



(86) Date de dépôt PCT/PCT Filing Date: 2011/09/12
(87) Date publication PCT/PCT Publication Date: 2013/03/21
(85) Entrée phase nationale/National Entry: 2014/03/07
(86) N° demande PCT/PCT Application No.: US 2011/051269
(87) N° publication PCT/PCT Publication No.: 2013/039477

(51) Cl.Int./Int.Cl. *G01N 33/50* (2006.01),
C12Q 1/68 (2006.01), *G01N 33/68* (2006.01)

(71) Demandeur/Applicant:
CREATICS LLC, US

(72) Inventeurs/Inventors:
PANNELL, LEWIS, US;
ROCKER, JANA, US;
CONTRERAS, CARLOS, US;
DI PALMA, JACK, US

(74) Agent: GOWLING LAFLEUR HENDERSON LLP

(54) Titre : PROCEDES NON INVASIFS DE DETECTION DE MOLECULES CIBLES

(54) Title: NON-INVASIVE METHODS OF DETECTING TARGET MOLECULES

colonic lavage glycans labeled with 4abee, HPLC eluted at 13.5min 0.1% TFA 6.4ul inj
jl_06211_glycans_1 516 (23.605) M3 [Ev-733016, It50, En1] (0.050,200.00,0.200,1400.00,4,Cmp); Cm(454:537)

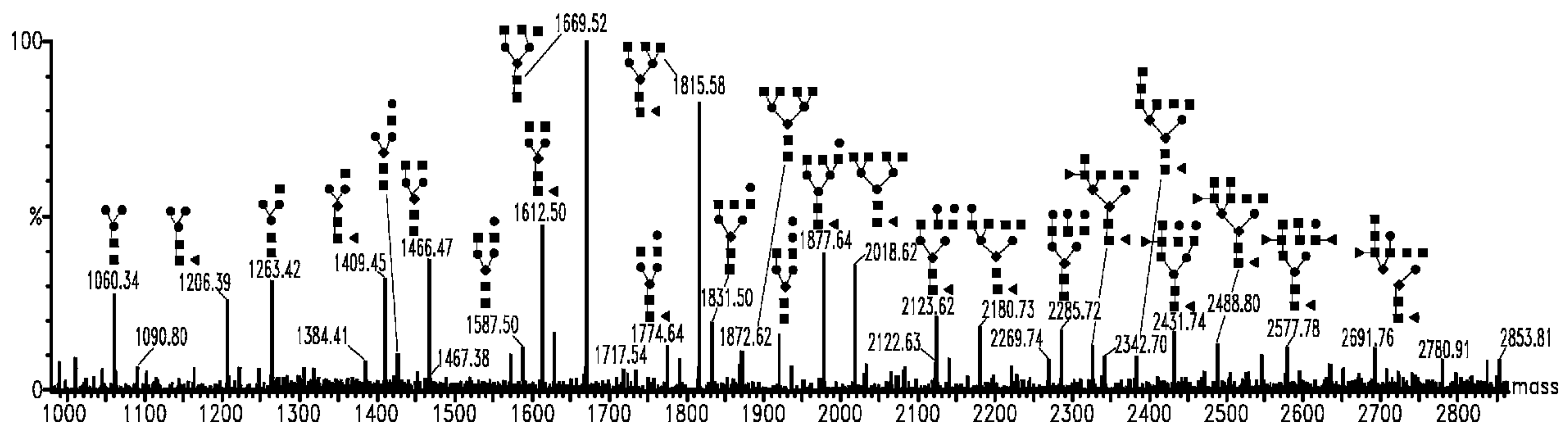


FIG. 1

◆ Sialic Acid
▲ Fucose
● Hexose (mannose)
■ HexNac

(57) Abrégé/Abstract:

Embodiments of the present invention relate to non-invasive methods and compositions for collecting detecting, measuring, and identifying target molecules. In some embodiments, methods and compositions relate to target molecules in gastrointestinal lavage fluid or feces.



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

(19) World Intellectual Property
Organization
International Bureau



(10) International Publication Number
WO 2013/039477 A1

(43) International Publication Date
21 March 2013 (21.03.2013)

(51) International Patent Classification:

G01N 33/50 (2006.01) C12Q 1/68 (2006.01)
G01N 33/68 (2006.01)

(21) International Application Number:

PCT/US2011/051269

(22) International Filing Date:

12 September 2011 (12.09.2011)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant (for all designated States except US): **UNIVERSITY OF SOUTH ALABAMA** [US/US]; 775 University Boulevard, Building Ii, Suite 150, Mobile, AL 36608 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PANNELL, Lewis** [US/US]; C/o 775 University Boulevard, Building Ii, Suite 150, Mobile, AL 36608 (US). **ROCKER, Jana** [US/US]; C/o 775 University Boulevard, Building Ii, Suite 150, Mobile, AL 36608 (US). **CONTRERAS, Carlos** [US/US]; C/o 775 University Boulevard, Building Ii, Suite 150, Mobile, AL 36608 (US). **DI PALMA, Jack** [US/US]; C/o 775 University Boulevard, Building Ii, Suite 150, Mobile, AL 36608 (US).

(74) Agent: **HART, Daniel**; Knobbe Martens Olson & Bear LLP, 2040 Main Street, Fourteenth Floor, Irvine, CA 92614 (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, QA, RO, RS, RU, RW, 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, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— with international search report (Art. 21(3))

