohydrate metabolic enzyme molecule or a functional carbohydrate transporter molecule to the gastrointestinal tract of the subject.
16. The method of claim 13, wherein the administering comprises feeding the human subject or child thereof a therapeutically effective amount of a functional carbohydrate metabolic enzyme molecule or a functional carbohydrate transporter molecule.
17. The method of claim 13, wherein administering occurs daily, weekly, twice weekly, monthly, twice monthly, or yearly.
18. An isolated nucleic acid composition, the composition comprising a nucleic acid molecule having at least about 90% identity to SEQ ID NO: 11, 12, 13, or 14.
19. A diagnostic kit for detecting the presence of *Sutterella sp.* in a sample, the kit comprising a nucleic acid molecule that specifically hybridizes to or a primer combination that amplifies a *Sutterella sp.* 16S nucleic acid sequence.
20. A diagnostic kit for determining whether a sample from a subject exhibits a presence of or a predisposition to autism or an autism spectrum disorder (ASD), the kit comprising a nucleic acid primer that specifically hybridizes to an autism

biomarker, wherein the primer will prime a polymerase reaction only when an autism biomarker is present.

21. The kit of claim 19, wherein the nucleic acid molecule comprises a nucleic acid primer or nucleic acid probe.
22. The kit of claim 19, wherein the 16S nucleic acid sequence comprises at least about 90% of SEQ ID NO: 59 or SEQ ID NO: 60.
23. The kit of claim 21, wherein the probe comprises a nucleotide sequence having SEQ ID NOS: 13 or 14 in Table 1, or the italicized nucleotide of sequence SEQ ID NO: 19.
24. The kit of claim 21, wherein the probe comprises at least 10 consecutive nucleotide bases comprising SEQ ID NO: 19, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C nucleotide.
25. The kit of claim 21, wherein the probe comprises a reverse complement of SEQ ID NOS: 11, 12, 15, 16, 17, 18, or 19, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C nucleotide.
26. The kit of claim 20 or 21, wherein the primer comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS: 11, 12, 15, 16, 17, or 18, wherein, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C nucleotide.

27. The kit of claim 20, wherein the autism biomarker is a carbohydrate transporter molecule, a carbohydrate metabolic enzyme molecule, or a gastrointestinal *Sutterella sp.* bacterium.
28. The kit of claim of 27, wherein the carbohydrate transporter molecule is GLUT2 or SGLT1.
29. The kit of claim of 27, wherein the carbohydrate metabolic enzyme molecule is SI, MGAM, or LCT.
30. The kit of claim 19 or 20, wherein the sample is from a human or non-human animal.
31. The kit of claim 19 or 20, wherein the sample comprises intestinal tissue, feces, blood, skin, or a combination thereof.
32. A method of treating or preventing a disease associated with elevated levels of Beta-proteobacteria, the method comprising administering to a subject in need thereof a therapeutic amount of an antimicrobial composition effective against Beta-proteobacteria for treating the disease.
33. The method of claim 32, wherein the antimicrobial composition is an antibiotic, a probiotic agent, or a combination thereof.
34. The method of claim 32, wherein the disease is ASD, autism, or a gastrointestinal disease.
35. The method of claim 34, wherein the gastrointestinal disease is diarrhea, inflammatory bowel disease, antimicrobial-associated colitis, or irritable bowel syndrome.
36. The method of claim 35, wherein the diarrhea or inflammatory bowel diseases is ulcerative colitis or Crohn's disease.

37. The method of claim 33, wherein the antibiotic comprises lincosamides, chloramphenicols, tetracyclines, aminoglycosides, beta-lactams, vancomycins, bacitracins, macrolides, amphotericins, sulfonamides, methenamin, nitrofurantoin, phenazopyridine, trimethoprim; rifampicins, metronidazoles, cefazolins, lincomycin, spectinomycin, mupirocins, quinolones, novobiocins, polymixins, gramicidins, antipseudomonals, or a combination thereof.
38. The method of claim 33, wherein the probiotic agent comprises *Bacteroides*, *Prevotella*, *Porphyromonas*, *Fusobacterium*, *Sutterella*, *Bilophila*, *Campylobacter*, *Wolinella*, *Butyrovibrio*, *Megamonas*, *Desulfomonas*, *Desulfovibrio*, *Bifidobacterium*, *Lactobacillus*, *Eubacterium*, *Actinomyces*, *Eggerthella*, *Coriobacterium*, *Propionibacterium*, other genera of non-sporeforming anaerobic gram-positive bacilli, *Bacillus*, *Peptostreptococcus*, newly created genera originally classified as *Peptostreptococcus*, *Peptococcus*, *Acidaminococcus*, *Ruminococcus*, *Megasphaera*, *Gaffkya*, *Coprococcus*, *Veillonella*, *Sarcina*, *Clostridium*, *Aerococcus*, *Streptococcus*, *Enterococcus*, *Pediococcus*, *Micrococcus*, *Staphylococcus*, *Corynebacterium*, species of the genera comprising the Enterobacteriaceae and Pseudomonadaceae, or a combination thereof.
39. A method of detecting a *Sutterella sp.* in a sample, the method comprising:
- (a) selecting a *Sutterella sp.*-specific primer pair, wherein the primer pair mediates amplification of a polynucleotide amplicon of a selected, known length from a nucleic acid of a *Sutterella sp.*;
 - (b) contacting a nucleic acid from the sample with the *Sutterella sp.*-specific primer pair in a reaction mixture under conditions that promote amplification of a polynucleotide amplicon, wherein the primer pair will prime a polymerase reaction only when the nucleic acid of a *Sutterella sp.* is present; and

- (c) detecting the amplicons, wherein the detection of an amplicon of a selected, known length is indicative of the sample containing the nucleic acid of a *Sutterlla sp.*
40. The method of claim 39, wherein the sample comprises intestinal tissue, feces, blood, skin, or a combination thereof.
41. The method of claim 39, wherein the primer pair comprises a forward primer and a reverse primer.
42. The method of claim 41, wherein the forward primer comprises SEQ ID NO: 11 or 17, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C nucleotide.
43. The method of claim 41, wherein the reverse primer comprises SEQ ID NO: 12 or 18, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C nucleotide.
44. The method of claim 41, wherein the forward primer comprises at least 10 consecutive nucleotide bases comprising SEQ ID NO: 17 or 19, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, wherein H is an A nucleotide and/or T nucleotide and/or C

nucleotide, wherein B is a T nucleotide, C nucleotide, or G nucleotide, wherein V is an A nucleotide, G nucleotide, or C nucleotide; wherein D is an A nucleotide, G nucleotide, or T nucleotide; and wherein K is a G nucleotide or T nucleotide.

45. The method of claim 41, wherein the reverse primer comprises at least 10 consecutive nucleotide bases comprising SEQ ID NO: 18 or 19, wherein S is a G nucleotide and/or a C nucleotide, wherein Y is a C nucleotide and/or T nucleotide, wherein R is an A nucleotide and/or G nucleotide, wherein W is an A nucleotide and/or T nucleotide, and wherein H is an A nucleotide and/or T nucleotide and/or C.
46. An isolated nucleic acid composition, the composition comprising a nucleic acid molecule having at least about 98% identity to SEQ. ID NO: 11, 12, 13, or 14.
47. An isolated nucleic acid composition, the composition comprising a nucleic acid molecule comprising SEQ. ID NO: 11, 12, 13, or 14.

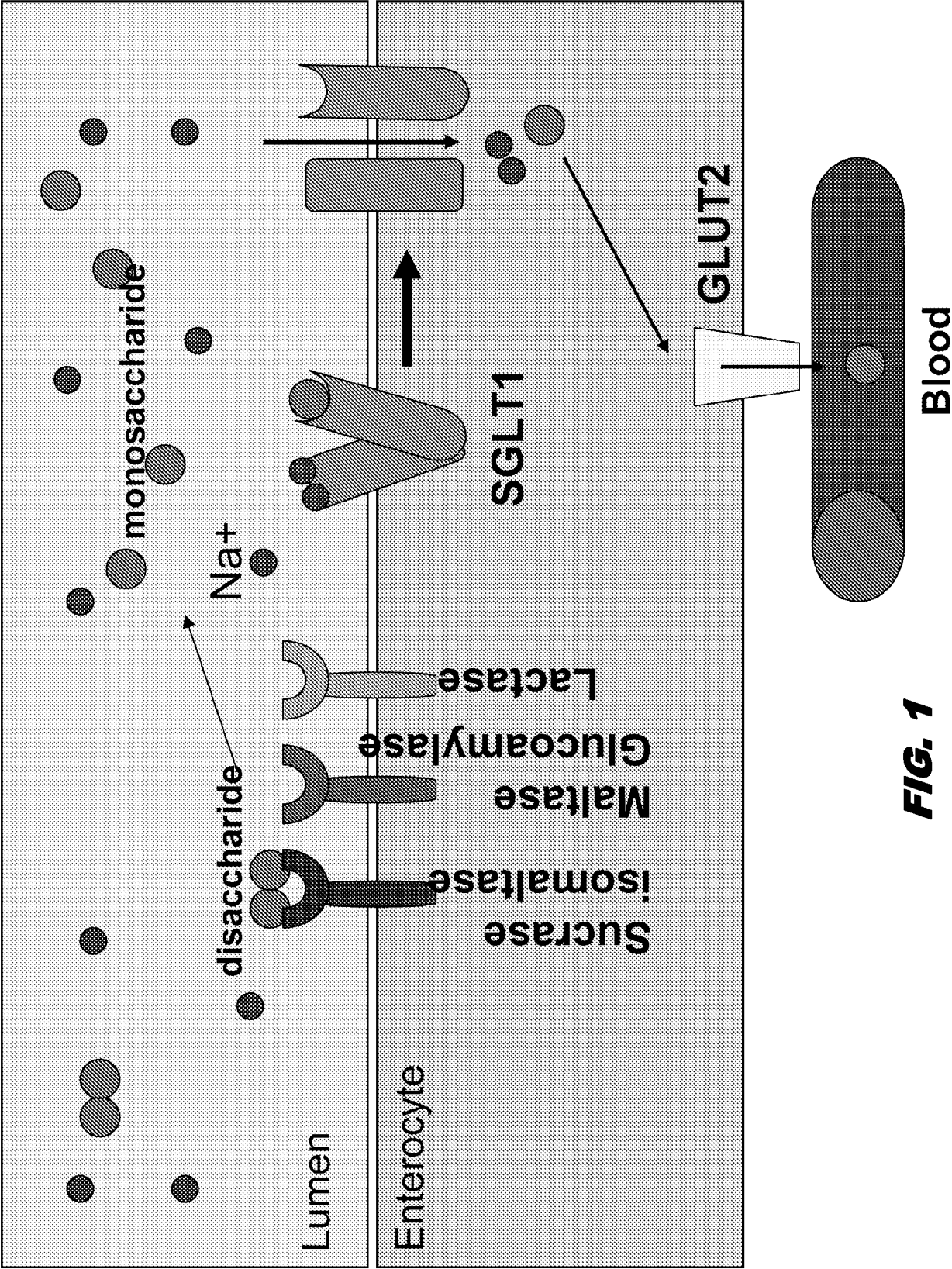


FIG. 1

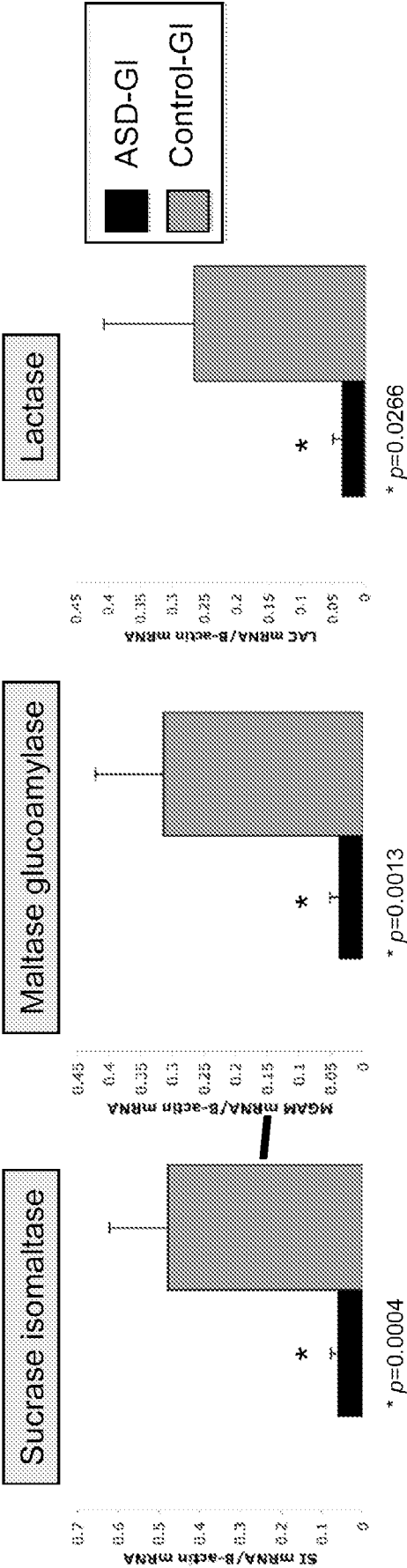
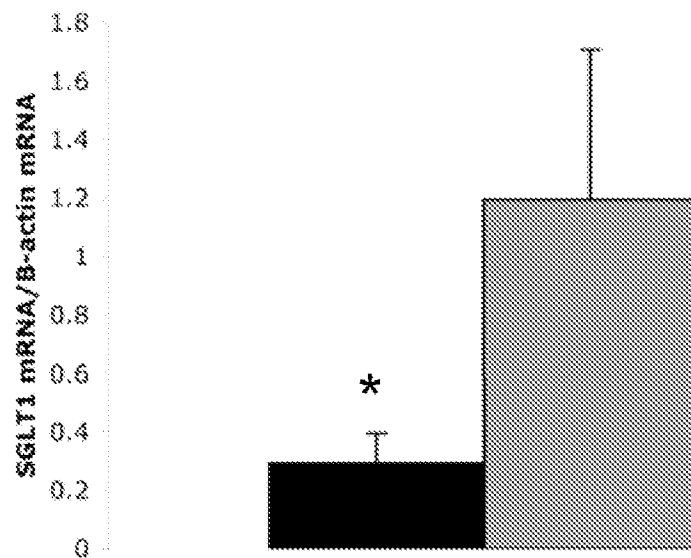


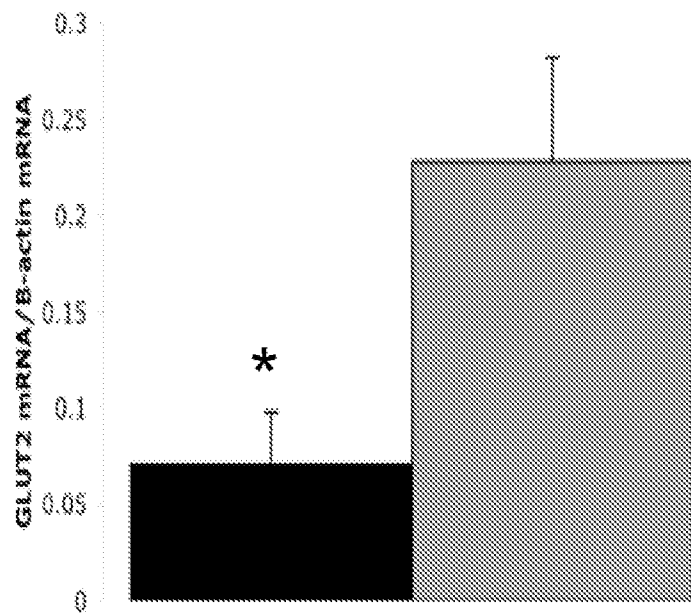
FIG. 2

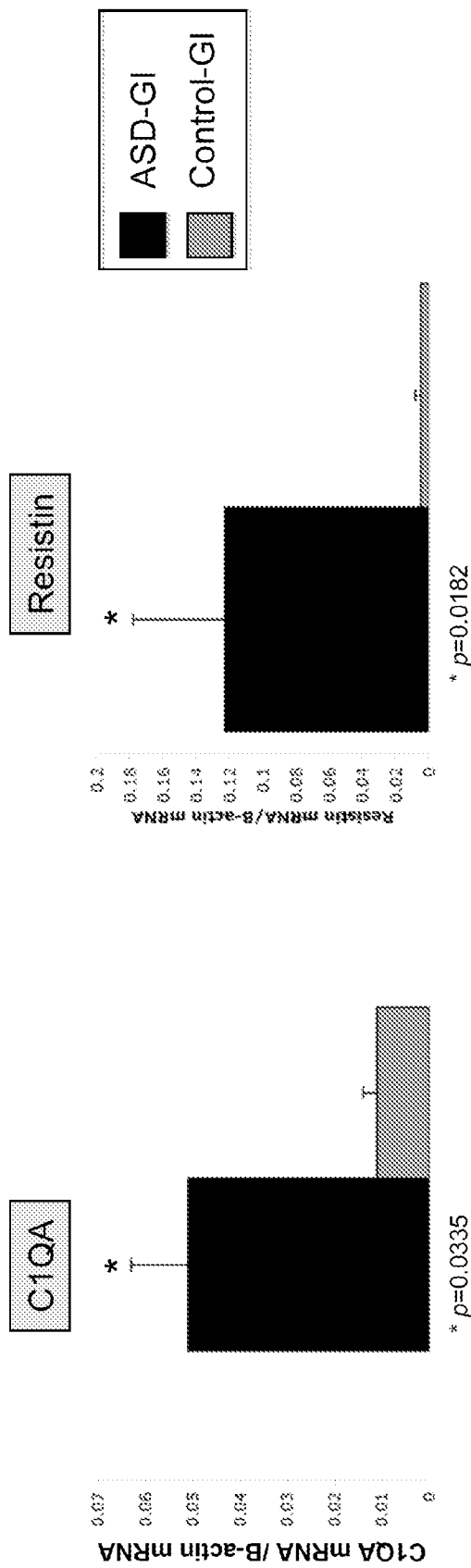
3 of 46

SGLT1

* $p=0.0252$

GLUT2

* $p=0.0083$ **FIG. 3**



IL17F mRNA per patient

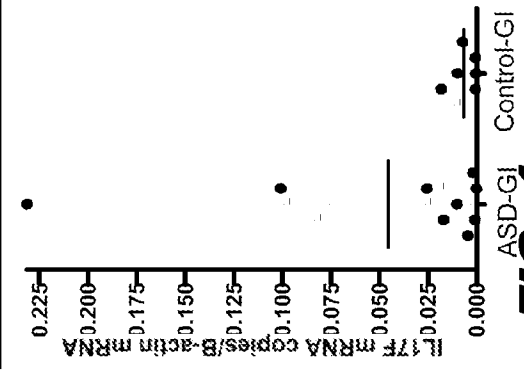
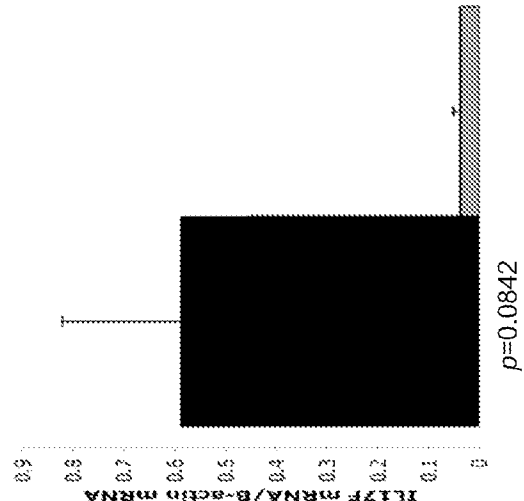