(54) Title: NON-INVASIVE METHODS OF DETECTING TARGET MOLECULES

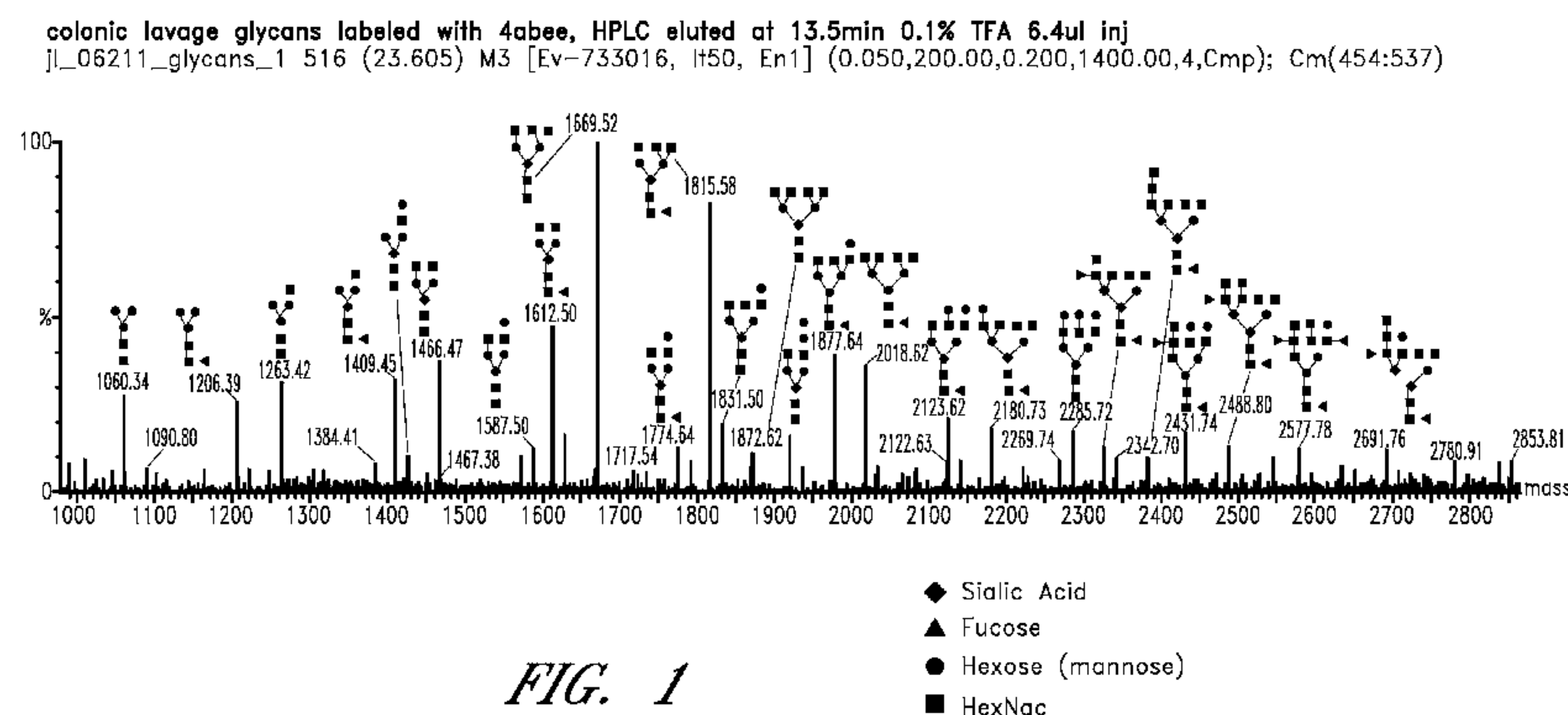


FIG. 1

(57) Abstract: Embodiments of the present invention relate to non-invasive methods and compositions for collecting detecting, measuring, and identifying target molecules. In some embodiments, methods and compositions relate to target molecules in gastrointestinal lavage fluid or feces.

NON-INVASIVE METHODS OF DETECTING TARGET MOLECULES

FIELD OF THE INVENTION

[0001] Embodiments of the present invention relate to non-invasive methods and compositions for collecting, detecting, measuring, and identifying target molecules. In some embodiments, methods and compositions relate to target molecules in gastrointestinal lavage fluid (GLF) or feces.

REFERENCE TO SEQUENCE LISTING

[0002] The present application is being filed along with a Sequence Listing in electronic format. The Sequence Listing is provided as a file entitled USA_007WO.TXT, created September 7, 2011, which is approximately 231 KB in size. The information in the electronic format of the Sequence Listing is incorporated herein by reference in its entirety.

BACKGROUND

[0003] Disorders associated with the gastrointestinal (GI) and hepatobiliary tracts and the organs/tissues associated with the GI tract include cancers such as gastric cancer, esophageal cancer, liver cancer, and pancreatic cancer. Pancreatic cancer (e.g., pancreatic adenocarcinoma), in particular, is a malignant growth of the pancreas that mainly occurs in the cells of the pancreatic ducts. This disease is the ninth most common form of cancer, yet it is the fourth and fifth leading cause of cancer deaths in men and women, respectively. Cancer of the pancreas is almost always fatal, with a five-year survival rate that is less than 3%.

[0004] The most common symptoms of pancreatic cancer include jaundice, abdominal pain, and weight loss, which, together with other presenting factors, are often nonspecific in nature. Thus, diagnosing pancreatic cancer at an early stage of tumor growth is often difficult and requires extensive diagnostic work-up, often times incidentally discovered during exploratory surgery. Endoscopic ultrasonography is an example non-surgical technique available for diagnosis of pancreatic cancer. However, reliable detection of small tumors, as well as differentiation of pancreatic cancer from focal pancreatitis, is difficult. The vast majority of patients with pancreatic cancer are

presently diagnosed at a late stage when the tumor has already extended beyond the pancreas to invade surrounding organs and/or has metastasized extensively. Gold *et al.*, Crit. Rev. Oncology/Hematology, 39:147-54 (2001), incorporated herein by reference in its entirety. Late detection of the disease is common with the majority of patients being diagnosed with advanced disease that often results in death; only a minority of patients are detected with early stage disease.

[0005] Invasive techniques to diagnose disorders and diseases related to the GI tract are inconvenient and expose a subject to significant risk. Accordingly, there is a need for non-invasive methods and compositions for the detection and identification of target molecules from the GI tract or associated organs/tissues. In some embodiments, the target molecules may be evaluated to determine whether they are useful as biomarkers associated with a particular characteristic, such as disease, predisposition to disease, positive response to a treatment regimen, or no response or negative response to a treatment regimen. In addition, biomarkers from the GI tract or associated organs/tissues may be used to determine whether an individual has any of the particular characteristics listed above.

SUMMARY

[0006] Embodiments of the present invention relate to non-invasive methods and compositions for collecting, detecting, measuring, and identifying target molecules. In some embodiments, methods and compositions relate to target molecules in gastrointestinal lavage fluid (GLF) or feces.

[0007] Some embodiments include a method for assessing the physiological state of a subject comprising: obtaining a gastrointestinal lavage fluid from the subject; and detecting a target molecule which originated from outside the gastrointestinal system in the gastrointestinal lavage fluid.

[0008] Some embodiments include a method for assessing the physiological state of a subject comprising: obtaining a fecal sample from the subject; and detecting a target molecule which originated from outside the gastrointestinal system in the fecal sample.

[0009] In some embodiments, the gastrointestinal lavage fluid is obtained from the subject by partially purging the subject's gastrointestinal system.

[0010] In some embodiments, the gastrointestinal lavage fluid comprises fecal matter. In some embodiments, the fecal sample comprises a gastrointestinal lavage fluid.