FIG. 4

IL17F real time RT-PCR



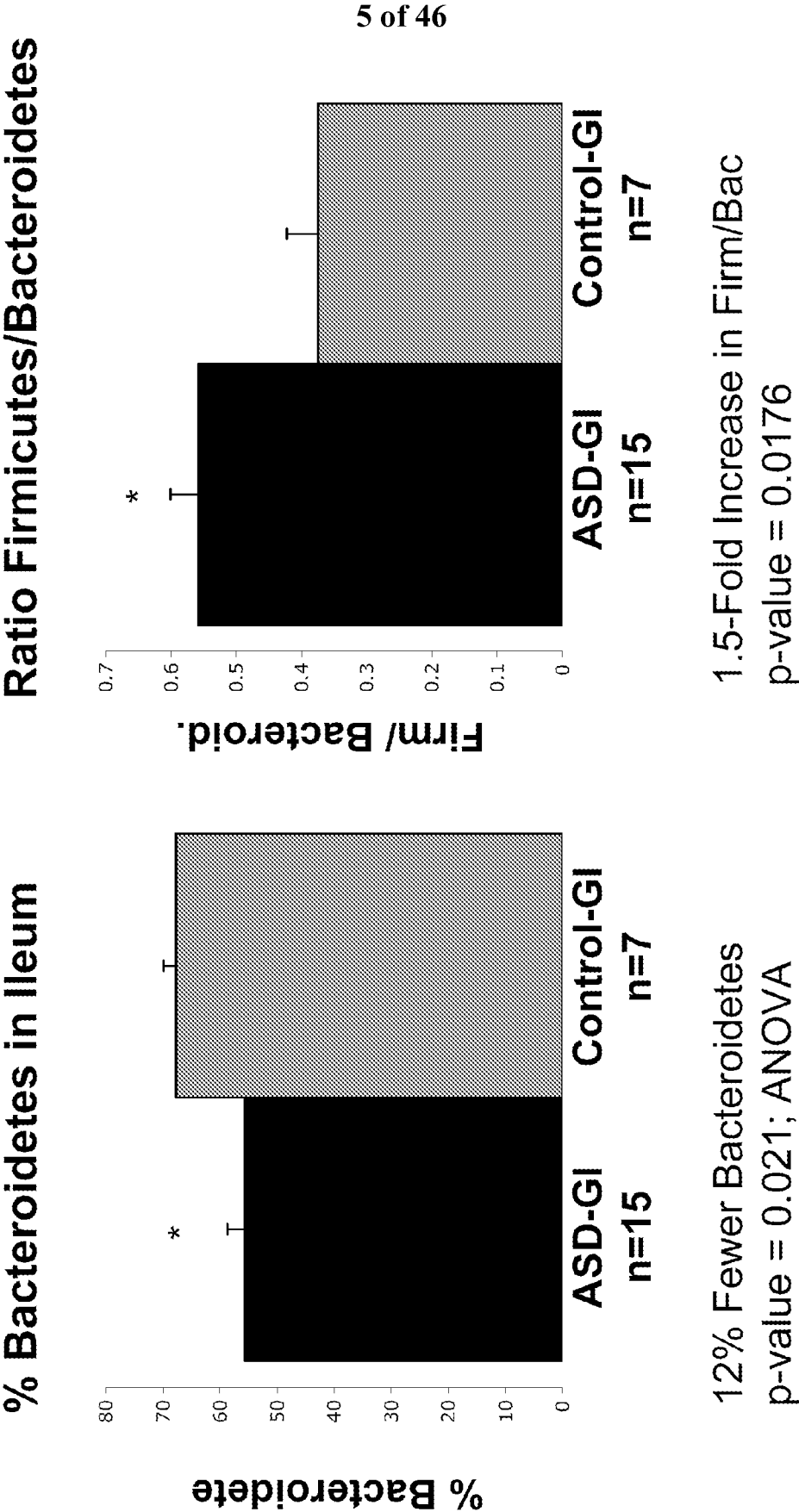
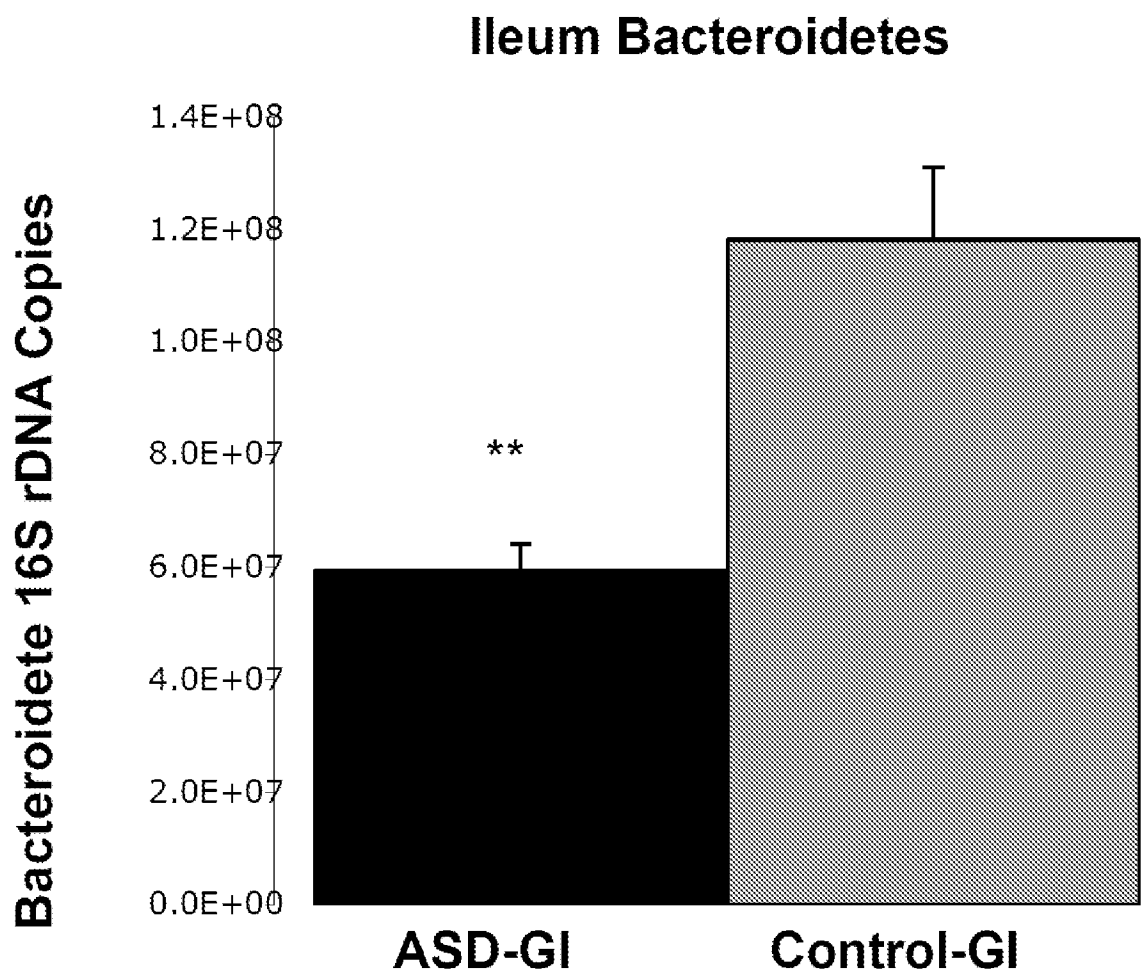


FIG. 5

6 of 46



50% Reduction in Bacteroidetes in ASD-GI
p-value < 0.0001

FIG. 6

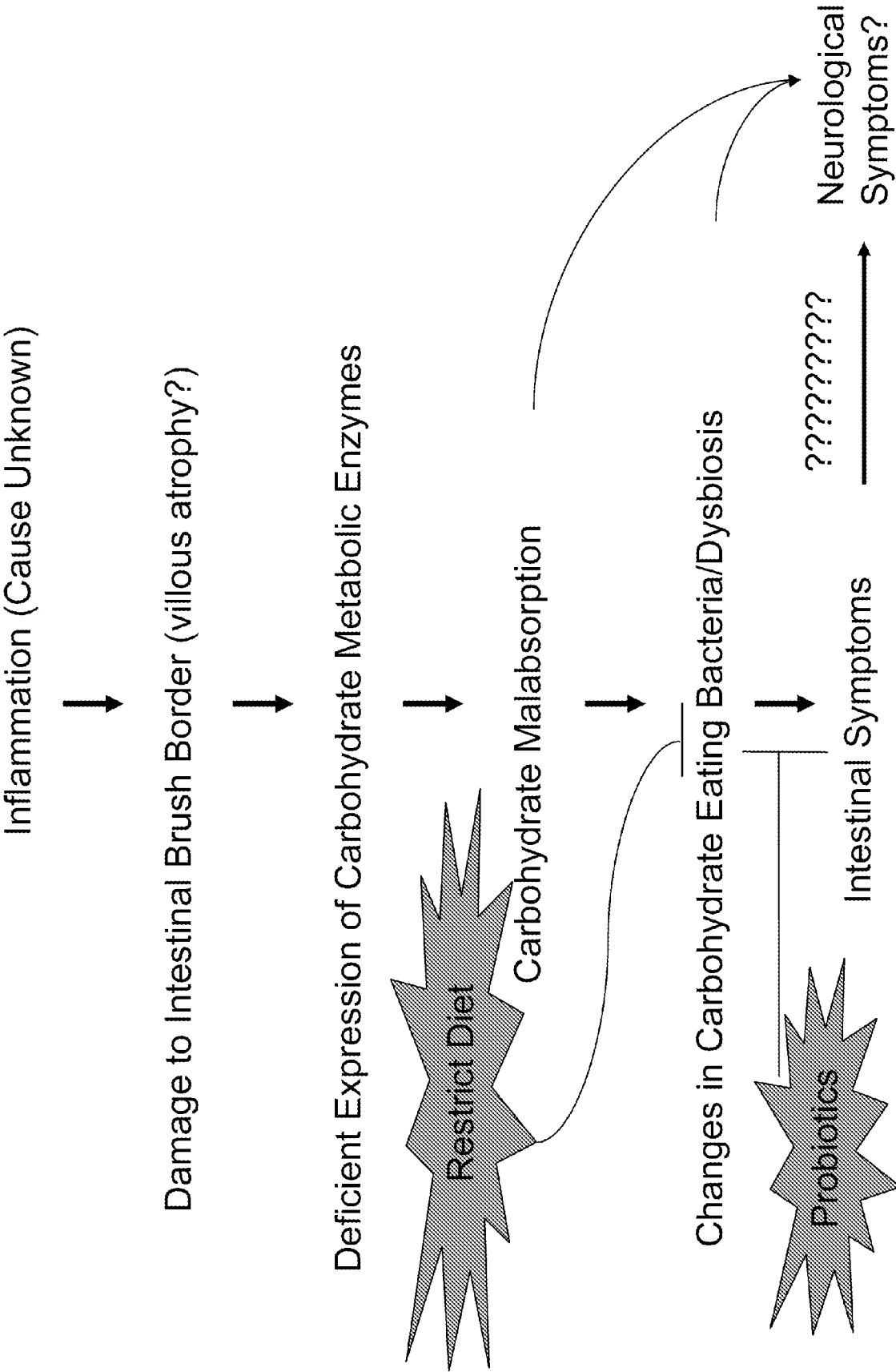
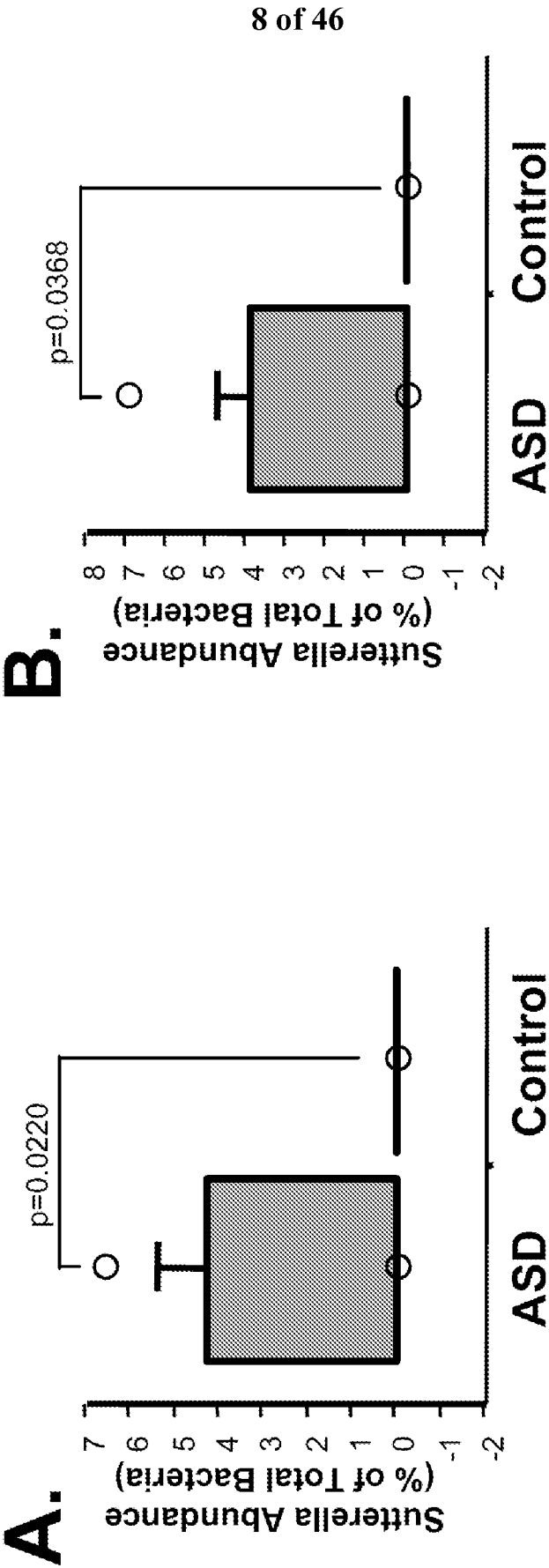
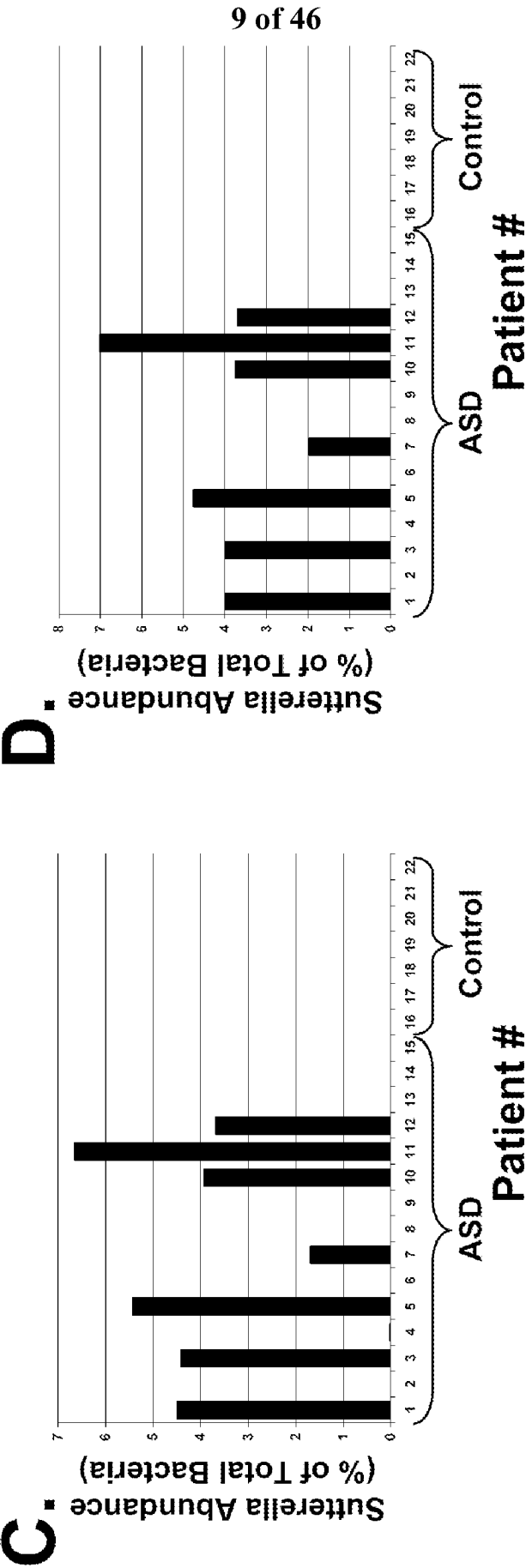


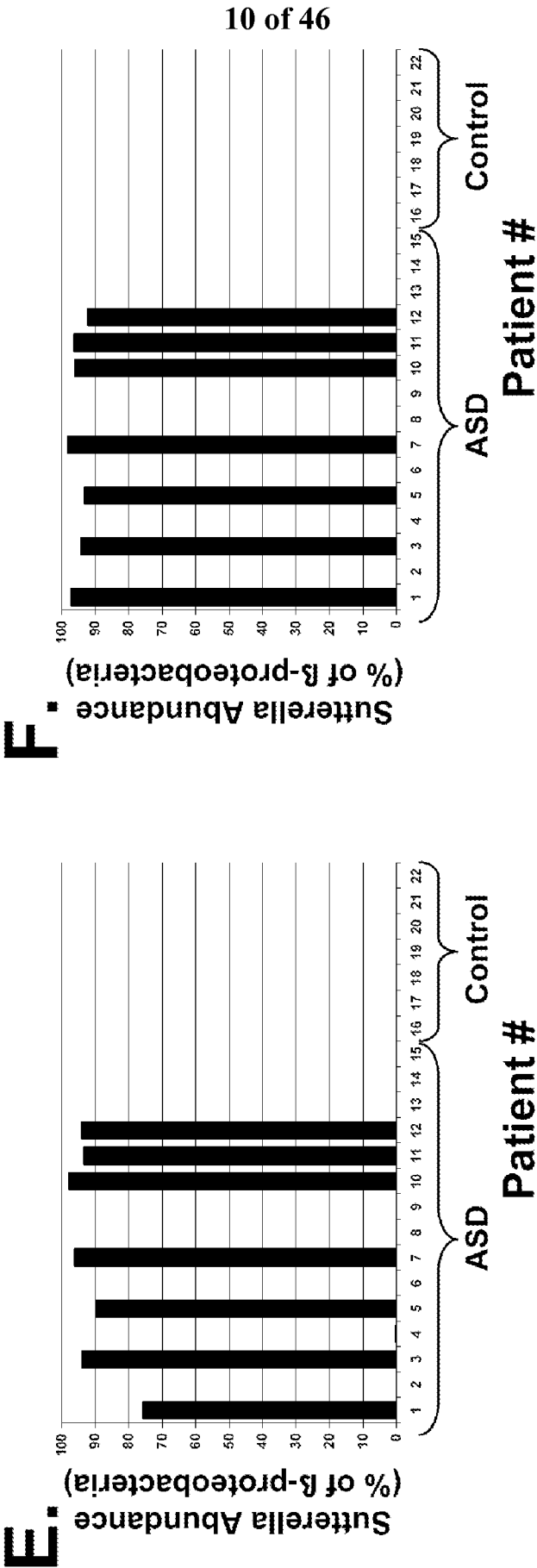
FIG. 7



FIGS. 8A-B

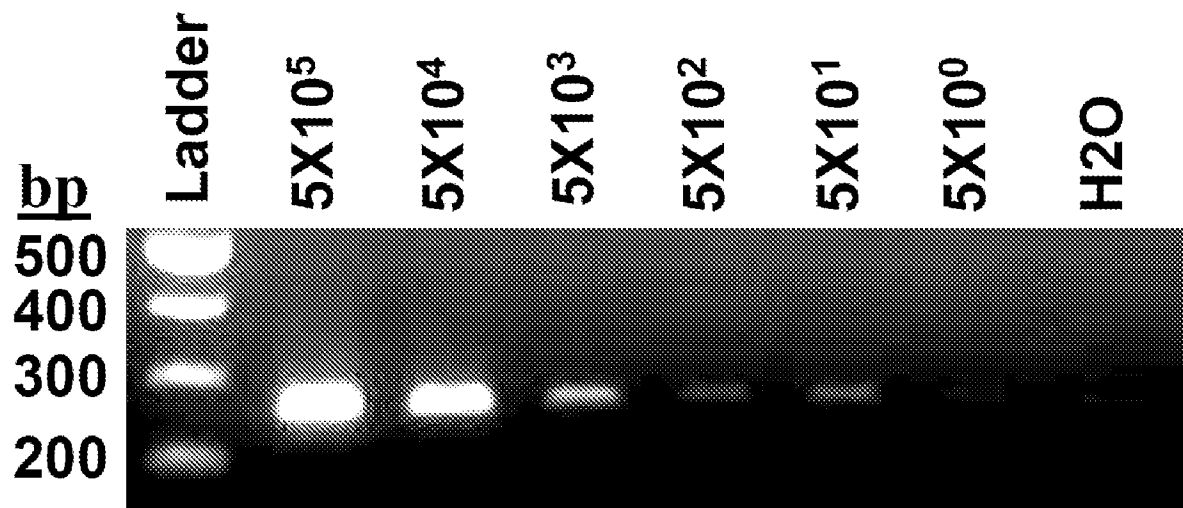
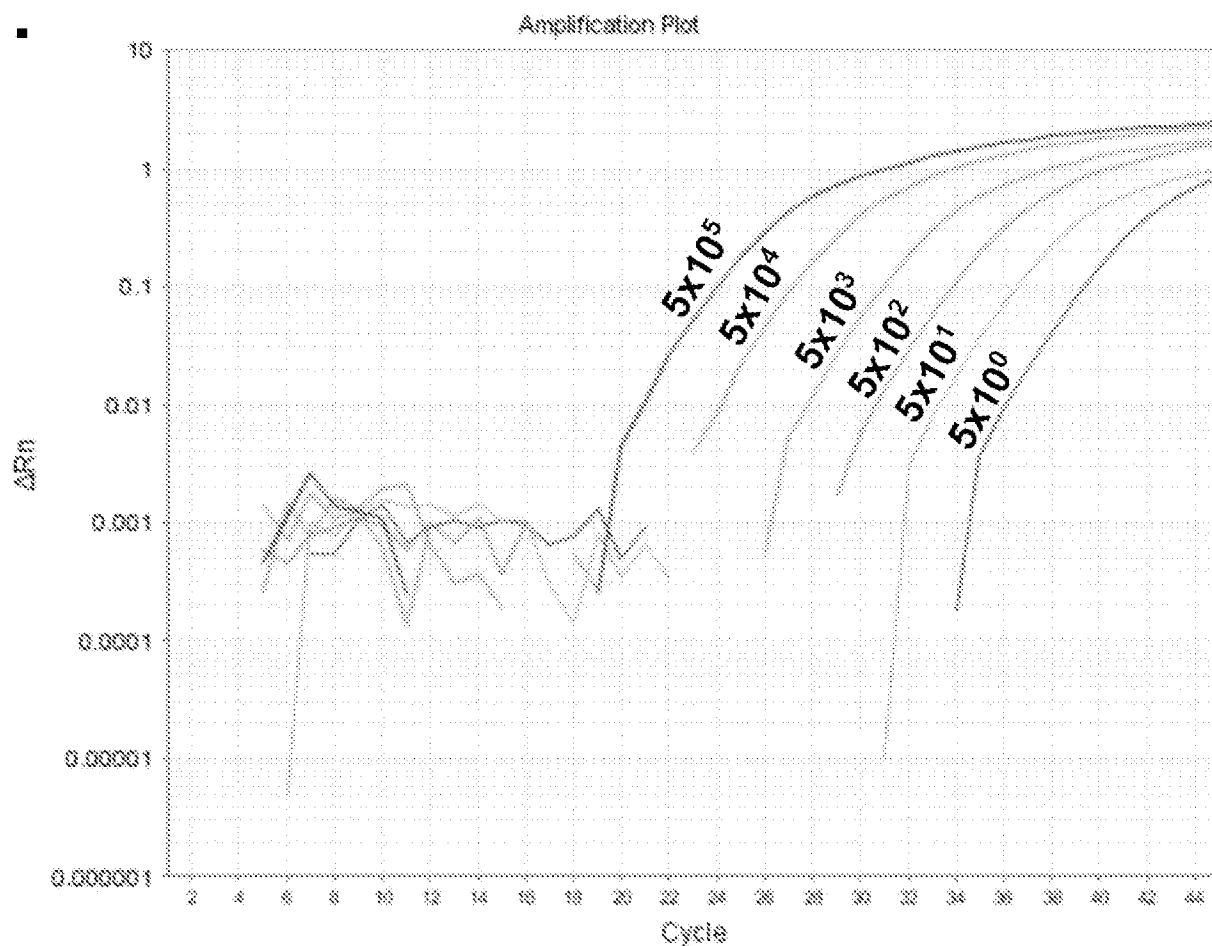


FIGS. 8C-D

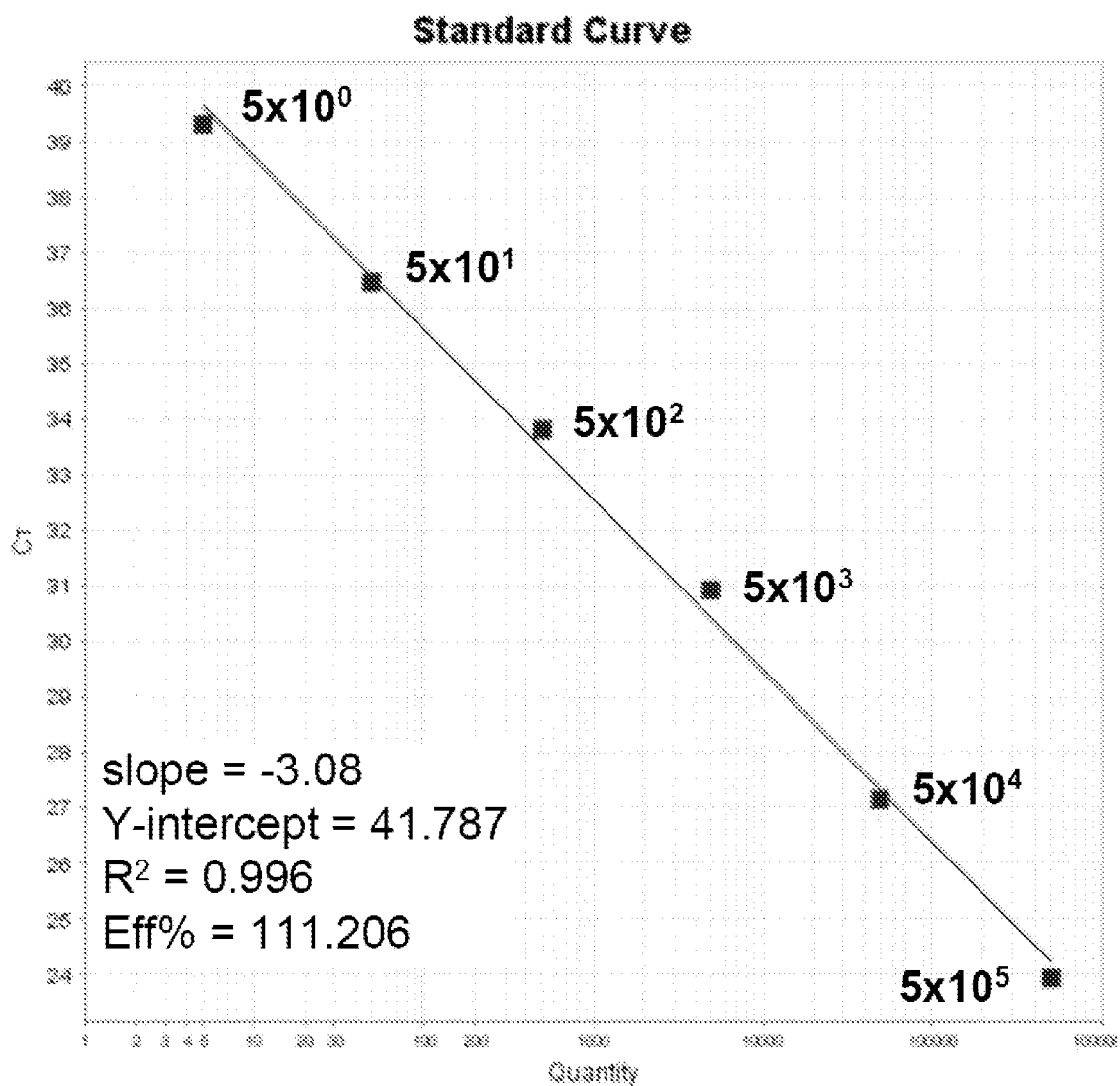


FIGS. 8E-F

11 of 46

**FIG. 9****FIG. 10A**

12 of 46

**FIG. 10B**

13 of 46

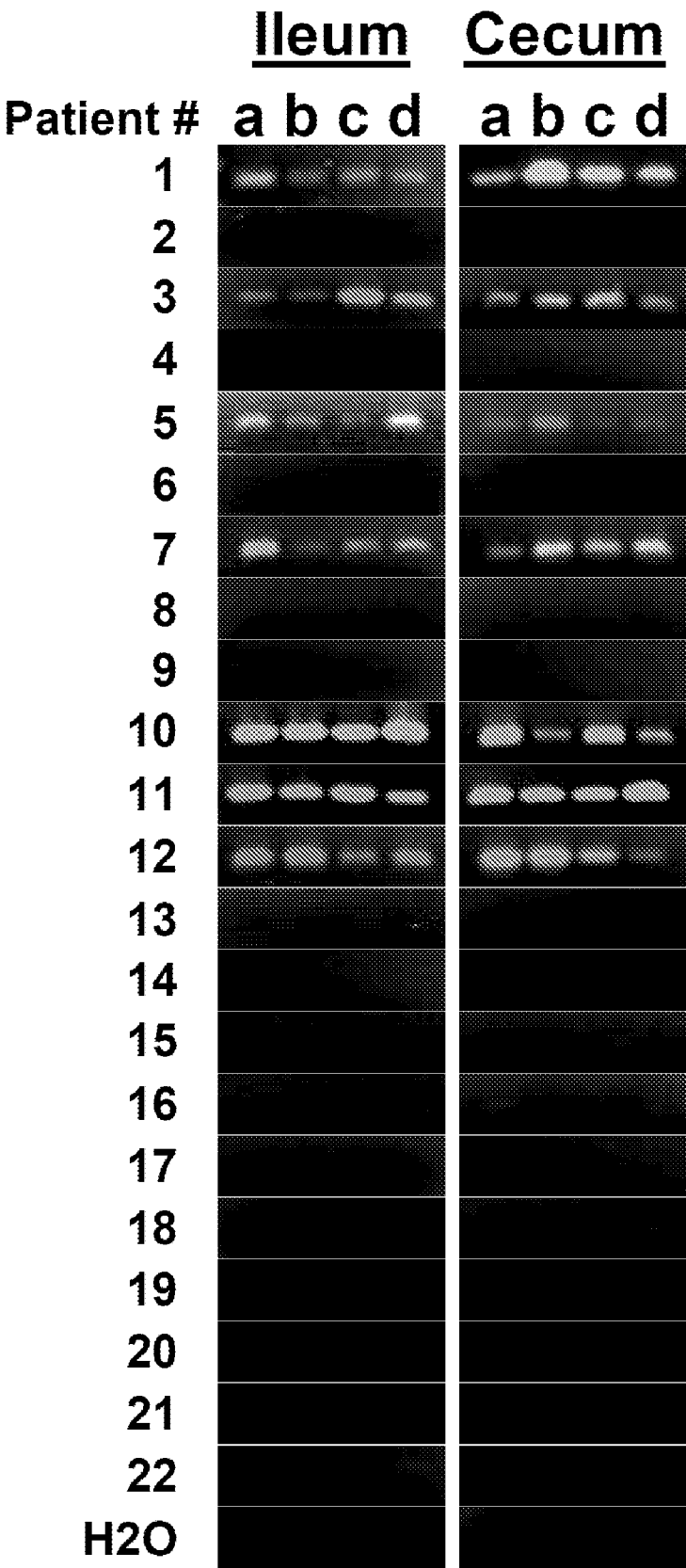
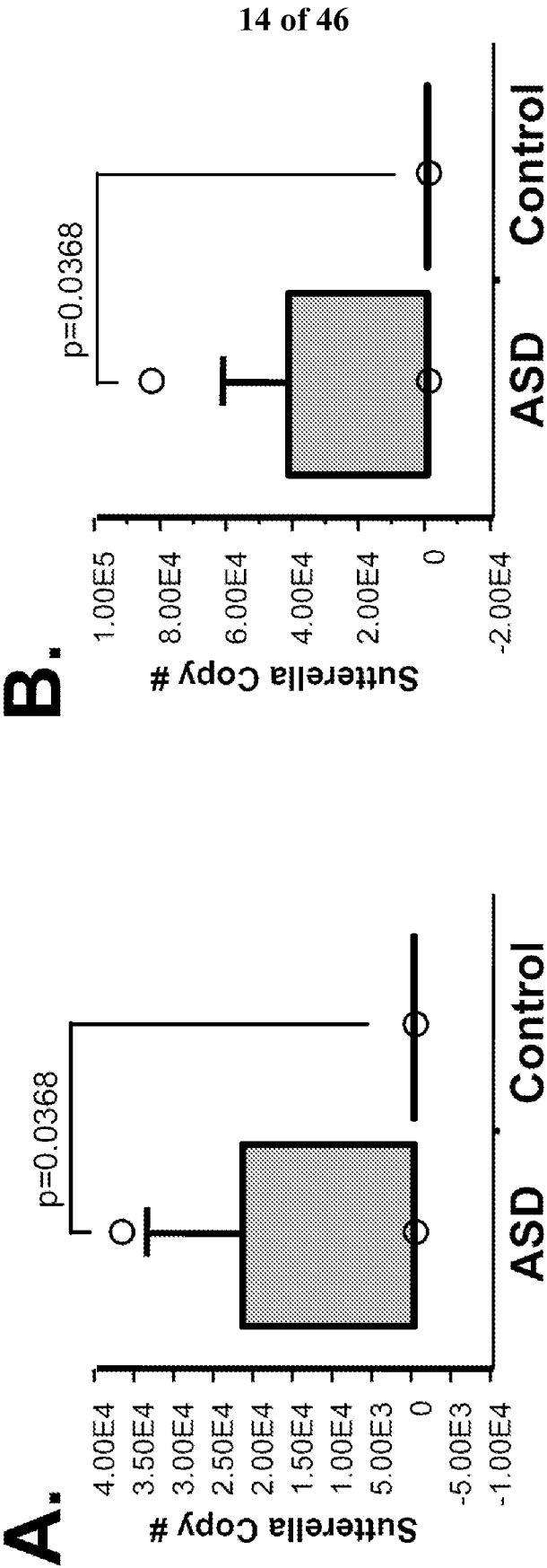
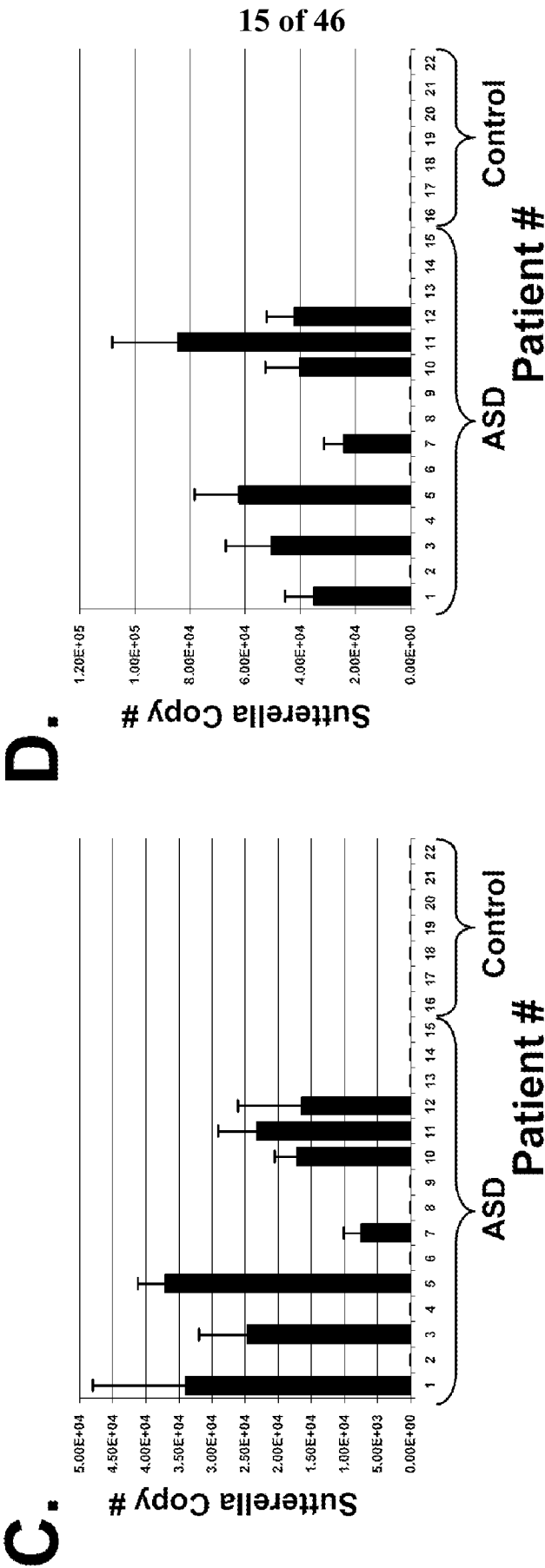


FIG. 11



FIGS. 12A-B



FIGS. 12C-D

[illegible]

FIG. 13

17 of 46



FIG. 14

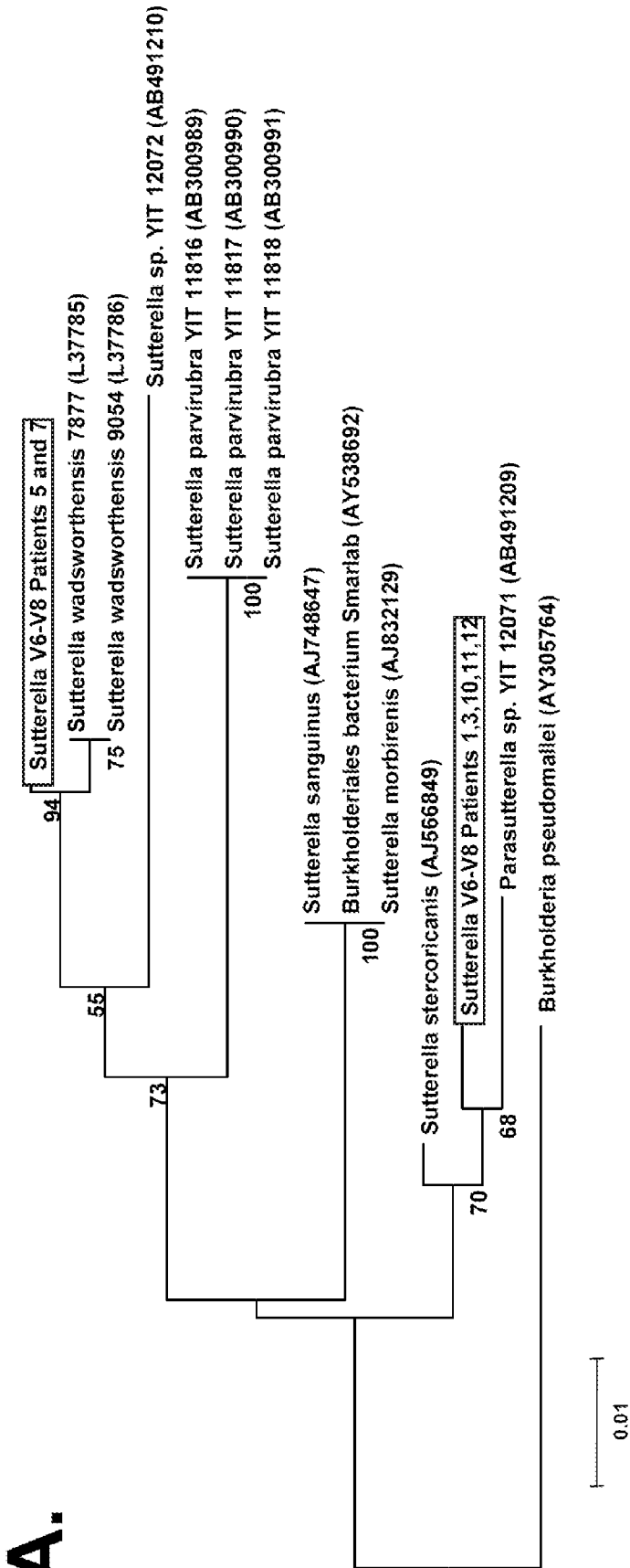


FIG. 15A

B.