[0011] In some embodiments, the target molecule comprises a polypeptide, antibody, bile acid, metabolite, or glycan. In some embodiments, the target molecule comprises a biomarker. In some embodiments, the biomarker is associated with a disease, a positive response to treatment, a partial response to treatment, a negative response to treatment, or no response to treatment. In some embodiments, the target molecule is associated with presence of a cancer or predisposition to a cancer. In some embodiments, the cancer is pancreatic cancer, colorectal cancer, liver cancer, or gastric cancer. In some embodiments, the target molecule originated from an accessory digestive gland. In some embodiments, the accessory digestive gland is salivary glands, pancreas, gallbladder, or liver.

[0012] Some embodiments include administering a lavage fluid to the subject. In some embodiments, lavage fluid is administered orally. In some embodiments, the lavage fluid comprises an ingredient selected from the group consisting of polyethylene glycol, magnesium sulfate, sodium sulfate, potassium sulfate, magnesium citrate, ascorbic acid, sodium picosulfate, and bisacodyl. In some embodiments, the lavage fluid is selected from the group consisting of GOLYTELY, HALFLYTELY, NULYTELY, SUPREP, FLEET'S PHOSPHO-SODA, magnesium citrate, and their generic equivalents.

[0013] Some embodiments include performing a colonoscopy on the subject.

[0014] In some embodiments, the subject is mammalian. In some embodiments, the subject is human.

[0015] Some embodiments include a method for identifying a biomarker comprising: obtaining a test gastrointestinal lavage fluid from a plurality of test subjects having a condition or physiological state of interest and a control gastrointestinal lavage fluid from a plurality of control subjects who do not have said condition or physiological state; determining the level of at least 5 target molecules in the test gastrointestinal lavage fluid and the control gastrointestinal lavage fluid, and identifying a target molecule which is present at significantly different levels in the test gastrointestinal lavage fluid relative to the levels in the control gastrointestinal lavage fluid, thereby identifying a biomarker.

[0016] In some embodiments, the gastrointestinal lavage fluid comprises fecal matter.

[0017] In some embodiments, the target molecules are selected from the group consisting of polypeptides, bile acids, antibodies, metabolites, glycans, and a combination thereof.

[0018] Some embodiments include determining the level of at least 10 target molecules in the test gastrointestinal lavage fluid and the control gastrointestinal lavage fluid. Some embodiments include determining the level of at least 20 target molecules in the test gastrointestinal lavage fluid and the control gastrointestinal lavage fluid. Some embodiments include determining the level of at least 30 target molecules in the test gastrointestinal lavage fluid and the control gastrointestinal lavage fluid. Some embodiments include determining the level of at least 50 target molecules in the test gastrointestinal lavage fluid and the control gastrointestinal lavage fluid. Some embodiments include determining the level of at least 100 target molecules in the test gastrointestinal lavage fluid and the control gastrointestinal lavage fluid.

[0019] In some embodiments, the biomarker is associated with a disease, a positive response to treatment, a partial response to treatment, a negative response to treatment or no response to treatment.

[0020] In some embodiments, the biomarker is associated with the presence of a cancer or predisposition to a cancer. In some embodiments, the cancer is pancreatic cancer, liver cancer, or gastric cancer.

[0021] In some embodiments, at least one target molecule originated from an accessory digestive gland. In some embodiments, the accessory digestive gland is salivary glands, pancreas, gallbladder, or liver.

[0022] Some embodiments include administering a lavage fluid to the test subjects and the control subjects. In some embodiments, the lavage fluid is administered orally. In some embodiments, the lavage fluid comprises an ingredient selected from the group consisting of polyethylene glycol, magnesium sulfate, sodium sulfate, potassium sulfate, magnesium citrate, ascorbic acid, sodium picosulfate, and bisacodyl. In some embodiments, the lavage fluid is selected from the group consisting of GOLYTELY, HALFLYTELY, NULYTELY, SUPREP, and FLEET'S PHOSPHO-SODA, magnesium citrate, and their generic equivalents.

[0023] Some embodiments include performing a colonoscopy on the test subjects and control subjects.

[0024] In some embodiments, the test subjects and control subjects are mammalian. In some embodiments, the test subjects and control subjects are human.

[0025] Some embodiments include a method for identifying a biomarker comprising: obtaining a test fecal sample from a plurality of test subjects having a condition of interest and a control fecal sample from a plurality of control subjects and; determining the level of at least 5 target molecules in the test fecal sample and the control fecal sample, identifying a target molecule which is present at significantly different levels in the test fecal sample relative to the levels in the control fecal sample, thereby identifying a biomarker.

[0026] In some embodiments, the fecal sample comprises a gastrointestinal lavage fluid.

[0027] In some embodiments, the target molecules are selected from the group consisting of polypeptides, bile acids, antibodies, metabolites, glycans, and a combination thereof.

[0028] Some embodiments include determining the level of at least 10 target molecules in the test fecal sample and the control fecal sample. Some embodiments include determining the level of at least 20 target molecules in the fecal sample and the control fecal sample. Some embodiments include determining the level of at least 30 target molecules in the fecal sample and the control fecal sample. Some embodiments include determining the level of at least 50 target molecules in the fecal sample and the control fecal sample. Some embodiments include determining the level of at least 100 target molecules in the fecal sample and the control fecal sample.

[0029] In some embodiments, the biomarker is associated with a disease, a positive response to treatment, or a negative response to treatment.

[0030] In some embodiments, the biomarker is associated with the presence of a cancer or predisposition to a cancer. In some embodiments, the cancer is pancreatic cancer, colorectal cancer, liver cancer, or gastric cancer.

[0031] In some embodiments, at least one target molecule originated from an accessory digestive gland. In some embodiments, the accessory digestive gland is salivary glands, pancreas, gallbladder, or liver.

[0032] In some embodiments, the test subjects and control subjects are mammalian. In some embodiments, the test subjects and control subjects are human.

[0033] Some embodiments include a kit for detecting a target molecule in a gastrointestinal lavage fluid comprising: a lavage fluid for oral administration to a subject; a vessel for collecting the gastrointestinal lavage fluid from the subject; and an agent for detecting a target molecule which originated from outside the gastrointestinal system.

[0034] Some embodiments include a kit for detecting a target molecule in a fecal sample comprising: a lavage fluid for oral administration to a subject; a vessel for collecting the fecal sample from the subject; and an agent for detecting a target molecule which originated from outside the gastrointestinal system.

[0035] Some embodiments include a protease inhibitor.