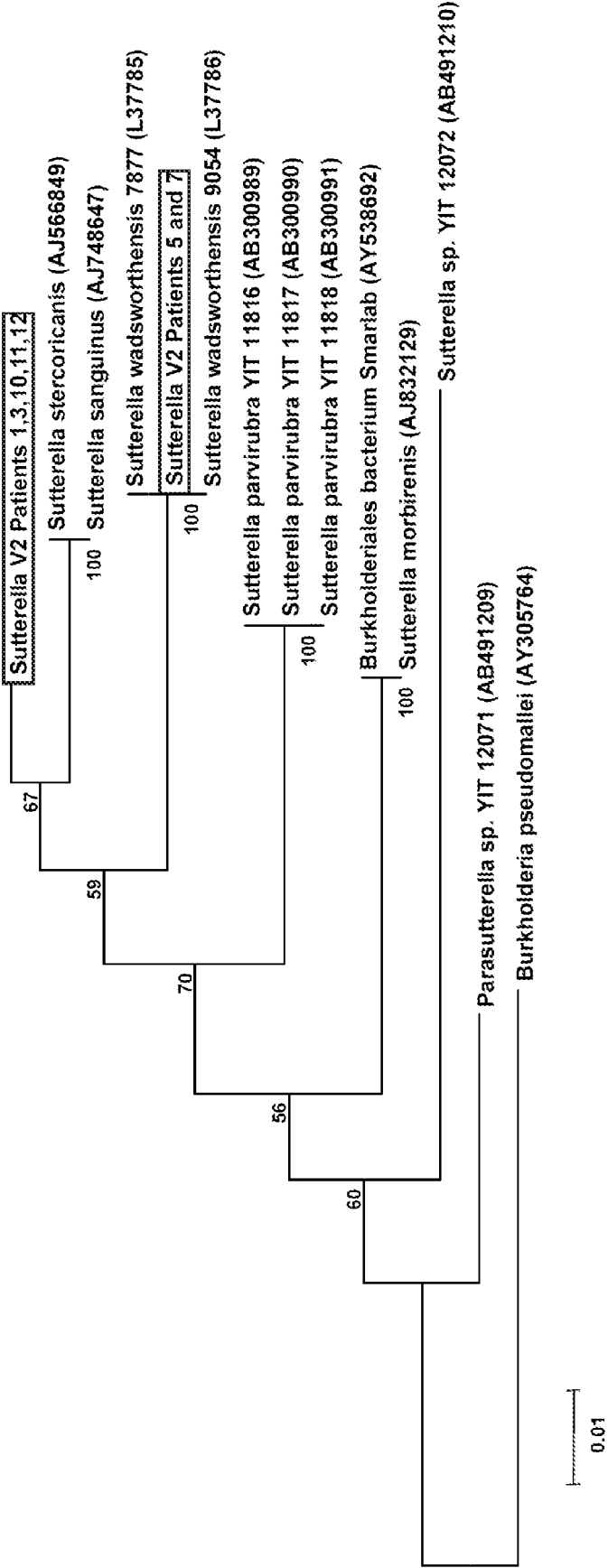
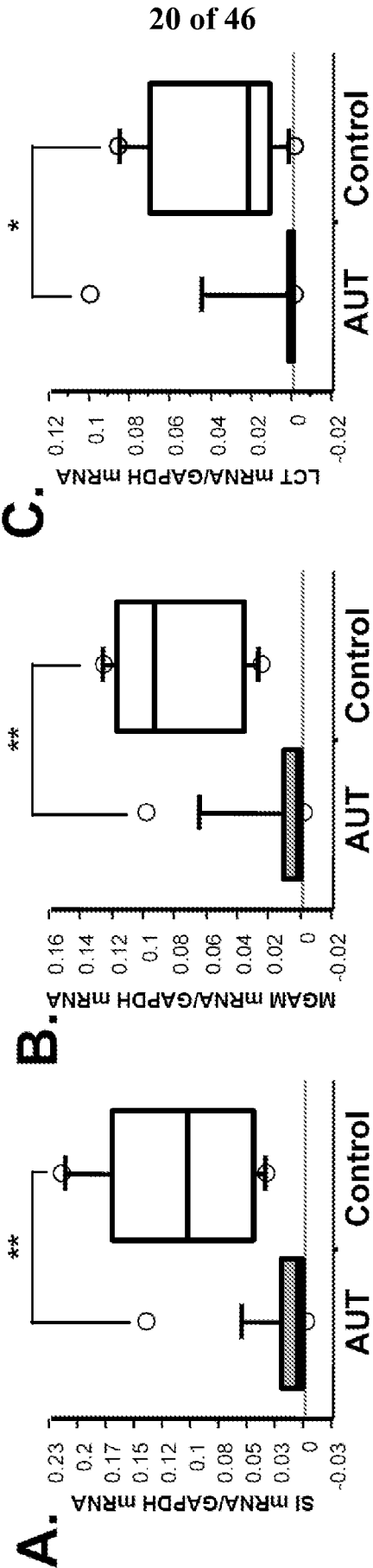
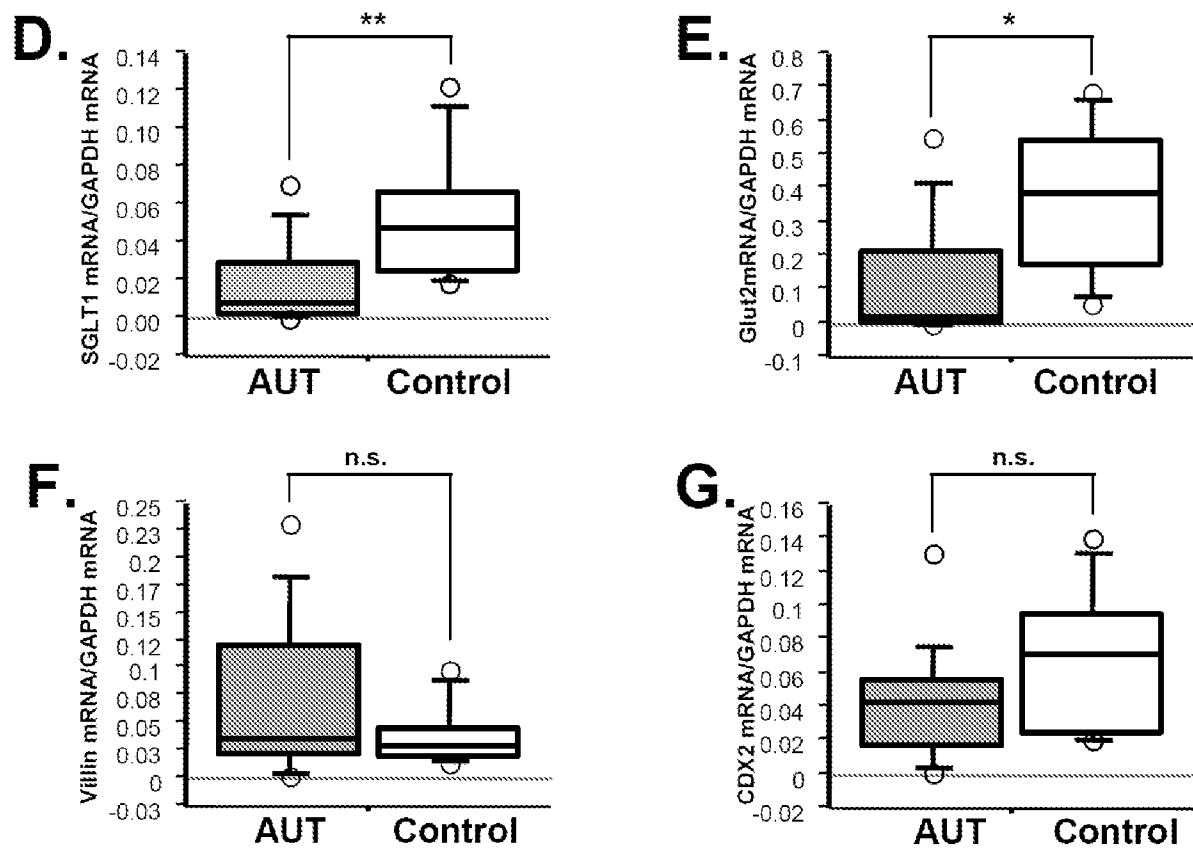


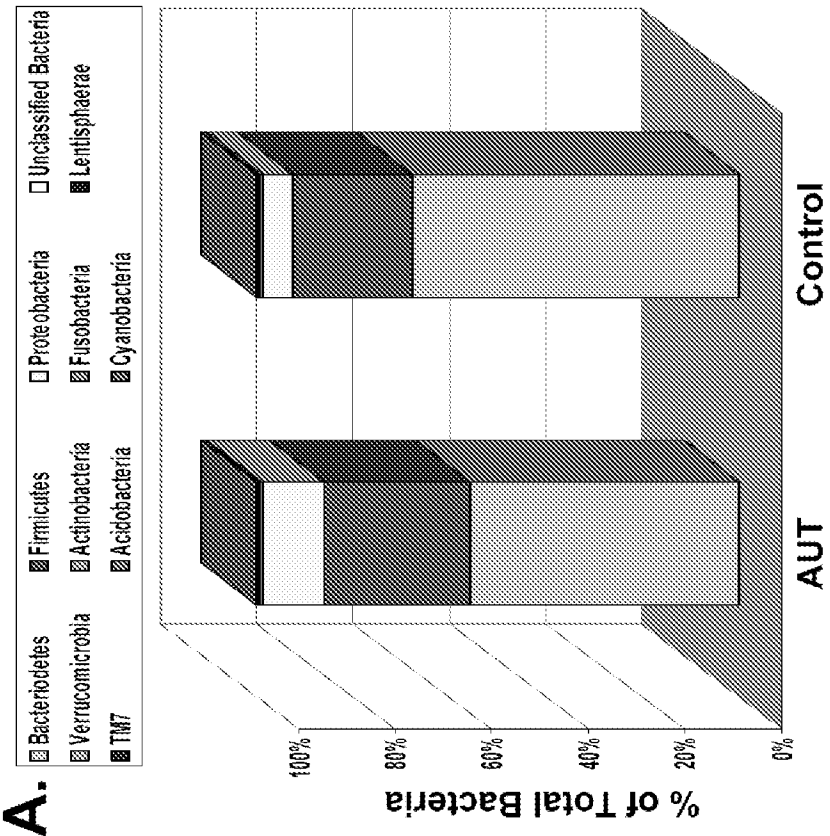
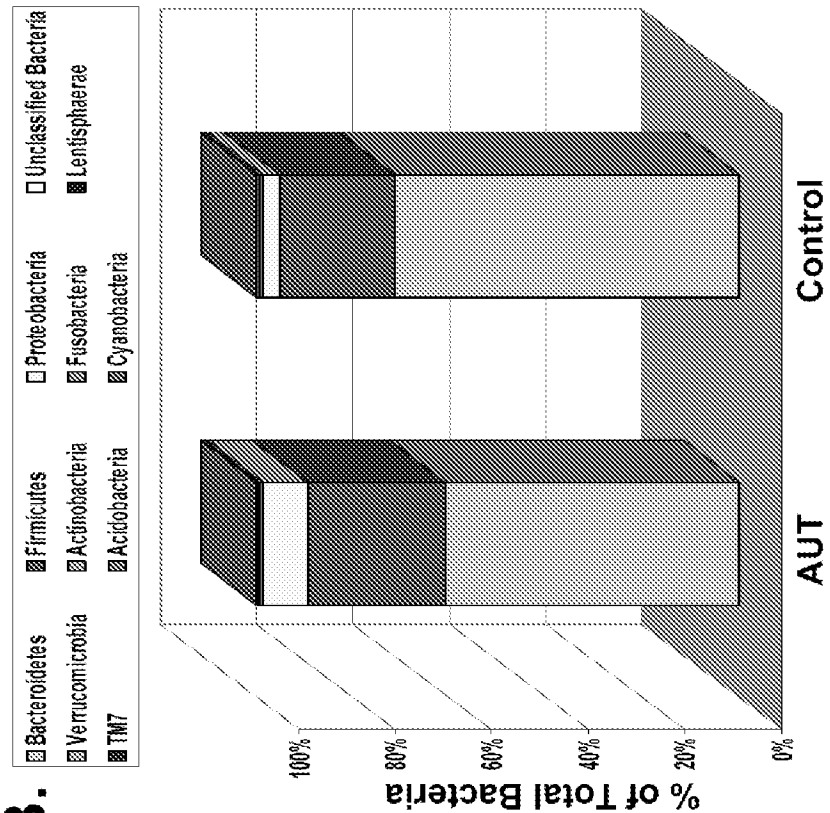
FIG. 15B



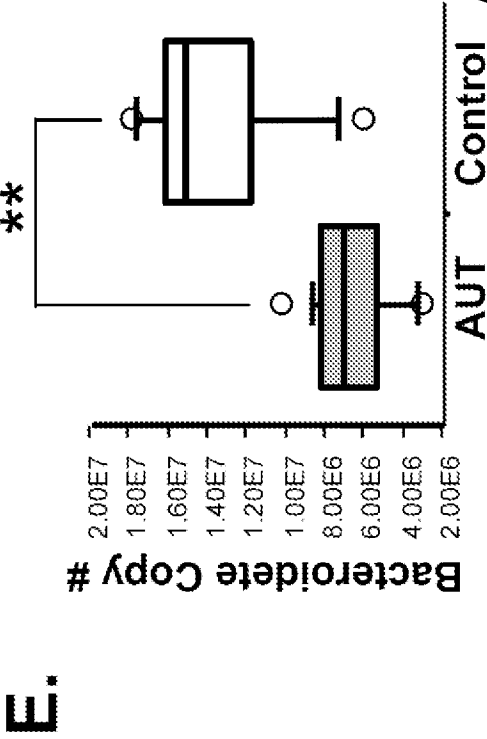
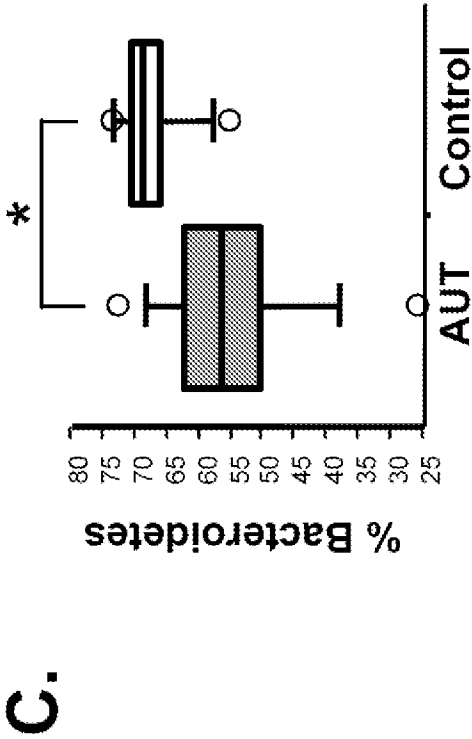
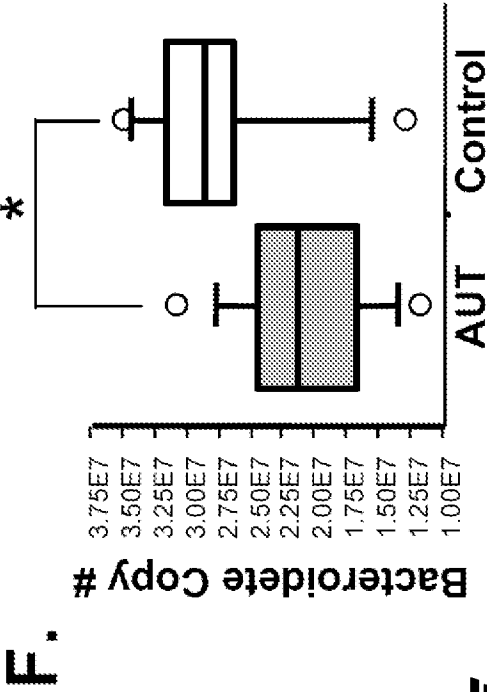
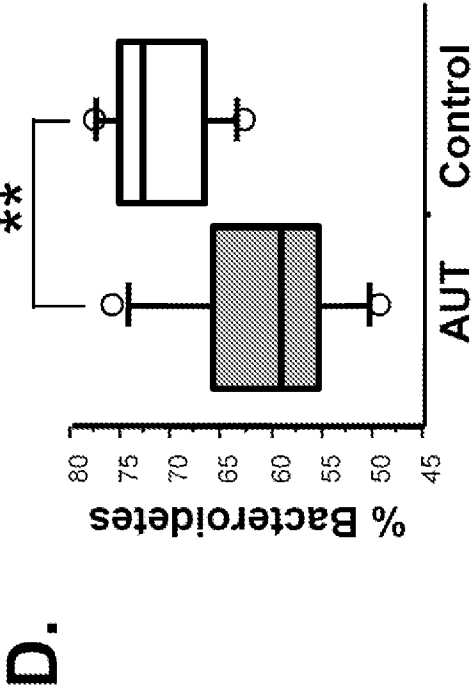
FIGS. 16A-C

21 of 46

**FIGS. 16D-G**

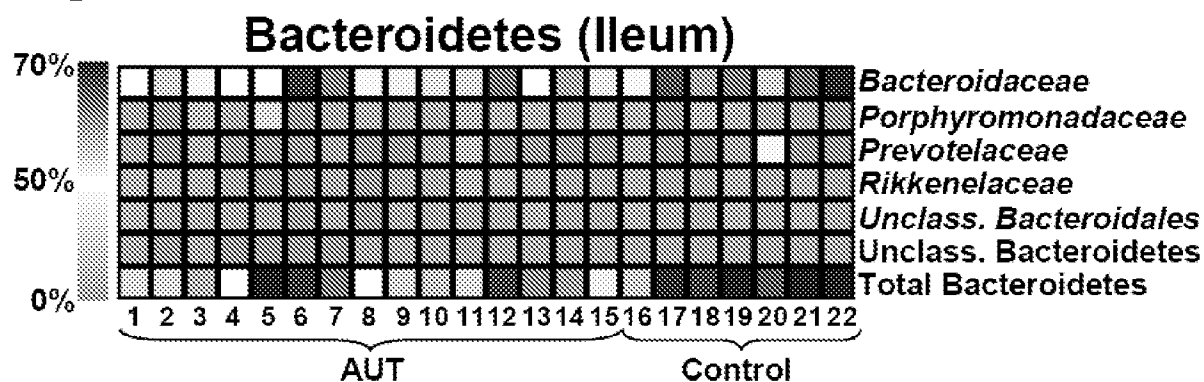
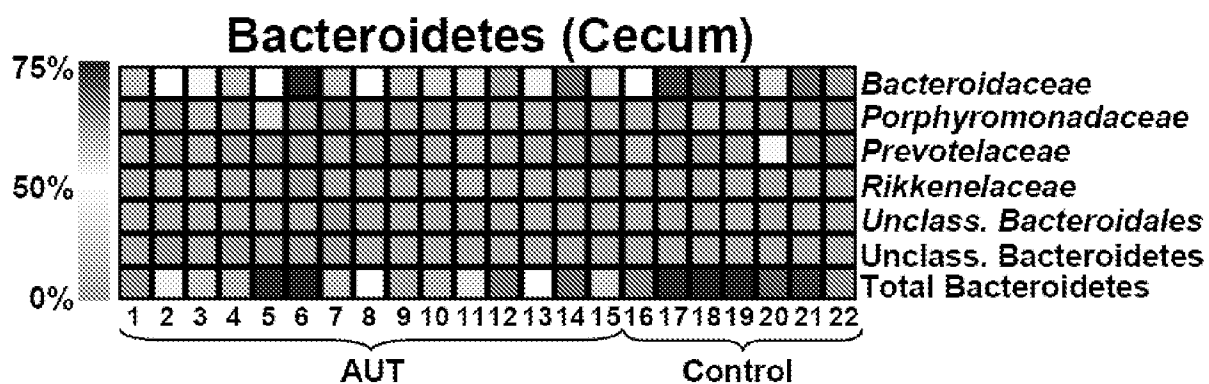


FIGS. 17A-B

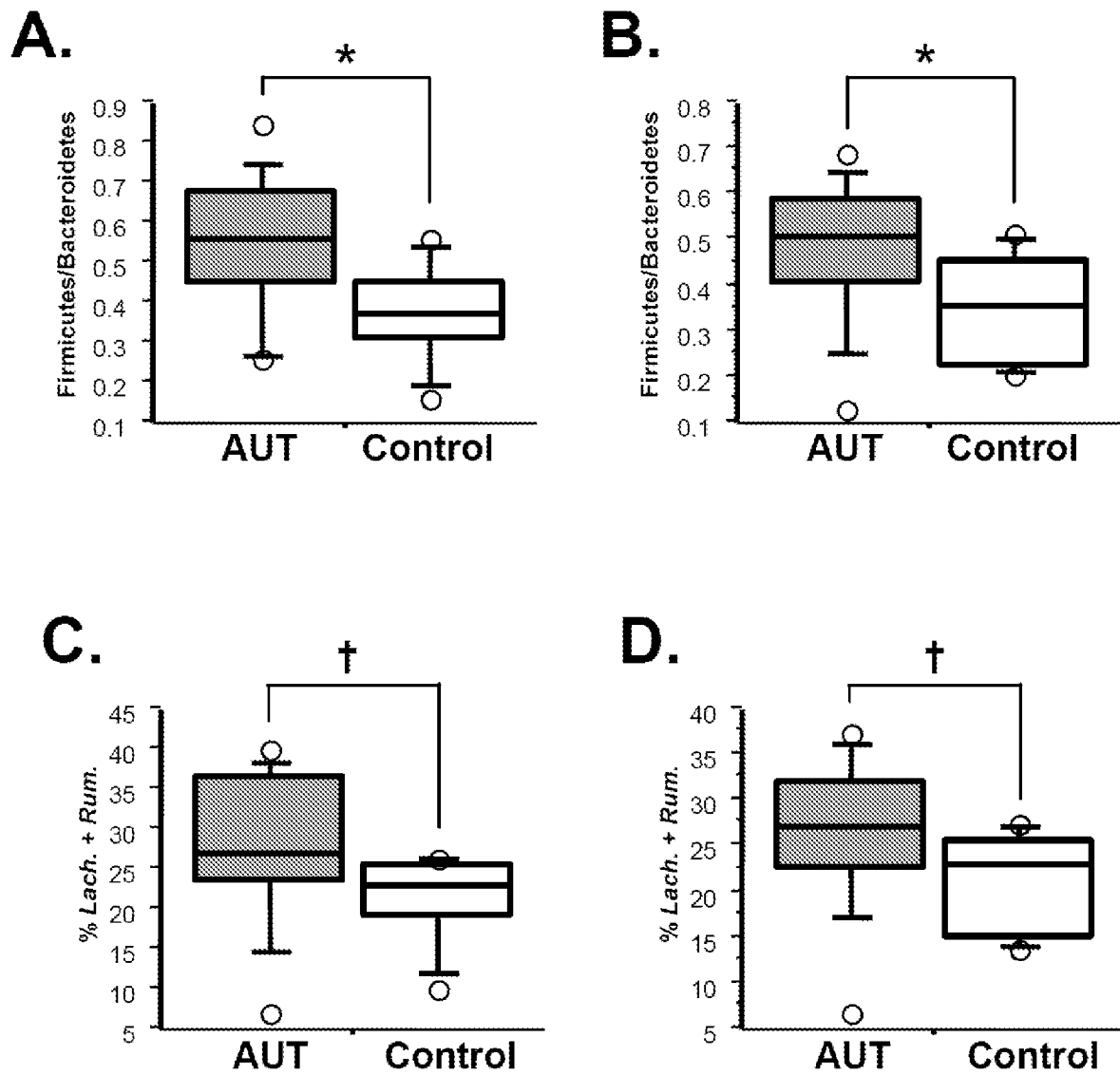


FIGS. 17C-F

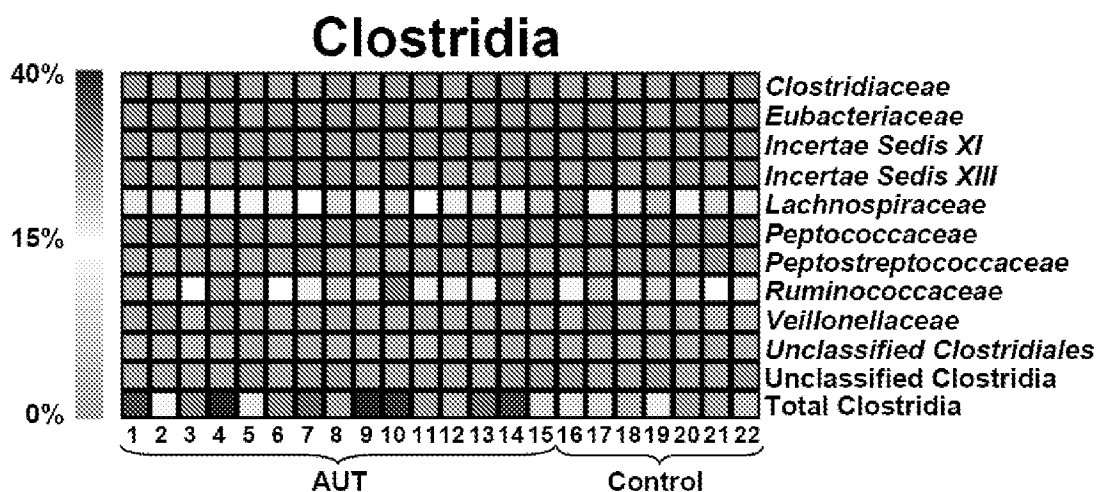
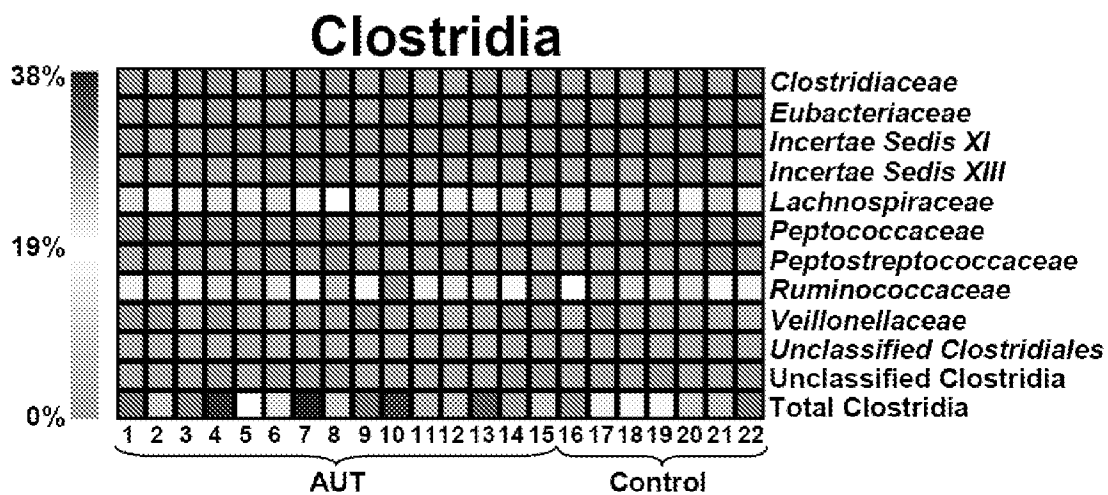
24 of 46

G.**H.****FIGS. 17G-H**

25 of 46

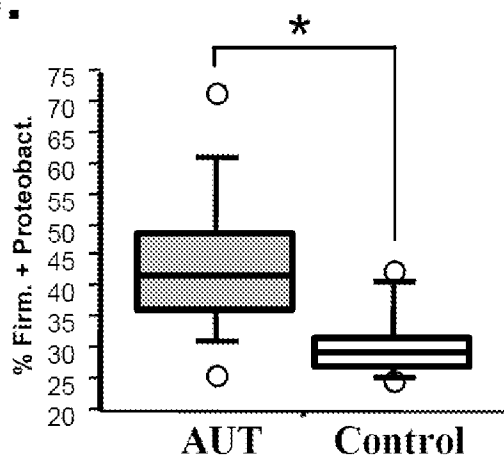
**FIGS. 18A-D**

26 of 46

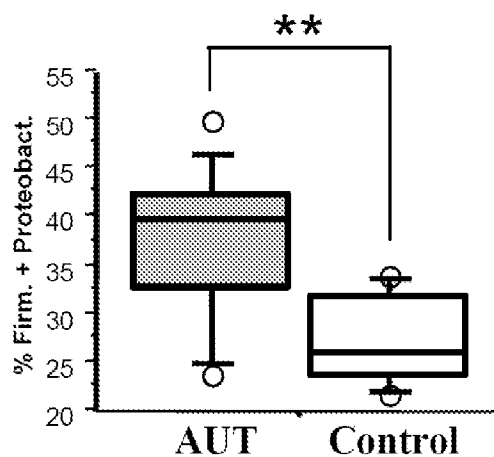
E. Ileum**F. Cecum****FIGS. 18E-F**

27 of 46

G.

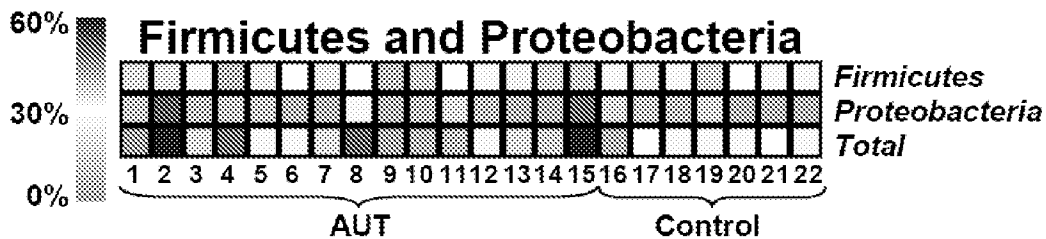


H.



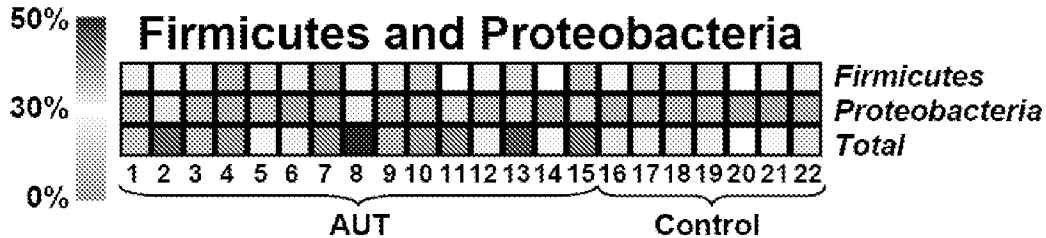
Ileum

I.



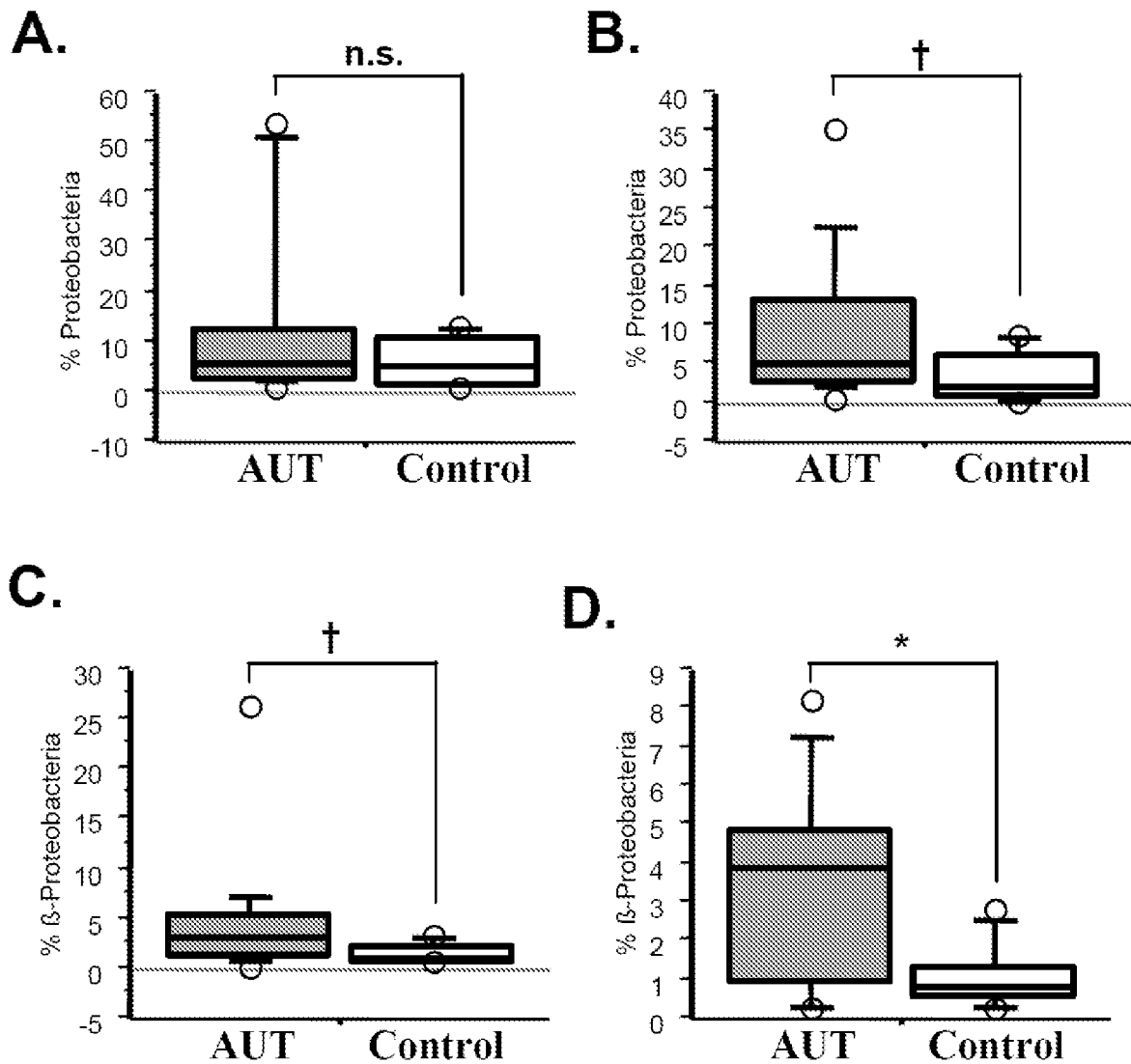
Cecum

J.