[0036] In some embodiments, the target molecule comprises a polypeptide, antibody, bile acid, metabolite, or glycan. In some embodiments, the target molecule comprises a biomarker. In some embodiments, the biomarker is associated with a disease, a positive response to treatment, or a negative response to treatment.

[0037] In some embodiments, the target molecule is associated with presence of a cancer or predisposition to a cancer. In some embodiments, the cancer is pancreatic cancer, liver cancer, colorectal cancer, or gastric cancer.

[0038] In some embodiments, the target molecule originated from an accessory digestive gland. In some embodiments, the accessory digestive gland is salivary glands, pancreas, gallbladder, or liver.

[0039] In some embodiments, the lavage fluid comprises an ingredient selected from the group consisting of polyethylene glycol, magnesium sulfate, sodium sulfate, potassium sulfate, magnesium citrate, ascorbic acid, sodium picosulfate, and bisacodyl. In some embodiments, the lavage fluid is selected from the group consisting of GOLYTELY, HALFLYTELY, NULYTELY, SUPREP, FLEET'S PHOSPHO-SODA, magnesium citrate, and their generic equivalents.

BRIEF DESCRIPTION OF THE DRAWINGS

[0040] FIG. 1 depicts a graph of the relative abundance of various glycoprotein derived glycan structures present in a fraction of a gastrointestinal lavage fluid. Derivatized glycans were eluted from a C18 reverse phase column on a Q-TOF MS at about 20-25% acetonitrile in 0.2% formic acid. The mass spectrometer was scanned in MS-only mode from m/z 150-2000 every second to acquire the derivatized glycan profile data.

[0041] FIG. 2 depicts a graph of the relative abundance of compounds, including metabolites such as cholic acid, present in a fraction of a gastrointestinal lavage fluid. Data were acquired on a Waters Q-TOF mass spectrometer using input from an LC system, and using MassLynx software. The MS scanned over the mass range from m/z 100 m/z to 2000 every second.

DETAILED DESCRIPTION

[0042] Embodiments of the present invention relate to non-invasive methods and compositions for collecting, detecting, measuring, and identifying target molecules. In some embodiments, methods and compositions relate to target molecules in gastrointestinal lavage fluid (GLF) or feces.

[0043] Gastrointestinal lavage is widely used as a lower gastrointestinal (GI) tract preparation for colonoscopy or colorectal surgery (*see e.g.*, DiPalma JA. *et al.*, (1984) *Gastroenterology* 86:856–60), incorporated herein by reference in its entirety). Particular pathophysiologies of intestinal diseases have been investigated by measuring proteins in GLF (Evgenikos N, *et al.* (2000) *Br J Surg* 87:808–13; Brydon WG, Ferguson A. (1992) *Lancet* 340: 1381–2; Choudari CP, *et al.* (1993) *Gastroenterology* 104: 1064–71; Ferguson A, *et al.* (1996) *Gut* 38:120–4; Handy LM, *et al.* (1996) *Scand J Gastroenterol* 31:700–5; and Stanley AJ, *et al.* (1996) *Gastroenterology* 111:1679–82), each incorporated herein by reference in its entirety).

[0044] Measurements of fecal proteins can be useful for investigating various pathophysiologies such as protein-losing enteropathy and mucosal inflammation. However, while feces may be used in some embodiments described herein, GLF is preferred over feces as a sample for detecting and identifying biomarkers because GLF contains smaller amounts of substances that interfere with assays, and destruction of protein by digestive enzymes and bacterial proteases is less in GLF than a fecal sample because of its quick transit through the GI tract. In addition, it is possible to estimate the rate of protein release from the mucosa, because the rate of fluid passage along the gut can be estimated.

[0045] In some embodiments, a GLF can be produced by orally administering a lavage fluid to a subject that causes a large volume of fluid to pass through the intestinal tract, the lavage fluid can contain a mixture of salts and other materials such as polyethylene glycol and bisacodyl. The lavage fluid causes an influx of liquid into the

colon that causes a flushing out of solids. Lavage fluids are commonly used to cause clearing of the GI tract as is commonly used in preparation for a colonoscopy and other methods used to examine the GI tract. These liquids that are flushed out or remain in the largely cleared GI tract are useful to evaluate a variety of diseases due to the continuity of the mouth to the anus along the GI tract. Consequently, any and all organs, including the GI tract, which deposit fluids into the GI tract are candidates for the methods and compositions provided herein.

GI tract and associated organs/tissues

[0046] Some of the methods and compositions provided herein relate to the GI tract and organs/tissues associated with the GI tract including accessory digestive glands. As is well known in the art, the GI tract includes the upper GI tract and lower GI tract. The upper GI tract includes the oral or buccal cavity, esophagus, stomach and duodenum. The lower GI tract includes the jejunum, ileum and the large intestine and the anus. The large intestine includes the appendix, cecum, colon, and rectum.

[0047] Organs and tissues associated with the GI tract include structures outside the GI tract. Examples of such structures include accessory digestive organs such as salivary glands, *e.g.*, parotid salivary glands, submandibular salivary glands, and sublingual salivary glands, pancreas, *e.g.*, exocrine pancreas, gallbladder, bile duct, and liver. More examples of structures associated with the GI tract and outside the GI tract include the pancreatic duct, biliary tree, and bile duct.

Gastrointestinal lavage fluid (GLF)

[0048] Generally, a lavage fluid can be orally administered to a subject, the oral lavage fluid passes through the GI tract of the subject, and the resulting GLF is collected from the subject. As used herein, the term “subject” can include an animal, such as a mammal, such as a human. As noted above, GLF provides a cleaner sampling of the GI tract than the examination of feces/stool samples. GLFs appear to mitigate variability related to food intake, type and digestive status.

[0049] Some embodiments described herein include analysis of a GLF for detecting a target molecule or for screening, triage, disease detection, diagnosis, prognosis, response to treatment, selection of treatment and personalized medicine for diseases and pathological conditions of the gastrointestinal tract or associated

organs/tissues. Some embodiments include analysis of a GLF sample for eliminating particular diseases and pathological conditions from the possible diseases or conditions from which a subject may be suffering. Some embodiments include analysis of a GLF for indicating the need for further testing for diagnosis. More embodiments include the analysis of GLF to establish a new disease diagnosis, further classifying a previous diagnosis, determining the sensitivity to potential treatment regimens, and/or evaluating the response to previous or ongoing treatment regimens.