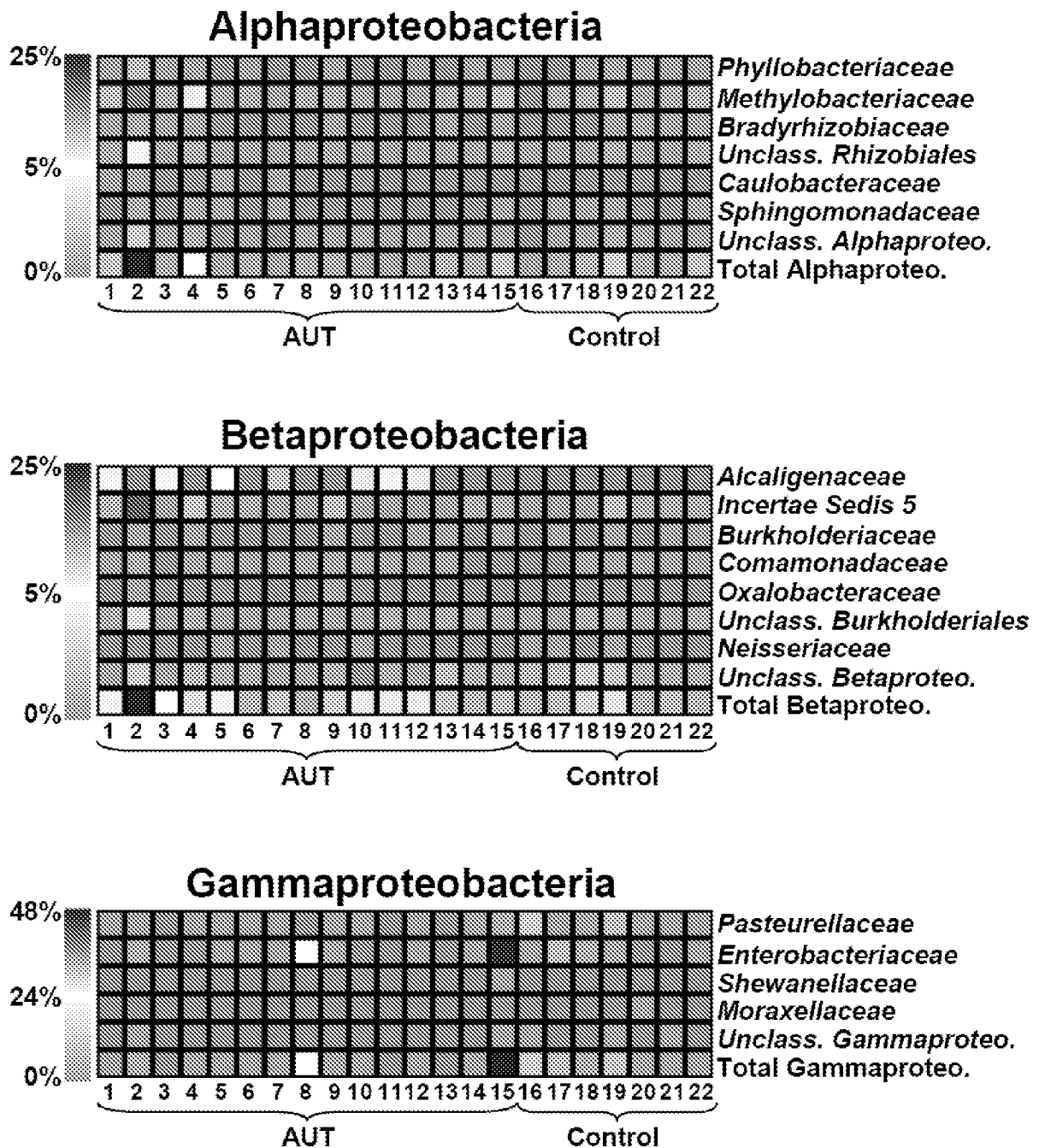
FIGS. 18G-J

28 of 46

**FIGS. 19A-D**

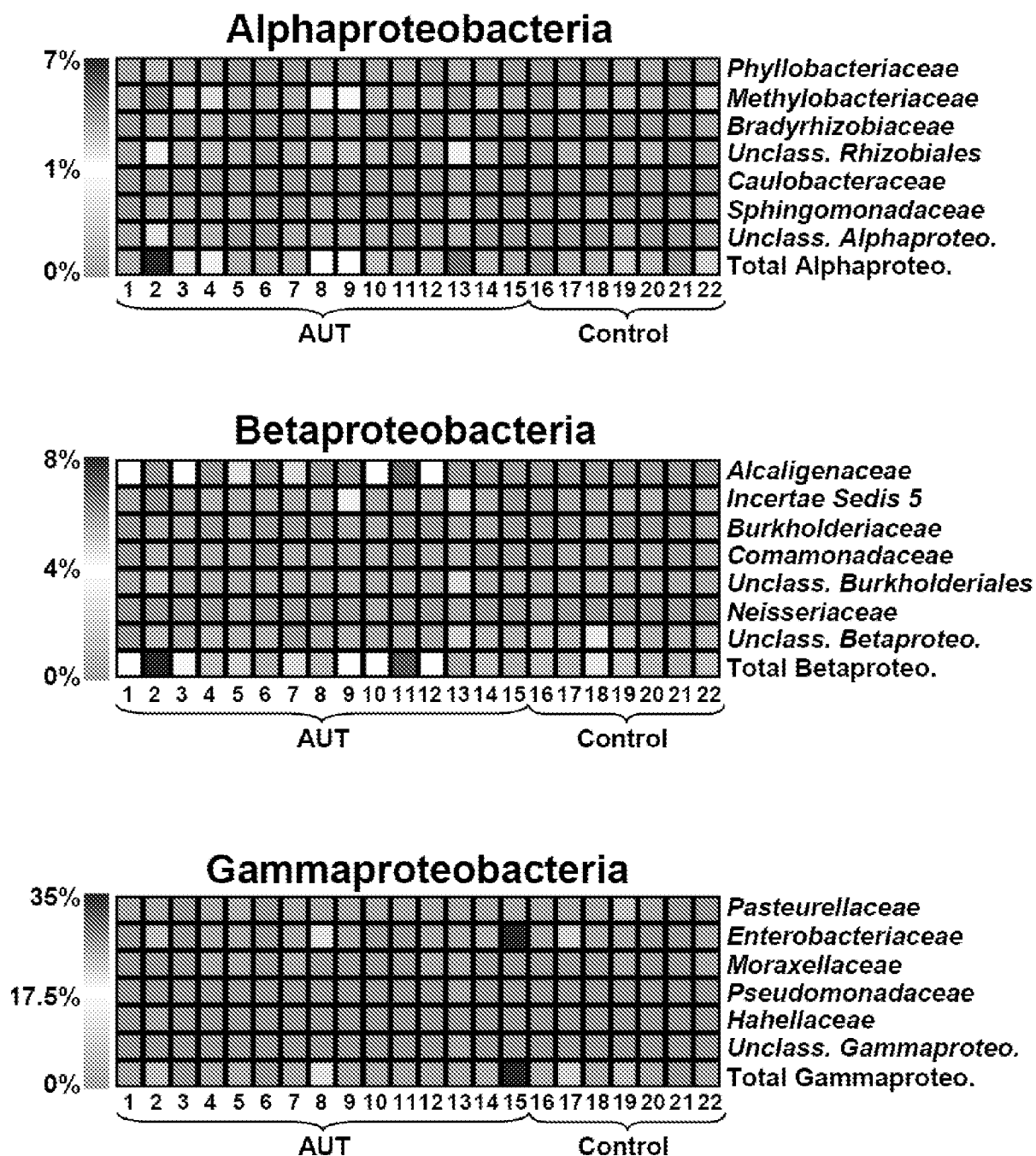
29 of 46

E. Ileum

**FIG. 19E**

30 of 46

F. Cecum

**FIG. 19F**

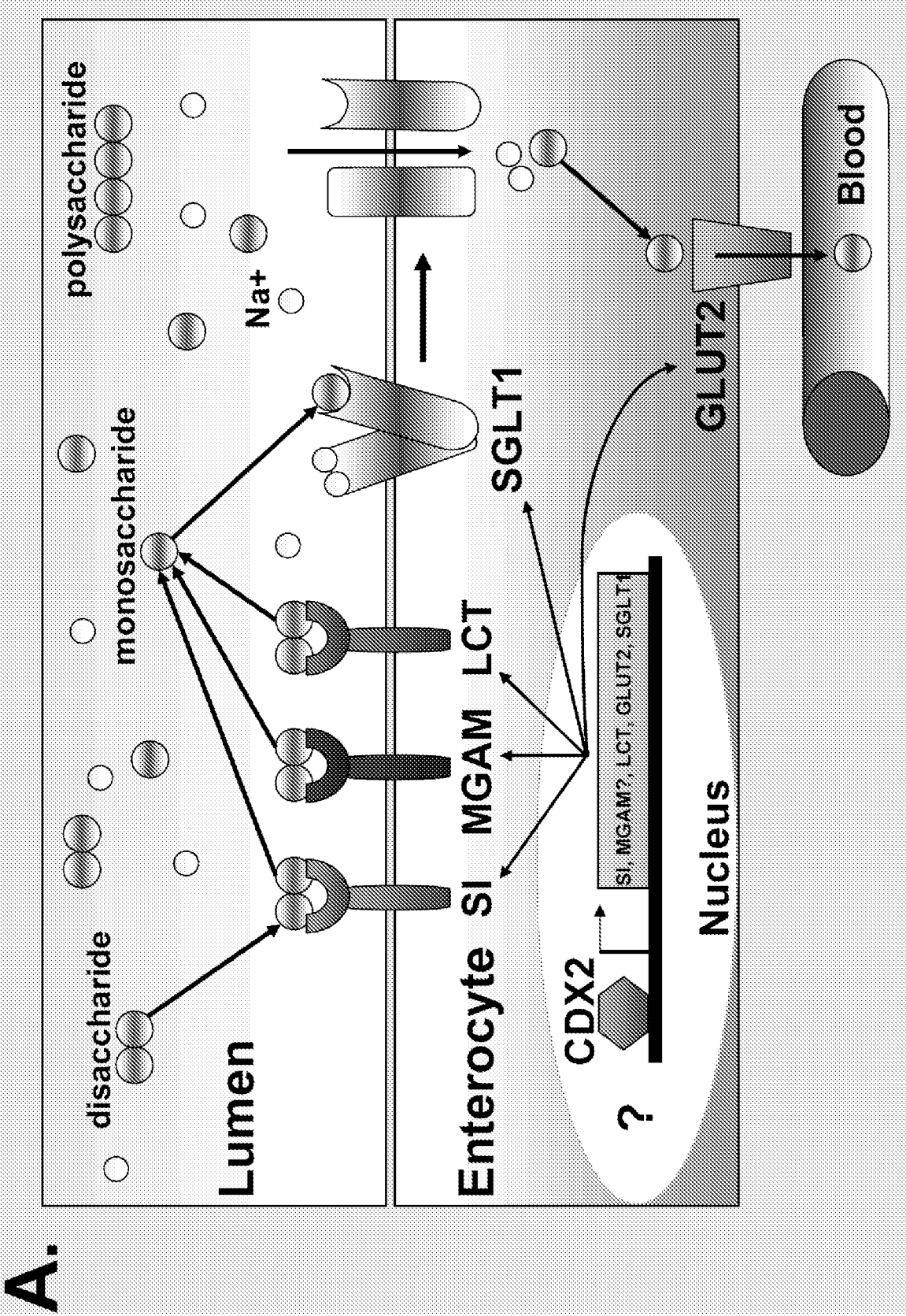
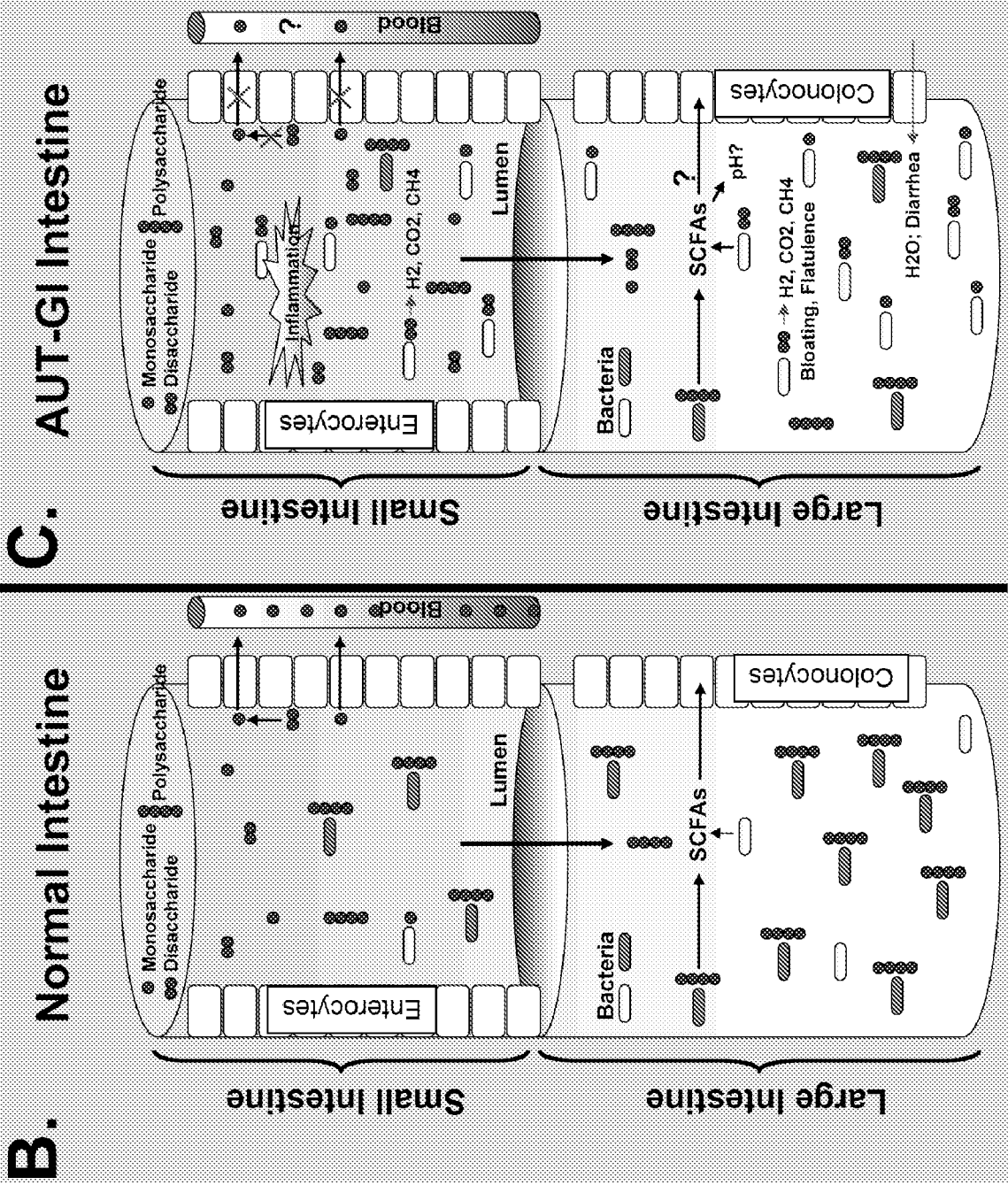
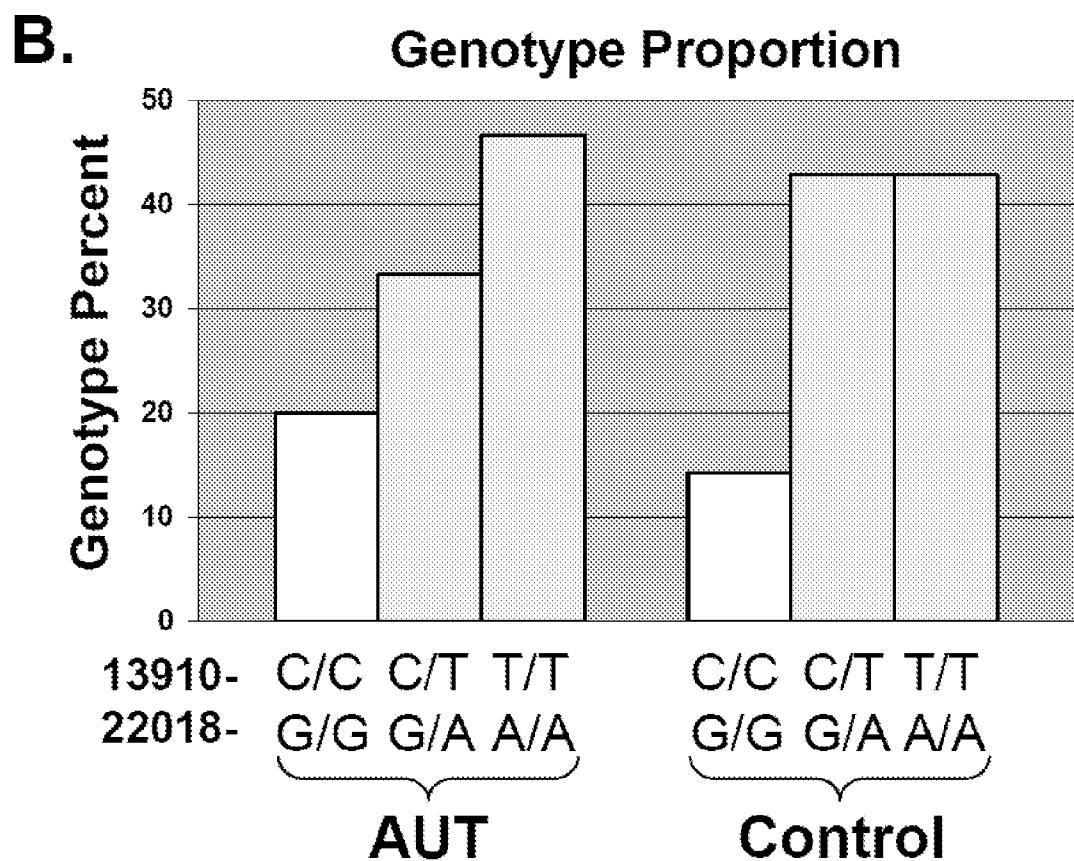
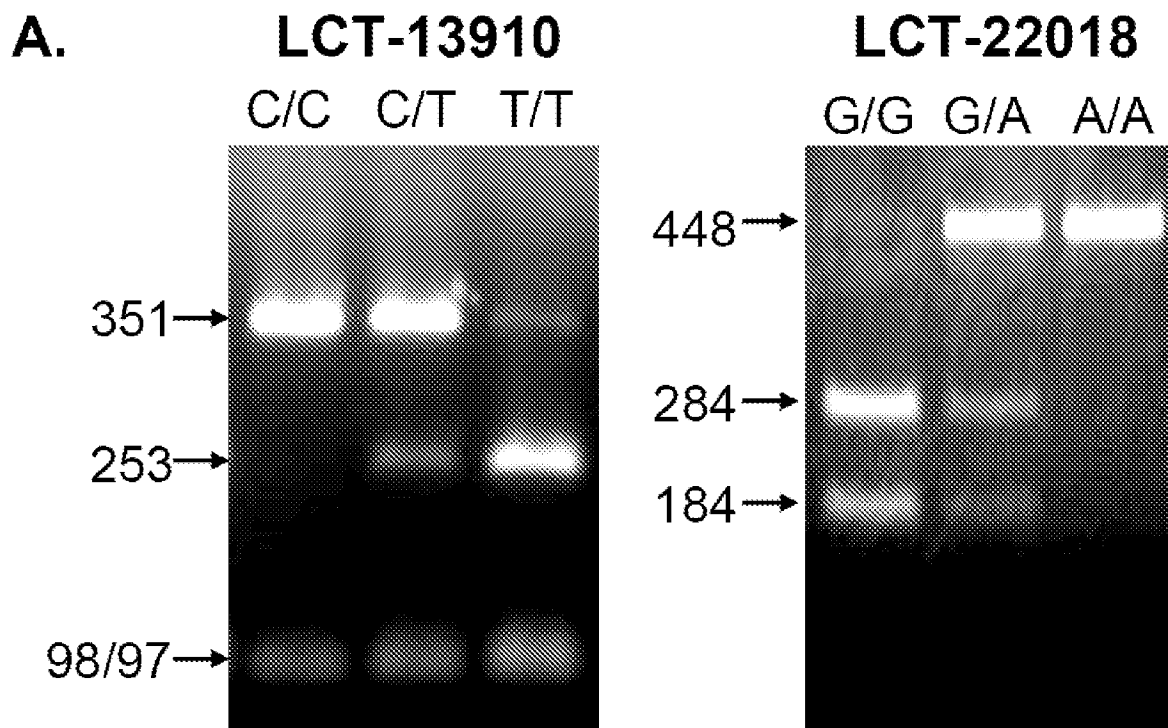
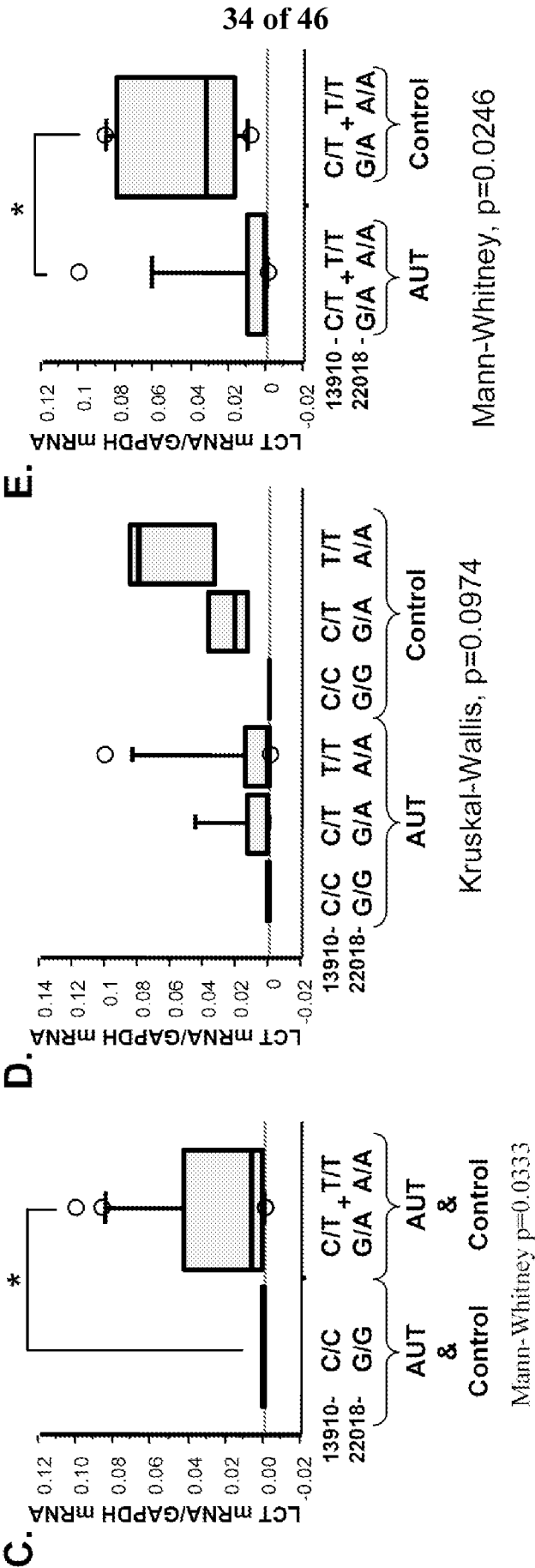


FIG. 20A



33 of 46

**FIGS. 21A-B**



FIGS. 21C-E

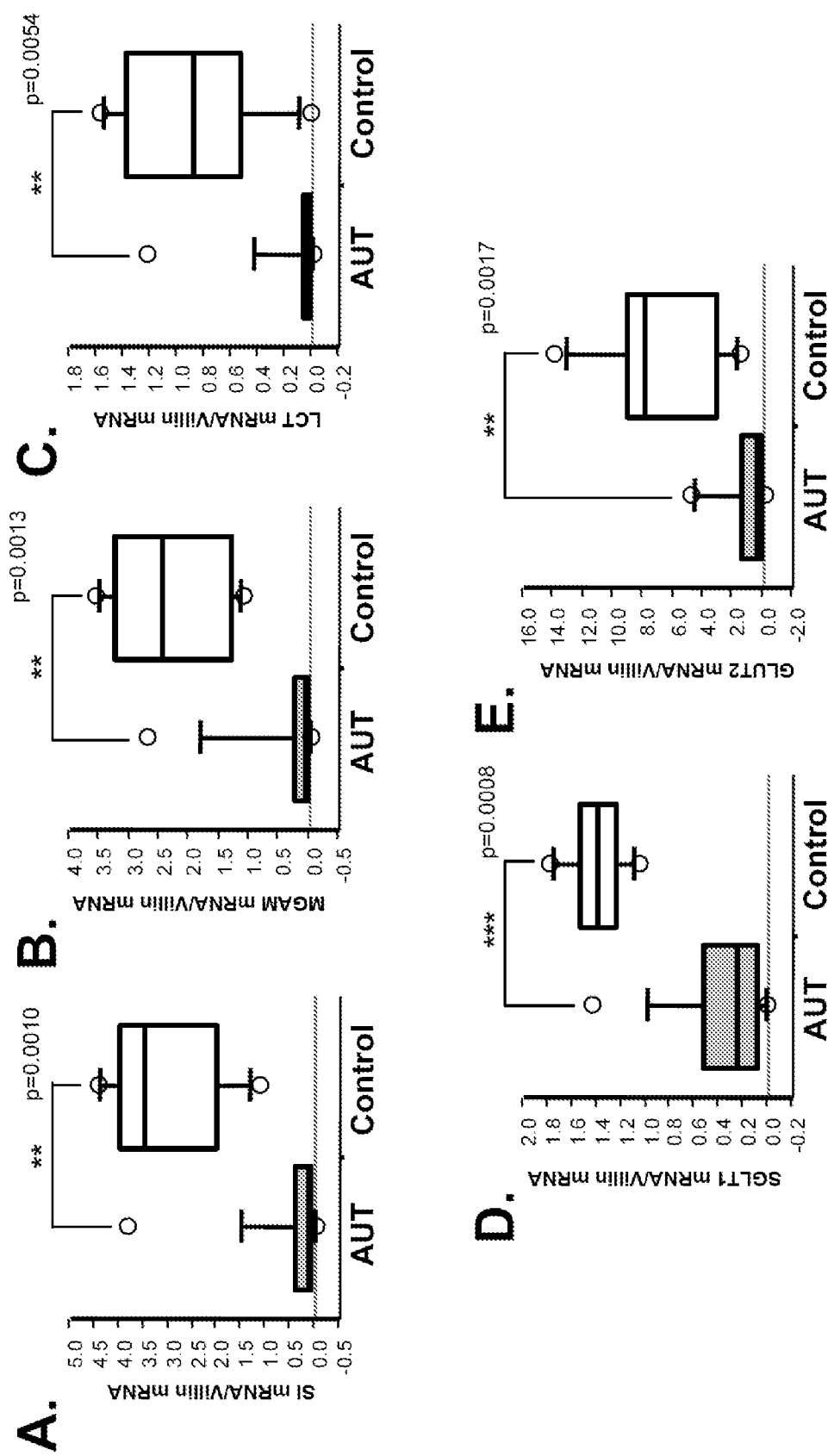
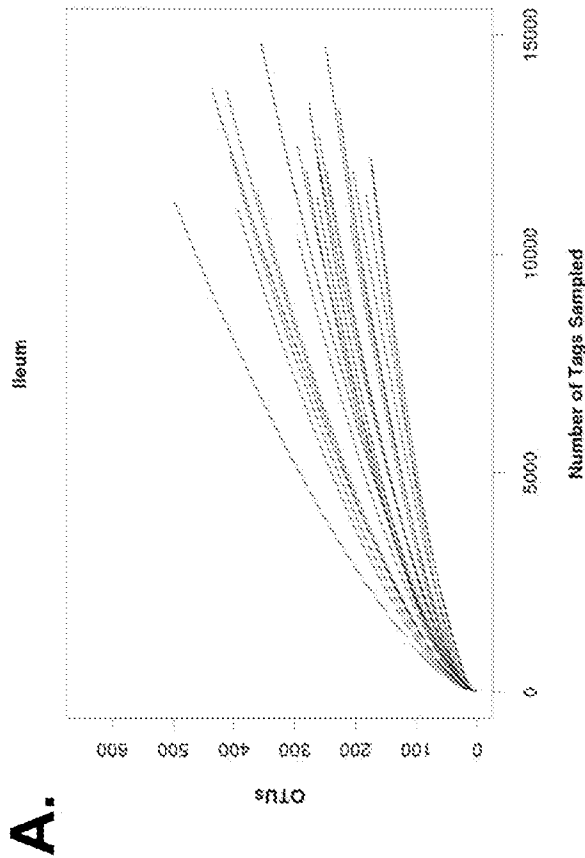
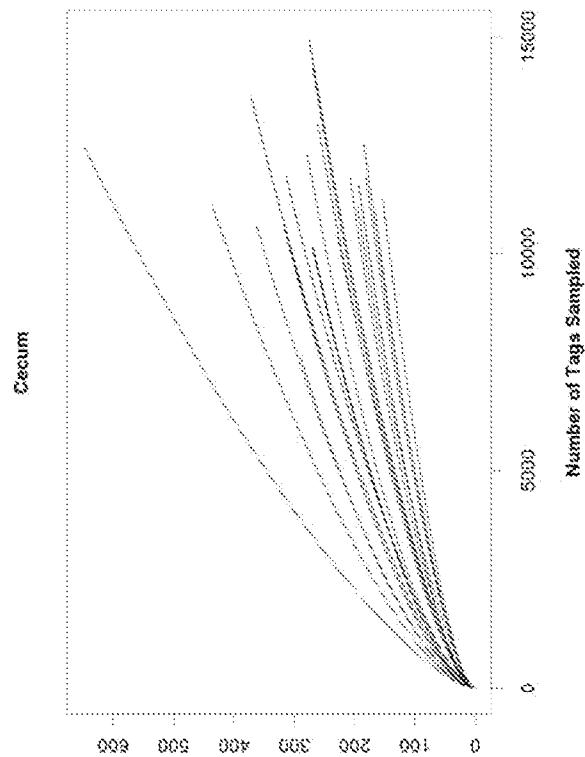
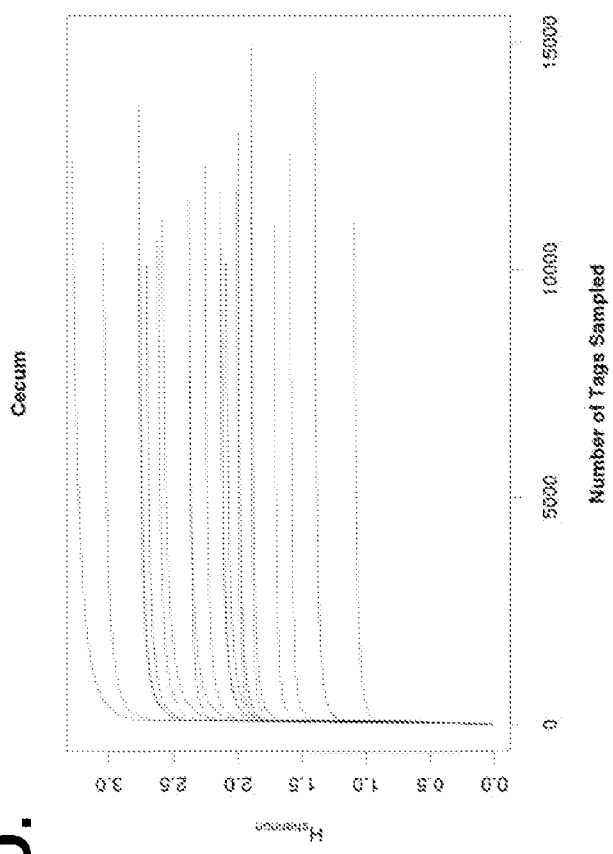


FIG. 22

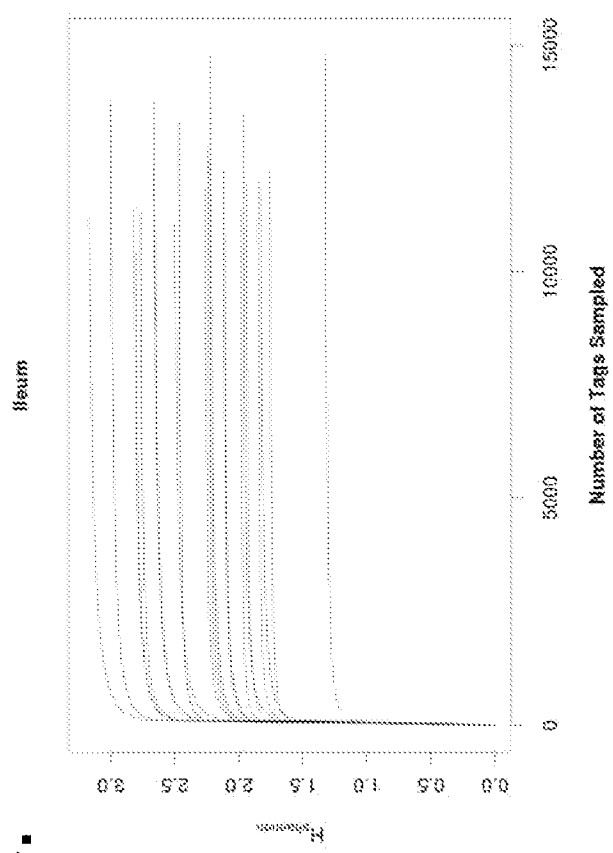


FIGS. 23A-B

D.

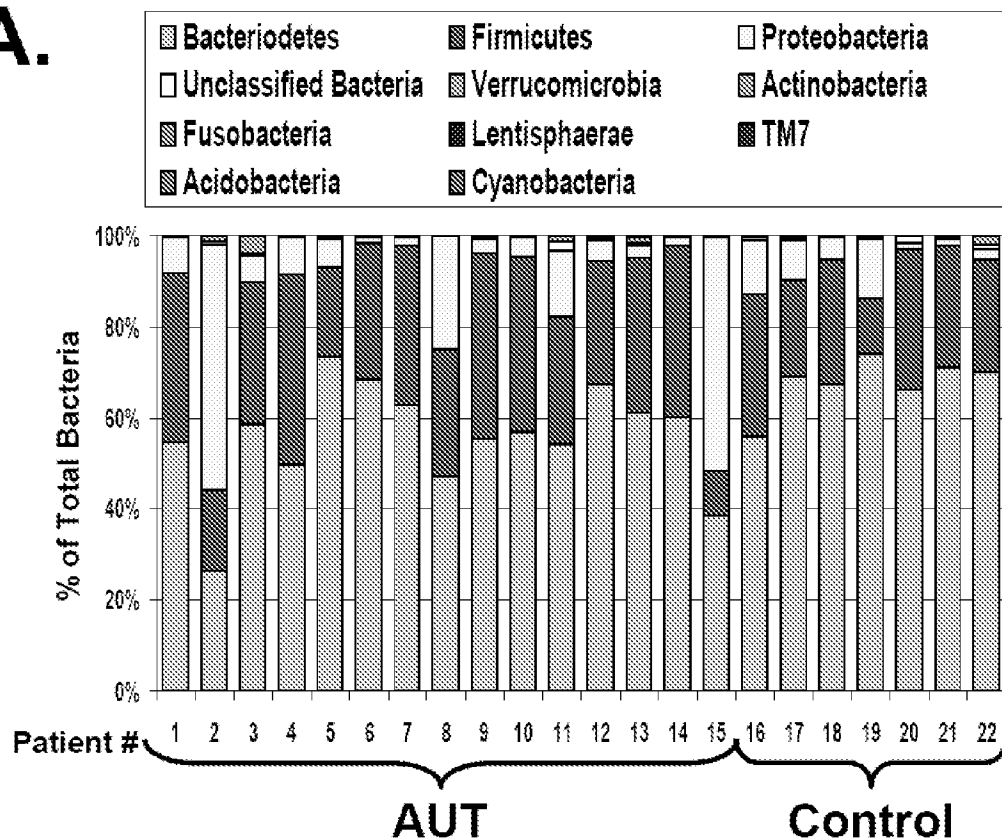
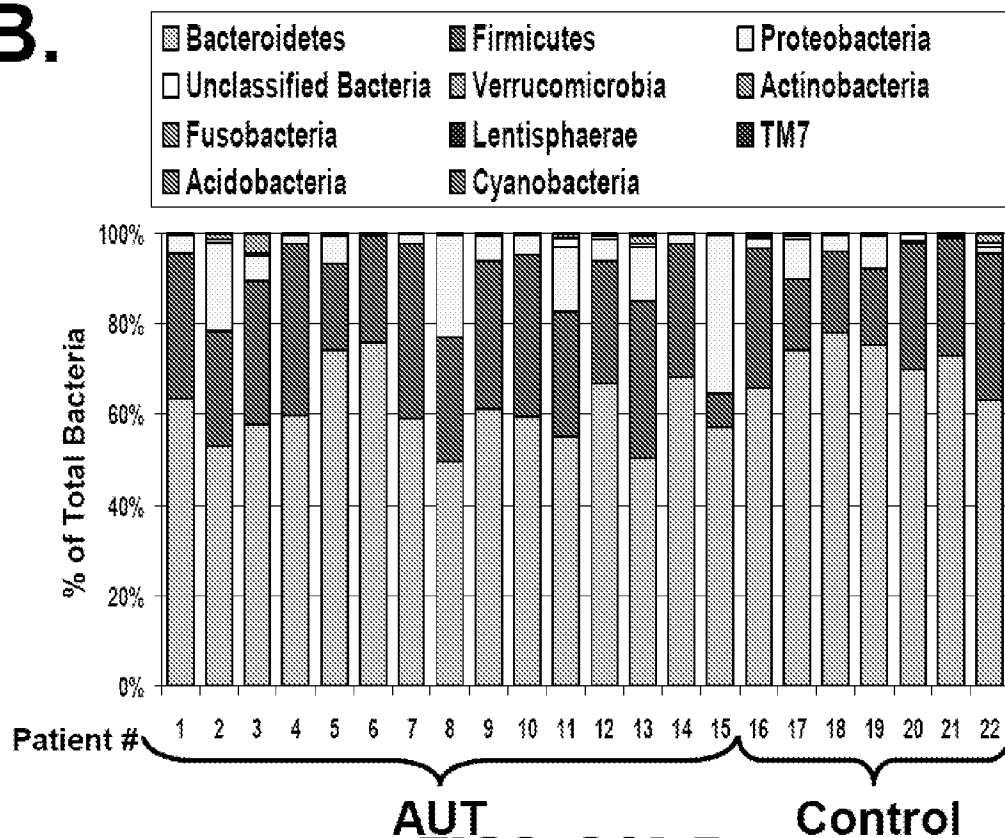


C.



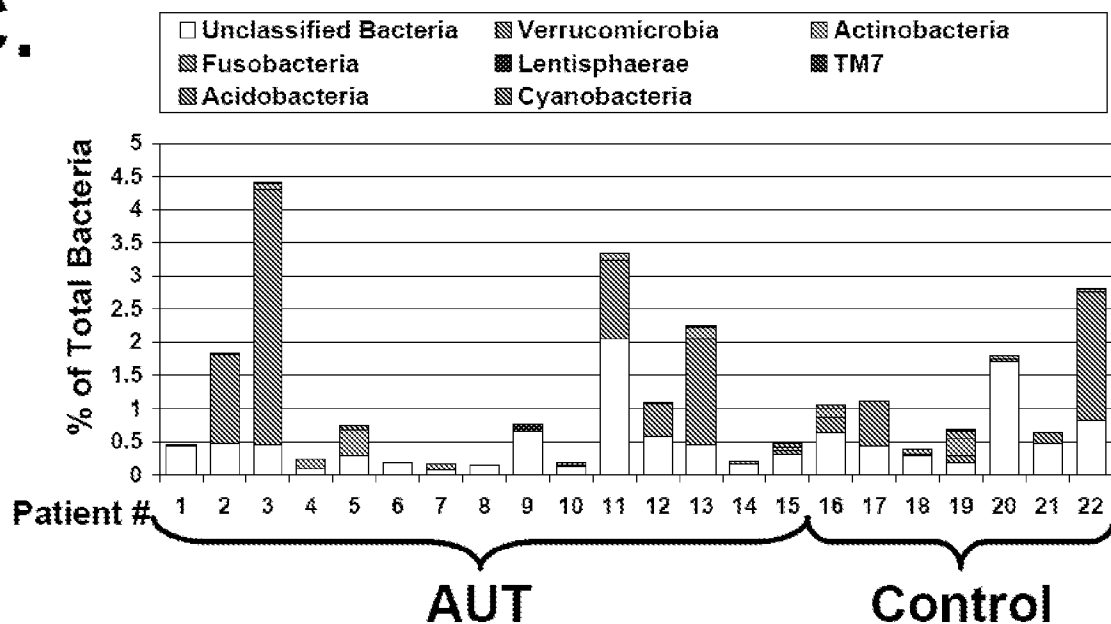
FIGS. 23C-D

38 of 46

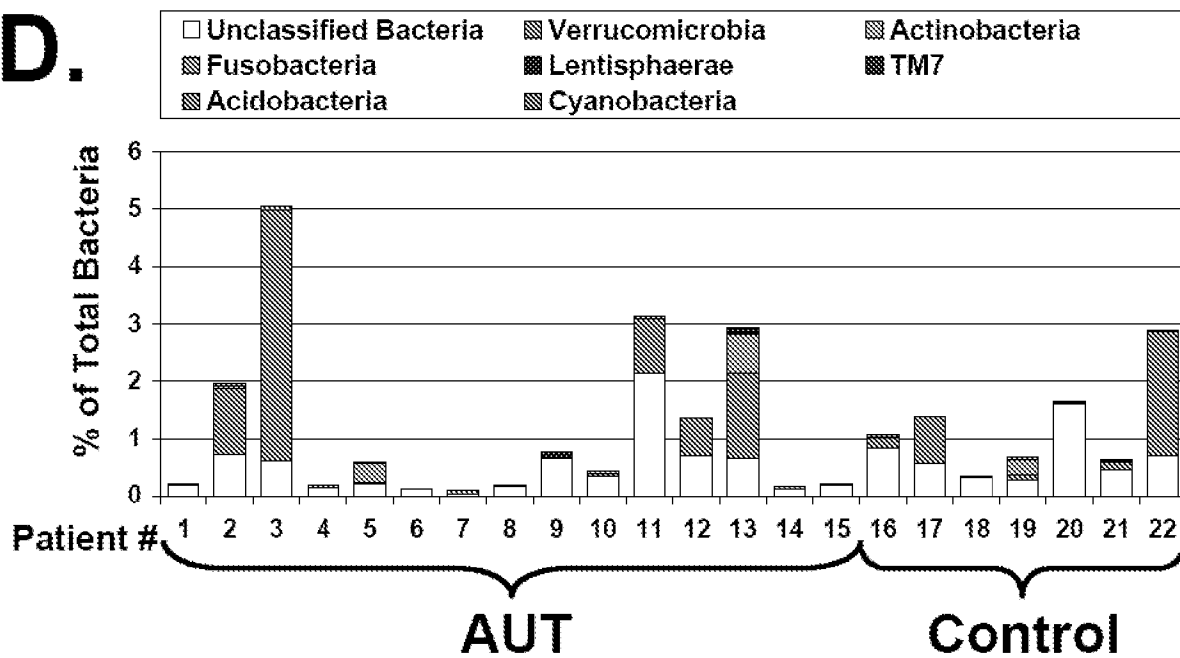
A.**B.****AUT****FIGS. 24A-B****Control**

39 of 46

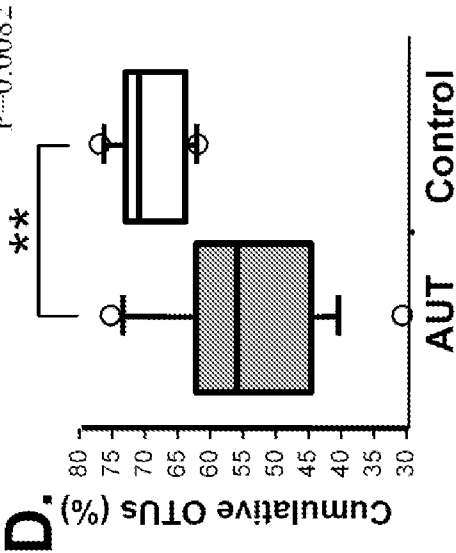
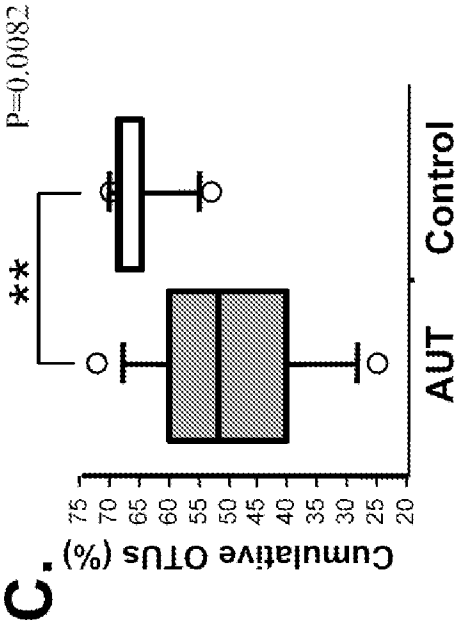
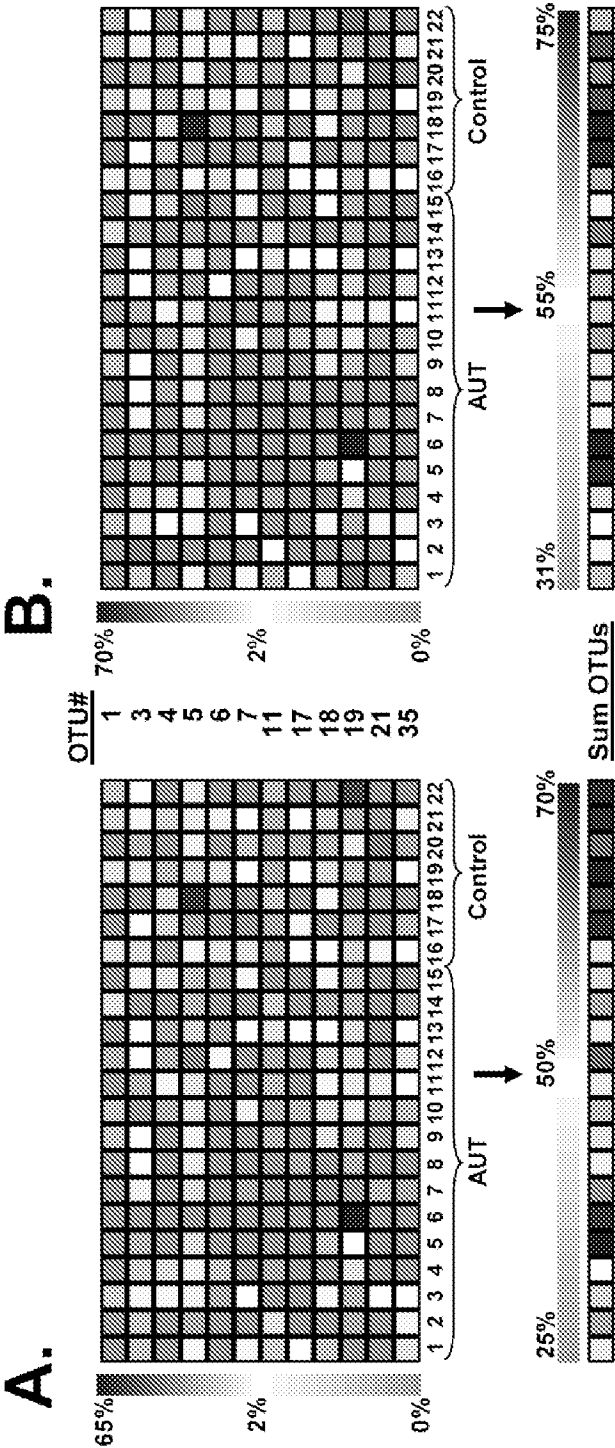
C.



D.



FIGS. 24C-D



FIGS. 25A-D

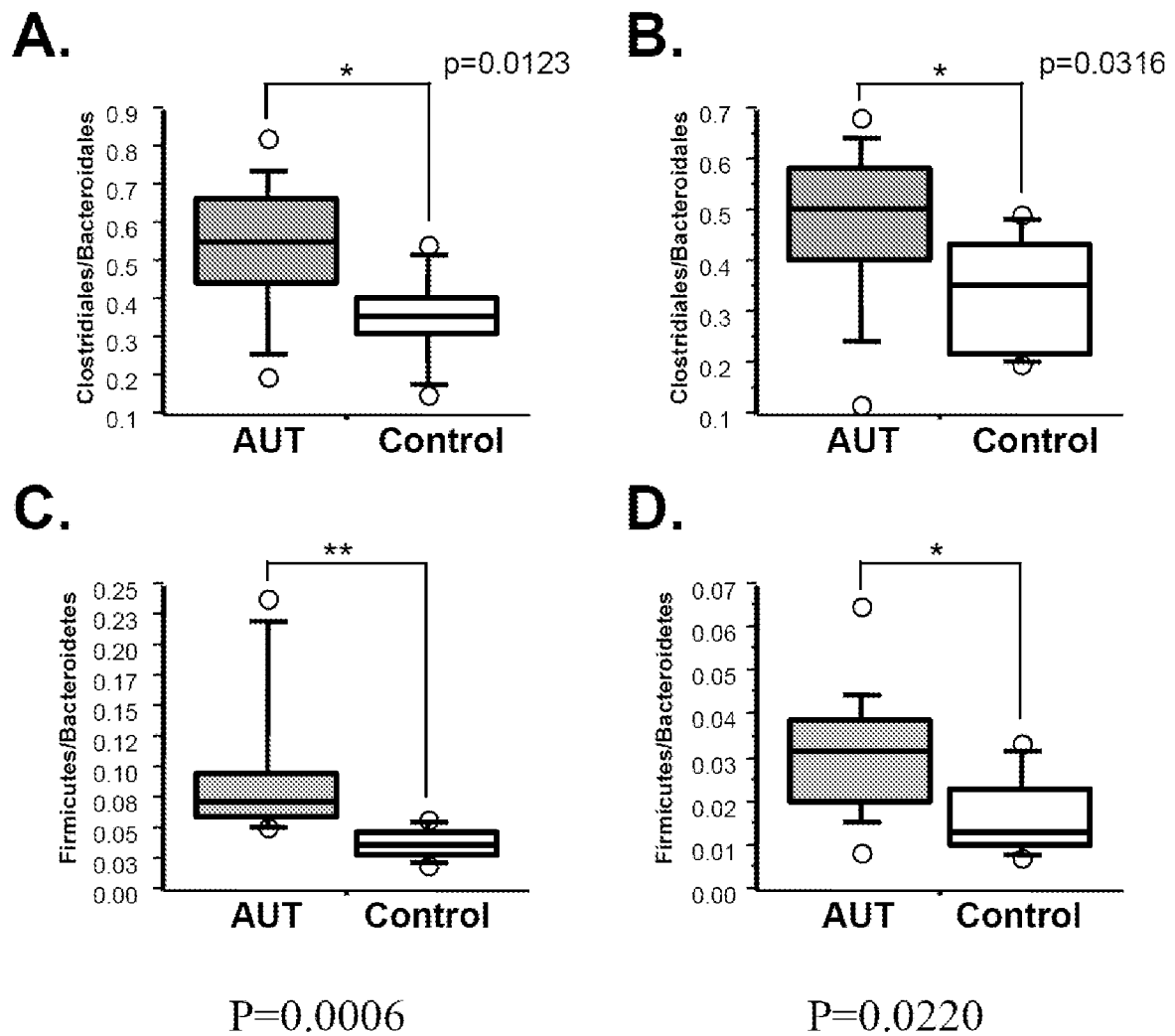
41 of 46

E.