Methods for obtaining a GLF

[0050] Some embodiments of the methods and compositions provided herein include obtaining a GLF from a subject. Methods of obtaining a GLF are well known in the art. For example, during medical and or diagnostic procedures such as sigmoidoscopy, colonoscopy, radiographic examination, preparation for patients undergoing bowel surgery, it is important that the bowels and colon be thoroughly purged and cleaned. In particular, it is essential that as much fecal matter as possible be removed from the colon to permit adequate visualization of the intestinal mucosa. This is important prior to, for example, diagnostic procedures such as flexible sigmoidoscopy or colonoscopy, diagnostic examinations widely performed to screen patients for diseases of the colon. In addition, it is important that the intestines be cleansed thoroughly in order to obtain satisfactory radiographs of the colon. The same condition also applies when the colon is preoperatively prepared for surgery, where removal of fecal waste materials is critically important for patient safety. To prepare the colon for endoscopic exam, current cleaning procedures include orthograde colonic lavage. Orthograde lavage can include orally administering a lavage composition to a subject comprising 4 L of a polyethylene glycol/electrolyte solution (U.S. Patent Application Publication No. 20070298008, incorporated by reference in its entirety). Some embodiments include antegrade lavage and retrograde lavage.

[0051] Generally, oral lavage compositions include solutions of electrolytes, such as sodium, potassium and magnesium salts of sulfate, bicarbonate, chloride, phosphate or citrate. Some such compositions may also include polyethylene glycol, which can act as a non-absorbable osmotic agent. Generic compositions include polyethylene glycol with an electrolyte solution, optionally also including bisacodyl, or ascorbic acid, and compositions including sulfate salts such as sodium sulfate, magnesium

sulfate, or potassium sulfate. In some embodiments, an oral lavage fluid can include magnesium citrate. In some embodiments, an oral lavage fluid can include sodium picosulfate. One example composition of an oral lavage solution comprising polyethylene glycol with an electrolyte solution is GOLYTELY (Braintree Labs. Inc.). GOLYTELY is formulated according to the following: polyethylene glycol 59 g, sodium sulfate 5.68 g, sodium bicarbonate 1.69 g, sodium chloride 1.46 g, potassium chloride 0.745 g and water to make up one liter (Davis *et al.* (1980) *Gastroenterology* 78:991-995, incorporated by reference in its entirety). Ingestion of GOLYTELY produces a voluminous, liquid stool with minimal changes in the subject's water and electrolyte balance. Another example of an oral lavage composition comprising polyethylene glycol with an electrolyte solution is NULYTELY (Braintree Labs. Inc.). An example oral lavage composition comprising polyethylene glycol with an electrolyte solution and bisacodyl is HALFLYTELY (Braintree Labs. Inc.). An example oral lavage composition comprising sulfate salts, such as sodium sulfate, magnesium sulfate, or potassium sulfate is SUPREP (Braintree Labs. Inc.). An example composition of an oral lavage solution comprising polyethylene glycol with an electrolyte solution and ascorbic acid is MOVIPREP (Salix Pharmaceuticals, Inc.).

[0052] Polyethylene glycol is effective as an oral lavage composition when large amounts of polyethylene glycol are administered in large volumes of a dilute salt solution. Usually about 250 - 400 g polyethylene glycol are administered to the subject in about 4 L of an electrolyte solution in water. Oral administration of polyethylene glycol can be used to produce a bowel movement over a period of time, e.g., overnight. The dose required will vary, but from about 10 - 100 g of polyethylene glycol in 8 oz. of water can be effective. A dose of from about 68 - 85 g of polyethylene glycol can be effective to produce an overnight bowel movement, without profuse diarrhea. A volume of a solution of polyethylene glycol in an isotonic fluid can be an effective amount of an osmotic laxative. Volumes from about 0.5 L to about 4 L can be effective. Preferably the effective volume is between about 1.5 L and about 2.5 L. Oral administration of 2 L of isotonic solution is effective.

[0053] More examples of oral lavage compositions include hypertonic solutions of non-phosphate salts with an osmotic laxative agent such as polyethylene glycol (U.S. Pat. App. No. 20090258090, incorporated by reference in its entirety). Mixtures of sulfate salts that omit phosphates, for example, effective amounts of one or

more of the following sulfate salts Na_2SO_4 , MgSO_4 , and K_2SO_4 can be effective (*e.g.*, SUPREP). Some embodiments include about 0.1 g to about 20.0 g Na_2SO_4 , and from about 1.0 g to 10.0 g Na_2SO_4 may be useful. Dosage amounts of MgSO_4 from about 0.01 g to about 40.0 g can be effective. Doses of from about 0.1 g to about 20.0 g Na_2SO_4 may also be advantageously used, as well as dosages of 1.0 to 10.0 g. Dosage amounts of K_2SO_4 from about 0.01 g to about 20.0 g can be effective to produce purgation, and doses of from about 0.1 g to about 10.0 g and from about 0.5 g to about 5.0 g K_2SO_4 may also be useful. Addition of an osmotic laxative agent, such as polyethylene glycol (PEG) may improve the effectiveness of the above salt mixtures. Doses of PEG from about 1.0 g to about 100 g PEG are effective. Doses from about 10.0 g to about 50 g of PEG are also effective, as is a dose of about 34 g. For ease of administration, the above mixture of salts can be dissolved in a convenient volume of water. A volume of less than one liter of water can be well tolerated by most subjects. The mixture can be dissolved in any small volume of water, and volumes of between 100 and 500 ml are useful. The effective dose may be divided and administered to the patient in two or more administrations over an appropriate time period. Generally, administration of two doses of equal portions of the effective dose, separated by 6 to 24 hours produces satisfactory purgation. Some embodiments include cessation of normal oral intake during a defined period before and during administration of an oral lavage composition.

[0054] Some lavage compositions include a laxative, such as bisacodyl. In some embodiments, a laxative can be co-administered to a subject with a lavage composition. As will be understood, such co-administration can include, for example, administration of a laxative up to several hours before administration of a lavage composition to a subject, administration of a laxative with the administration of a lavage composition to a subject, or administration of a laxative up to several hours after administration of a lavage composition to a subject. Examples of laxatives and their effective doses include Aloe, 250-1000 mg.; Bisacodyl, about 5-80 mg.; Casanthranol, 30-360 mg.; Cascara aromatic fluid extract, 2-24 ml.; Cascara sagrada bark, 300-4000 mg.; Cascara sagrada extract, 300-2000 mg.; Cascara sagrada fluid extract, 0.5-5.0 ml.; Castor oil, 15-240 ml.; Danthron, 75-300 mg.; Dehydrocholic Acid, 250-2000 mg; Phenolphthalein, 30-1000 mg.; Sennosides A and B, 12-200 mg.; and Picosulfate, 1-100 mg.

[0055] More examples of lavage compositions include aqueous solutions of concentrated phosphate salts. The aqueous phosphate salt concentrate produces an osmotic effect on the intra-luminal contents of the GI tract, evacuation of the bowel occurs with a large influx of water and electrolytes into the colon from the body. One exemplary composition comprises 480 g/L monobasic sodium phosphate and 180 g/L dibasic sodium phosphate in stabilized buffered aqueous solution (FLEET'S PHOSPHO-SODA, C. S. Fleet Co., Inc.). Subjects are typically required to take 2 - 3 oz doses of this composition, separated by a three to 12 hour interval for a total of 6 ounces (180 ml).

[0056] GLF may be collected from a subject before, during, or after a medical or diagnostic procedure. In some embodiments, a subject may collect GLF, for example, using a receptacle such as a toilet insert which captures the fluid. Enzyme inhibitors and denaturants may be used to preserve the quality of the GLF. In some embodiments, the pH of the sample may be adjusted to help stabilize the samples. In some embodiments, GLF samples may be further treated to remove some or all solids and/or bacteria, such as by centrifugation. In some embodiments, the GI tract may not be fully purged by administration of an oral lavage composition. For example, a portion of a complete dose of an oral lavage composition required to fully purge the GI tract of a subject can be administered to the subject. In some embodiments, a GLF can comprise fecal matter. In more embodiments, fecal matter can comprise a GLF.

Target molecules

[0057] Some embodiments described herein relate to methods of detecting target molecules in samples obtained from the GI tract or compositions useful for such detection. As used herein, "target molecule" includes any molecule that can be detected or measured or identified in a sample from the GI tract. Such samples include a GLF from a subject, and a fecal sample from a subject. Examples of target molecules include molecules such as peptides, polypeptides, proteins, mutant proteins, proteins generated from alternative splicing, modified proteins, such as post-translationally modified proteins *e.g.*, glycosylated proteins, phosphorylated proteins, antibodies (*e.g.*, autoantibodies, IgG, IgA, and IgM), antibody fragments, sugars, *e.g.*, monosaccharides, disaccharides, oligosaccharides, and glycans, lipids, small molecules, *e.g.* metabolites, pharmaceutical compositions, metabolized pharmaceutical compositions, and pro-drugs. More examples of target molecules include bile salts and bile acids, *e.g.*, cholic acid. More examples

include chenodeoxycholic acid, glycocholic acid, taurocholic acid, deoxycholic acid, and lithocholic acid. Target molecules can originate in the GI tract and outside the GI tract, *e.g.*, from organs and/or tissues associated with the GI tract, such as accessory digestive glands. In some embodiments, cells including their fragments and their other biproducts, *e.g.*, red blood cells, white blood cells, and endothelial cells, organisms, *e.g.*, bacteria, protozoans, and viruses and viral particles can be detected in a GLF or fecal samples. In some embodiments, the target molecules may be any of the proteins or portions thereof listed in any of Tables 1-10 herein or a portion thereof. In some embodiments, the portion of the proteins listed in any of Tables 1-10 can comprise at least 10, at least 15, at least 20, at least 25, at least 50, or more than 50 consecutive amino acids of any the proteins listed in Tables 1-10. In some embodiments, the target molecules may comprise, consist essentially of, or consist of a polypeptide of one of SEQ ID NO.s: 01-804. A polypeptide consisting essentially of one of SEQ ID NO.s: 01-804 may include additional amino acids or substituents beyond those in SEQ ID NO.s: 01-804 where such additional amino acids or substituents do not prevent the polypeptide from being detectable.

[0058] Target molecules also include biomarkers. As used herein, the term “biomarker” includes any target molecule present in a GLF or fecal sample that is associated with a disease, predisposition to disease, positive response to a particular treatment regimen, no response to a particular treatment regimen, or negative response to a particular treatment regimen. In some embodiments, a biomarker can be identified, measured and/or correlated with a diagnosis or prognosis of a disease.

[0059] In some embodiments, a target molecule is a component of a fluid of the subject selected from the group consisting of blood, saliva, gastric juice, hepatic secretion, bile, duodenal juice, and pancreatic juice. In some embodiments, a target molecule is expressed in the upper gastrointestinal tract of the subject, or the lower gastrointestinal tract of the subject. In some embodiments, a target molecule is expressed at a location in the subject selected from the group consisting of buccal cavity, esophagus, stomach, biliary tree, gallbladder, duodenum, jejunum, ileum, appendix, cecum, colon, rectum, and anal canal.

[0060] In some embodiments, a target molecule does not include a protein or other compound found in a GLF, for example, lactoferrin, eosinophil-derived neurotoxin, eosinophil cationic protein, bilirubin (Bil), alkaline phosphatase (ALP), aspartate aminotransferase, hemoglobin, or eosinophil peroxidase. In some embodiments, a target

molecule does not include a protein found in feces, for example, heptaglobulin, hemopexin, α -2-macroglobulin, cadherin-17, calprotectin, carcinoembryogenic antigen, metalloproteinase-1 (TIMP-1), S100A12, K-ras, or p53. In some embodiments, a target molecule does not include a protein found in pancreatic juice, for example, anterior gradient-2 (AGR2), insulin-like growth factor binding protein-2, CEACAM6, MUC1, CA19-9, serine proteinase-2 (PRSS2) preproprotein, pancreatic lipase-related protein-1 (PLRP1), chymotrypsinogen B (CTRB), elastase 3B (ELA3B), tumor rejection antigen (pg96), azurocidin, hepatocarcinoma-intestine-pancreas/pancreatitis-associated-protein I (HIP/PAP-I), matrix metalloproteinase-9 (MMP-9), oncogene DJ1 (DJ-1), or alpha-1B-glycoprotein precursor (A1BG).

Methods for characterizing target molecules

[0061] Some embodiments of the methods and compositions provided herein include characterizing a target molecule in a GLF or fecal sample. Characterizing a target molecule can include, for example, identifying a target molecule, detecting a target molecule, and/or quantifying a target molecule. Methods to identify, detect and quantify target molecule are well known in the art.