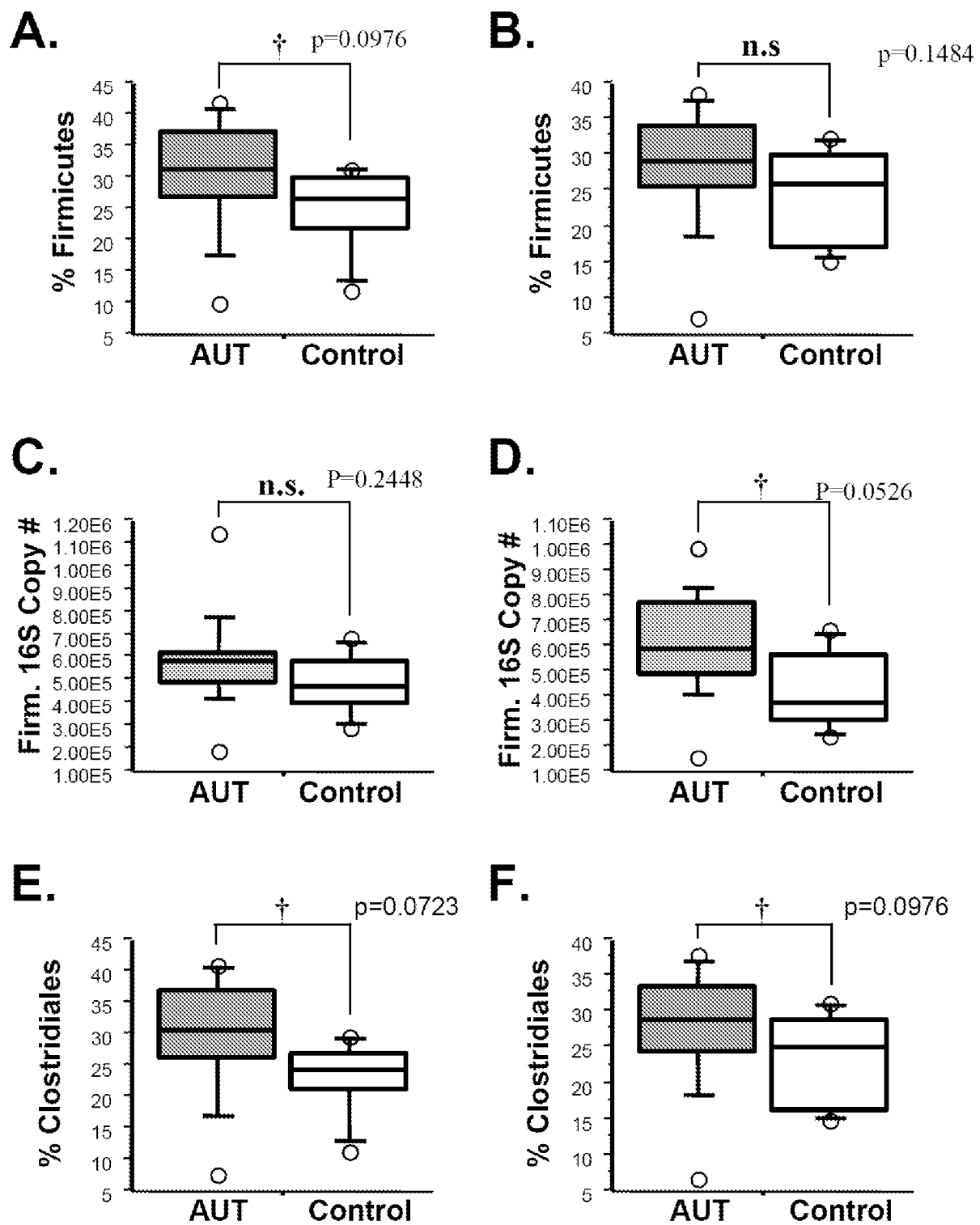
OTU#	Sequence Match (Accession#)	%Identity	E-value
1	Human colon mucosal biopsy clone LW72 (AY976241.1)	100.0	e-103
3	<i>Bacteroides ovatus</i> (AAXF02000043.1)	100.0	e-107
4	Human fecal clone RL204 aa158c11 (DQ824027.1)	100.0	e-103
5	<i>Bacteroides thetaiotaomicron</i> (AY895203.1)	98.4	e-104
6	Human fecal clone B2_414 (EU764955.1)	100.0	e-107
7	Human fecal clone RL306aa192h04 (DQ805937.1)	100.0	e-105
11	<i>Alistipes Finegoldii</i> (AY643083.1)	99.2	e-110
17	Human colon mucosal biopsy clone Z468 (AY979269.1)	100.0	e-106
18	<i>Parabacteroides distasonis</i> (EU722736.1)	99.6	1e-96
19*	<i>Bacteroides vulgatus</i> (NC_009614)	98.8	4e-117
21	<i>Prevotella copri</i> (AB244773.1)	100.0	e-106
35*	<i>Alistipes putredinis</i> (NZ_ABFK02000016)	99.6	3e-118

 = *Bacteroidaceae*
 = *Porphyromonadaceae*
 = *Rikenellaceae*
 = *Prevotellaceae*
FIG. 25E

42 of 46

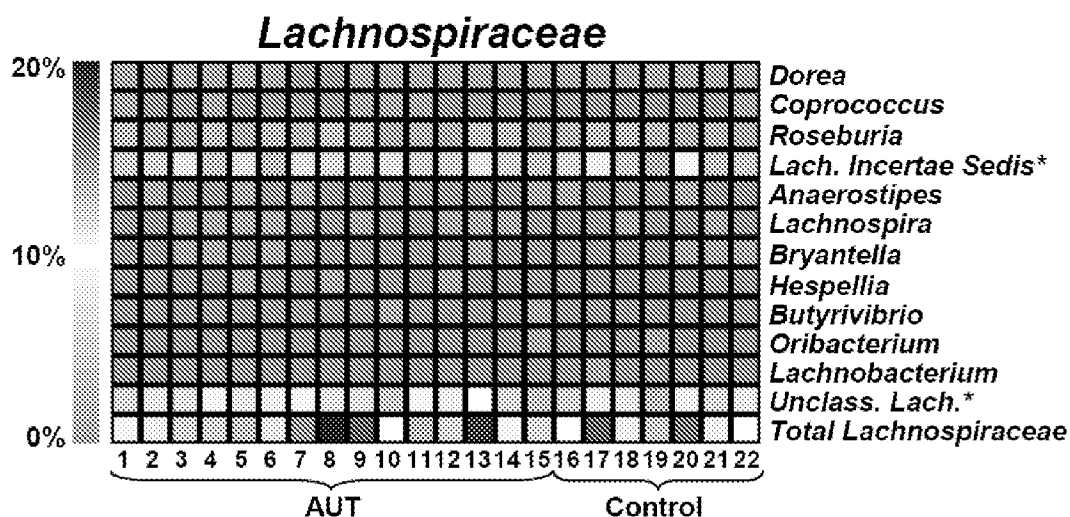
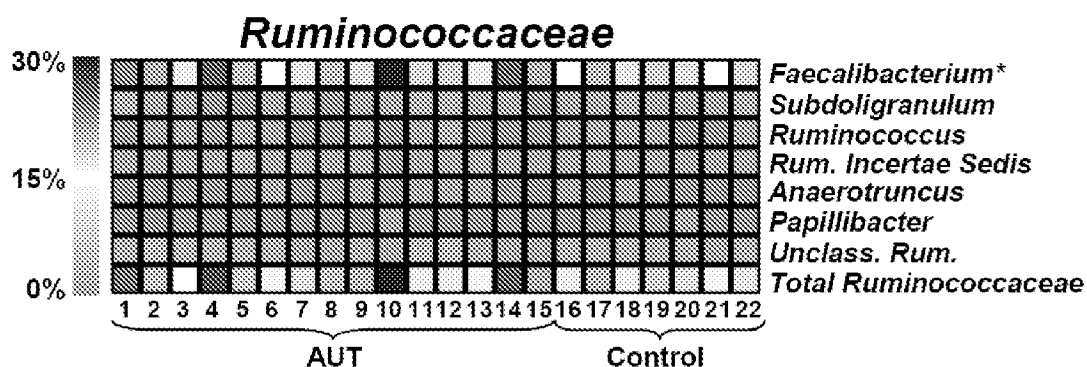
**FIG. 26**

43 of 46

**FIG. 27**

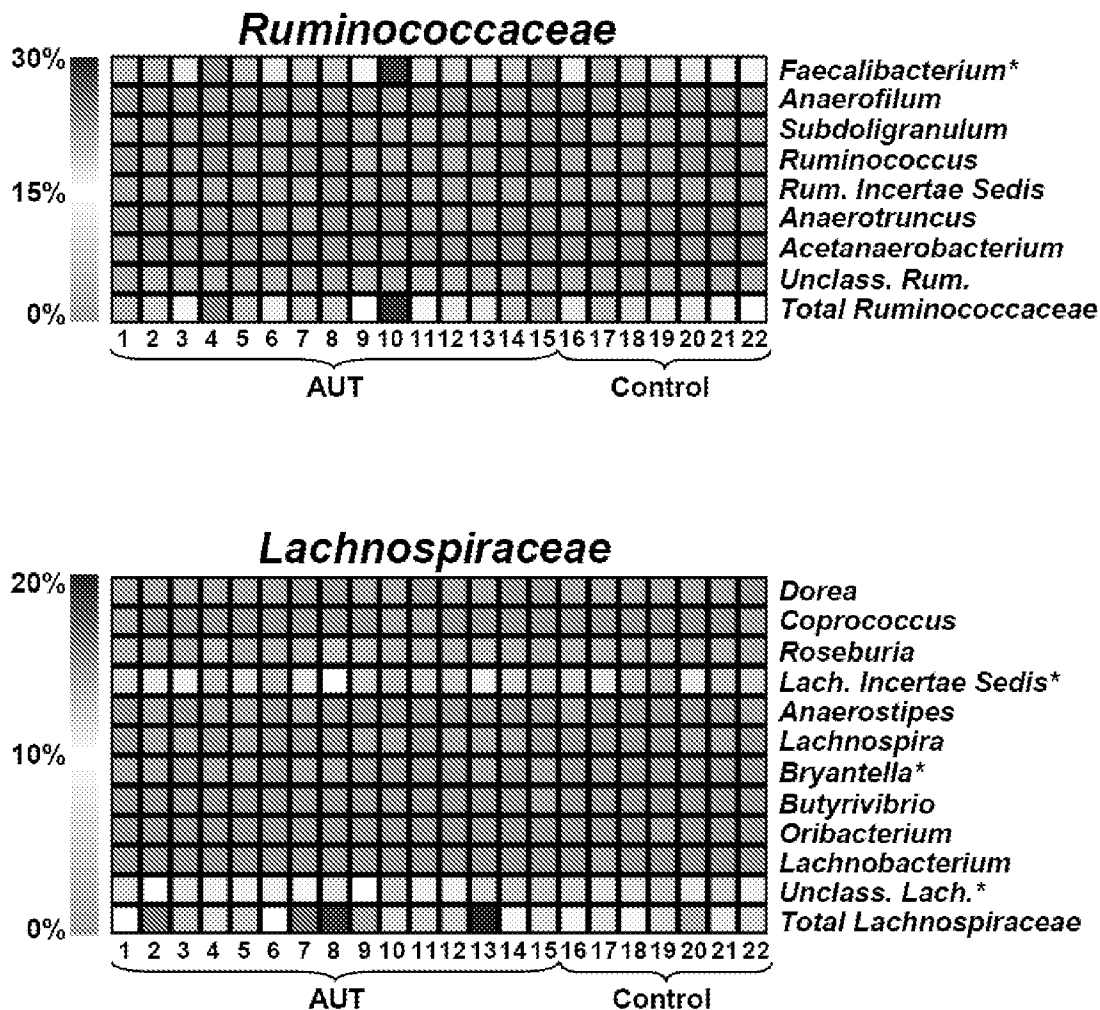
44 of 46

A. Ileum

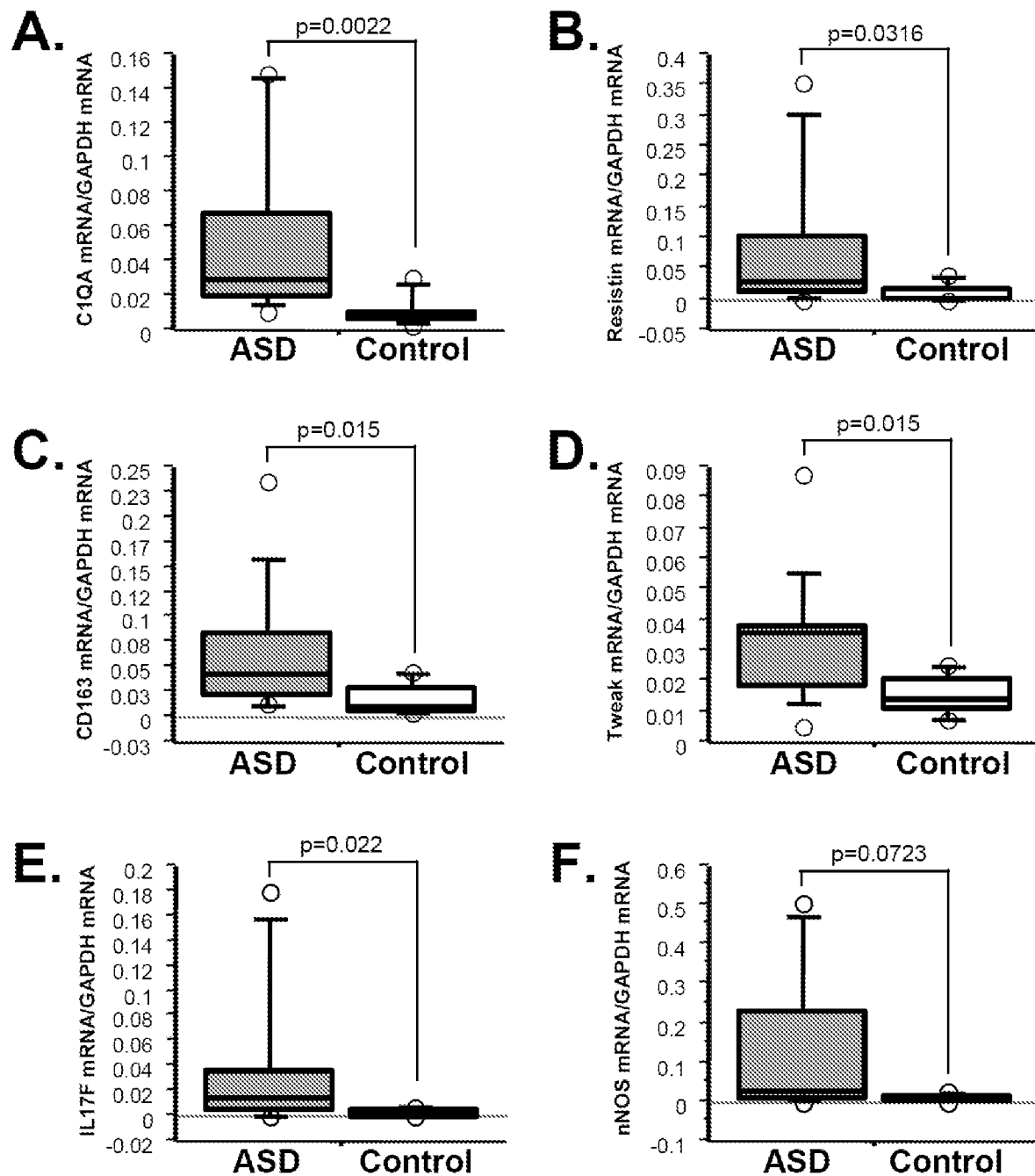
**FIG. 28A**

45 of 46

B. Cecum

**FIG. 28B**

46 of 46

**FIG. 29**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/34254

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/68 (2010.01)

USPC - 435/6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8): C12Q 1/68 (2010.01)

USPC: 435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 435/7.1

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST (PGPB,USPT,EPAB,JPAB): autism, sucrase, maltase, lactase, embryo, fetus, sutterella, mRNA

Google Scholar: autism sutterella gene locus mRNA lactase

esp@cenet: Columbia, autism, Ian Lipkin, Mady Hornig, Brent Williams, biomarkers

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2006/0115467 A1 (PANGBORN et al.) 1 June 2006 (01.06.2006), abstract; para [0002], [0011], [0012], [0035], [0047], [0048].	1, 3, 4, 6-8, 10-12
Y	US 2007/0180544 A1 (TAYLOR et al.) 2 August 2003 (02.08.2003), abstract; para [0052], [0098], [0125].	1, 3, 4, 6-8, 10-12
Y	US 2004/0170617 A1(FINEGOLD) 2 September 2004 (02.09.2004), abstract; para [0050], [0051], [0053].	6, 7
Y	US 2009/0011414 A1 (PHILIPPI et al.) 8 January 2009 (08.01.2009), abstract; para [0017], [0036], [0043], [0098], [0125], [0127].	8, 10

☐ Further documents are listed in the continuation of Box C.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

30 September 2010 (30.09.2010)

Date of mailing of the international search report

15 OCT 2010

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/34254

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1, 3, 4, 6-8, and 10-12, drawn to a method for detecting the presence of or a predisposition to autism or an autism spectrum disorder (ASD) in a human subject by detecting an alteration in the expression of a carbohydrate metabolic enzyme protein.

Group II: Claims 2, 3, 5-7, and 9-12, drawn to a method for detecting the presence of or a predisposition to autism or an autism spectrum disorder (ASD) in a human subject by detecting an alteration in the expression of a carbohydrate transporter protein.

--- Please see continuation on attached additional sheet ---

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1, 3, 4, 6-8, 10-12

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/34254

Continuation of Box (III): Lack of Unity -

Group III+: Claims 13-17, drawn to a method for treating or preventing autism or an ASD in a subject in need thereof by administering an agent. Group III+ contains multiple inventions, each drawn to a different species. The species are defined as: the method of claim 13, wherein the agent is limited to i) a functional carbohydrate metabolic enzyme molecule; or ii) a carbohydrate transporter molecule or a combination of a carbohydrate transporter molecule and a functional carbohydrate metabolic enzyme. Applicant must pay an additional search fee for each elected species.

Group IV+: Claims 18-31 and 39-47, drawn to an isolated nucleic acid composition, a diagnostic kit, and a method of detecting a *Sutterella* sp. in a sample, where each invention of Groups IV+ includes one or more of SEQ ID NOS: 11-19. Due to the number of sequences in this application, Groups IV+ will be defined as necessary depending on Applicant's ultimate payment of additional fees. The additional sequences will be searched if applicant pays for each additional sequence or shows that the sequences share a special technical feature, i.e. a common structure or feature that defines a contribution over the prior art. Note that each additional sequence to be searched must be specified by the Applicant in the response to this invitation and must either (1) have an additional invention fee paid or (2) have a showing that the sequences share a common structure or feature that defines a contribution over the prior art.

Group V: 32-38, drawn to a method of treating or preventing a disease associated with elevated levels of Beta-proteobacteria.

The groups listed above do not relate to a single general inventive concept under PCT Rule 13.1 because under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons.

Groups I and II share the technical feature of a method for detecting the presence of or a predisposition to autism or an autism spectrum disorder (ASD) in a human subject by detecting an alteration in the expression of an agent which modulates carbohydrate levels.

The inventions of Groups III+ share the technical feature of a method for treating or preventing autism or an ASD in a subject in need thereof by administering an agent that modulates carbohydrate levels such as a carbohydrate transporter or carbohydrate metabolic enzyme.

However, neither of these shared technical features represent an improvement over the prior art. Pangborn et al. (US 2006/0115467 A1) teach a method for treating or preventing autism or an ASD in a subject in need thereof by administering a carbohydrate metabolic enzyme (para [0011]; [0012]). Although Pangborn et al. do not expressly teach a method of detecting the presence of or a predisposition to autism by detecting an alteration in the expression of an agent which modulates carbohydrate levels such as a carbohydrate metabolic enzyme, such would have been obvious because Pangborn et al. teach that autism is characterized by a deficiency in carbohydrate metabolic enzymes (para [0011]).

Further, neither of the above-mentioned shared technical features are present in Group IV+ or V, which have special technical features of methods and compositions for detecting *Sutterella* sp and a method of treating or preventing a disease associated with elevated levels of Beta-proteobacteria, respectively.

Accordingly, unity of invention is lacking.