[0062] Some embodiments include identifying, determining the presence or absence of a target molecule, and/or quantifying a target molecule, wherein the target molecule comprises a peptide, polypeptide, and/or protein. Such target molecules may be characterized by a variety of methods such as immunoassays, including radioimmunoassays, enzyme-linked immunoassays and two-antibody sandwich assays as described herein. A variety of immunoassay formats, including competitive and non-competitive immunoassay formats, antigen capture assays and two-antibody sandwich assays also are useful (Self and Cook, (1996) Curr. Opin. Biotechnol. 7:60-65, incorporated by reference in its entirety). Some embodiments include one or more antigen capture assays. In an antigen capture assay, antibody is bound to a solid phase, and sample is added such that antigen, *e.g.*, a target molecule in GLF or a fecal sample, is bound by the antibody. After unbound proteins are removed by washing, the amount of bound antigen can be quantitated, if desired, using, for example, a radioassay (Harlow and Lane, (1988) Antibodies A Laboratory Manual Cold Spring Harbor Laboratory: New York, incorporated by reference in its entirety). Immunoassays can be performed under

conditions of antibody excess, or as antigen competitions, to quantitate the amount of antigen and, thus, determine a level of a target molecule in GLF or a fecal sample.

[0063] Enzyme-linked immunosorbent assays (ELISAs) can be useful in certain embodiments provided herein. An enzyme such as horseradish peroxidase (HRP), alkaline phosphatase (AP), β -galactosidase or urease can be linked, for example, to an anti-HMGB1 antibody or to a secondary antibody for use in a method of the invention. A horseradish-peroxidase detection system can be used, for example, with the chromogenic substrate tetramethylbenzidine (TMB), which yields a soluble product in the presence of hydrogen peroxide that is detectable at 450 nm. Other convenient enzyme-linked systems include, for example, the alkaline phosphatase detection system, which can be used with the chromogenic substrate p-nitrophenyl phosphate to yield a soluble product readily detectable at 405 nm. Similarly, a β -galactosidase detection system can be used with the chromogenic substrate o-nitrophenyl- β -D-galactopyranoside (ONPG) to yield a soluble product detectable at 410 nm, or a urease detection system can be used with a substrate such as urea-bromocresol purple (Sigma Immunochemicals). Useful enzyme-linked primary and secondary antibodies can be obtained from a number of commercial sources such as Jackson Immuno-Research (West Grove, Pa.) as described further herein.

[0064] In certain embodiments, a target molecule in GLF or a fecal sample can be detected and/or measured using chemiluminescent detection. For example in certain embodiments, specific antibodies to a particular target molecule are used to capture the target molecule present in the biological sample, *e.g.*, GLF or a fecal sample and an antibody specific for the target molecule-specific antibodies and labeled with an chemiluminescent label is used to detect the target molecule present in the sample. Any chemiluminescent label and detection system can be used in the present methods. Chemiluminescent secondary antibodies can be obtained commercially from various sources such as Amersham. Methods of detecting chemiluminescent secondary antibodies are known in the art.

[0065] Fluorescent detection also can be useful for detecting a target molecule in certain methods provided herein. Useful fluorochromes include, DAPI, fluorescein, Hoechst 33258, R-phycoyanin, B-phycoerythrin, R-phycoerythrin, rhodamine, Texas red and lissamine. Fluorescein or rhodamine labeled antibodies, or fluorescein- or rhodamine-labeled secondary antibodies can be useful in the invention.

[0066] Radioimmunoassays (RIAs) also can be useful in certain methods provided herein. Such assays are well known in the art. Radioimmunoassays can be performed, for example, with ^{125}I -labeled primary or secondary antibody (Harlow and Lane, *supra*, 1988).

[0067] A signal from a detectable reagent can be analyzed, for example, using a spectrophotometer to detect color from a chromogenic substrate; a radiation counter to detect radiation, such as a gamma counter for detection of ^{125}I ; or a fluorometer to detect fluorescence in the presence of light of a certain wavelength. Where an enzyme-linked assay is used, quantitative analysis of the amount of a target molecule can be performed using a spectrophotometer such as an EMAX Microplate Reader (Molecular Devices; Menlo Park, Calif.) in accordance with the manufacturer's instructions. The assays of the invention can be automated or performed robotically, if desired, and that the signal from multiple samples can be detected simultaneously.

[0068] In some embodiments, capillary electrophoresis based immunoassays (CEIA), which can be automated if desired, may be used to detect and/or measure the target molecule. Immunoassays also can be used in conjunction with laser-induced fluorescence as described, for example, in Schmalzing and Nashabeh, *Electrophoresis* 18:2184-93 (1997), and Bao, J. *Chromatogr. B. Biomed. Sci.* 699:463-80 (1997), each incorporated by reference in its entirety. Liposome immunoassays, such as flow-injection liposome immunoassays and liposome immunosensors, also can be used to detect target molecules or to determine a level of a target molecule according to certain methods provided herein (Rongen *et al.*, (1997) *J. Immunol. Methods* 204:105-133, incorporated by reference in its entirety).

[0069] Sandwich enzyme immunoassays also can be useful in certain embodiments. In a two-antibody sandwich assay, a first antibody is bound to a solid support, and the antigen is allowed to bind to the first antibody. The amount of a target molecule is quantitated by measuring the amount of a second antibody that binds to it.

[0070] Quantitative Western blotting also can be used to detect a target molecule or to determine a level of target molecule in a method provided herein. Western blots can be quantitated by well known methods such as scanning densitometry. As an example, protein samples are electrophoresed on 10% SDS-PAGE Laemmli gels. Primary murine monoclonal antibodies, for example, against a target molecule are reacted with the blot, and antibody binding confirmed to be linear using a preliminary slot blot

experiment. Goat anti-mouse horseradish peroxidase-coupled antibodies (BioRad) are used as the secondary antibody, and signal detection performed using chemiluminescence, for example, with the Renaissance chemiluminescence kit (New England Nuclear; Boston, Mass.) according to the manufacturer's instructions. Autoradiographs of the blots are analyzed using a scanning densitometer (Molecular Dynamics; Sunnyvale, Calif.) and normalized to a positive control. Values are reported, for example, as a ratio between the actual value to the positive control (densitometric index). Such methods are well known in the art as described, for example, in Parra *et al.*, J. Vasc. Surg. 28:669-675 (1998), incorporated herein by reference in its entirety.

[0071] As described herein, immunoassays including, for example, enzyme-linked immunosorbent assays, radioimmunoassays and quantitative western analysis, can be useful in some embodiments for detecting a target molecule or determining a level of a target molecule. Such assays typically rely on one or more antibodies. As would be understood by the skilled artisan, methods described herein can be used to readily distinguish proteins with alternative forms of post-translation modifications, *e.g.*, phosphorylated proteins, and glycosylated proteins.

[0072] Target molecules, such as protein target molecules, can be characterized by a variety of methods. Proteins, polypeptides and peptides can be isolated by a variety of methods well known in the art, such as protein precipitation, chromatography (*e.g.*, reverse phase chromatography, size exclusion chromatography, ion exchange chromatography, liquid chromatography), affinity capture, and differential extractions.

[0073] Isolated proteins can undergo enzymatic digestion or chemical cleavage to yield polypeptide fragments and peptides. Such fragments can be identified and quantified. A particularly useful method for analysis of polypeptide/peptide fragments and other target molecules is mass spectrometry (U.S. Pat. App. No. 20100279382, incorporated by reference in its entirety). A number of mass spectrometry-based quantitative proteomics methods have been developed that identify the proteins contained in each sample and determine the relative abundance of each identified protein across samples (Flory *et al.*, Trends Biotechnol. 20:S23-29 (2002); Aebersold, J. Am. Soc. Mass Spectrom. 14:685-695 (2003); Aebersold, J. Infect. Dis. 187 Suppl 2:S315-320 (2003); Patterson and Aebersold, Nat. Genet. 33 Suppl, 311-323 (2003); Aebersold and Mann, Nature 422:198-207 (2003); Aebersold, R. and Cravatt, Trends Biotechnol. 20:S1-

2 (2002); Aebersold and Goodlett, Chem. Rev. 101, 269-295 (2001); Tao and Aebersold, Curr. Opin. Biotechnol. 14:110-118 (2003), each incorporated by reference in its entirety). Generally, the proteins in each sample are labeled to acquire an isotopic signature that identifies their sample of origin and provides the basis for accurate mass spectrometric quantification. Samples with different isotopic signatures are then combined and analyzed, typically by multidimensional chromatography tandem mass spectrometry. The resulting collision induced dissociation (CID) spectra are then assigned to peptide sequences and the relative abundance of each detected protein in each sample is calculated based on the relative signal intensities for the differentially isotopically labeled peptides of identical sequence.

[0074] More techniques for identifying and quantifying target molecules label-free quantitative proteomics methods. Such methods include: (i) sample preparation including protein extraction, reduction, alkylation, and digestion; (ii) sample separation by liquid chromatography (LC or LC/LC) and analysis by MS/MS; (iii) data analysis including peptide/protein identification, quantification, and statistical analysis. Each sample can be separately prepared, then subjected to individual LC-MS/MS or LC/LC-MS/MS runs (Zhu W. *et al.*, J. of Biomedicine and Biotech. (2010) Article ID 840518, 6 pages, incorporated by reference in its entirety). An example technique includes LC-MS in which the mass of a peptide coupled with its corresponding chromatographic elution time as peptide properties that uniquely define a peptide sequence, a method termed the accurate mass and time (AMT) tag approach. Using LC coupled with Fourier transform ion cyclotron resonance (LC-FTICR) MS to obtain the chromatographic and high mass accuracy information, peptide sequences can be identified by matching the AMT tags to previously acquired LC-MS/MS sequence information stored in a database. By taking advantage of the observed linear correlation between peak area of measured peptides and their abundance, these peptides can be relatively quantified by the signal intensity ratio of their corresponding peaks compared between MS runs (Tang, K., *et al.*, (2004) J. Am. Soc. Mass Spectrom. 15:1416–1423; and Chelius, D. and Bondarenko, P. V. (2002) J. Proteome Res. 1: 317–323, incorporated by reference in their entireties). Statistics tools such as the Student's t-test can be used to analyse data from multiple LC-MS runs for each sample (Wiener, M. C., *et al.*, (2004) Anal. Chem. 76:6085–6096, incorporated by reference in its entirety). At each point of acquisition time and m/z, the amplitudes of

signal intensities from multiple LC-MS runs can be compared between two samples to detect peptides with statistically significant differences in abundance between samples.

[0075] As will be understood, a variety of mass spectrometry systems can be employed in the methods for identifying and/or quantifying a polypeptide/peptide fragments. Mass analyzers with high mass accuracy, high sensitivity and high resolution include, ion trap, triple quadrupole, and time-of-flight, quadrupole time-of-flight mass spectrometers and Fourier transform ion cyclotron mass analyzers (FT-ICR-MS). Mass spectrometers are typically equipped with matrix-assisted laser desorption (MALDI) or electrospray ionization (ESI) ion sources, although other methods of peptide ionization can also be used. In ion trap MS, analytes are ionized by ESI or MALDI and then put into an ion trap. Trapped ions can then be separately analyzed by MS upon selective release from the ion trap. Fragments can also be generated in the ion trap and analyzed. Sample molecules such as released polypeptide/peptide fragments can be analyzed, for example, by single stage mass spectrometry with a MALDI-TOF or ESI-TOF system. Methods of mass spectrometry analysis are well known to those skilled in the art (*see, e.g.*, Yates, J. (1998) *Mass Spect.* 33:1-19; Kinter and Sherman, (2000) *Protein Sequencing and Identification Using Tandem Mass. Spectrometry*, John Wiley & Sons, New York; and Aebersold and Goodlett, (2001) *Chem. Rev.* 101:269-295, each incorporated by reference in its entirety).

[0076] For high resolution polypeptide fragment separation, liquid chromatography ESI-MS/MS or automated LC-MS/MS, which utilizes capillary reverse phase chromatography as the separation method, can be used (Yates *et al.*, *Methods Mol. Biol.* 112:553-569 (1999), incorporated by reference in its entirety). Data dependent collision-induced dissociation (CID) with dynamic exclusion can also be used as the mass spectrometric method (Goodlett, *et al.*, *Anal. Chem.* 72:1112-1118 (2000), incorporated by reference in its entirety).

[0077] Once a peptide is analyzed by MS/MS, the resulting CID spectrum can be compared to databases for the determination of the identity of the isolated peptide. Methods for protein identification using single peptides have been described previously (Aebersold and Goodlett, *Chem. Rev.* 101:269-295 (2001); Yates, J. *Mass Spec.* 33:1-19 (1998), David N. *et al.*, *Electrophoresis*, 20 3551-67 (1999), each incorporated by reference in its entirety). In particular, it is possible that one or a few peptide fragments can be used to identify a parent polypeptide from which the fragments were derived if the

peptides provide a unique signature for the parent polypeptide. Moreover, identification of a single peptide, alone or in combination with knowledge of a site of glycosylation, can be used to identify a parent glycopolypeptide from which the glycopeptide fragments were derived. As will be understood, methods that include MS can be used to characterize proteins, fragments thereof, as well as other types of target molecules described herein.

[0078] Some embodiments can include enriching proteins and/or protein fractions of a GLF. Example methods can include protein precipitation, chromatography, such as reverse phase chromatography, size exclusion chromatography, ion exchange chromatography, liquid chromatography, as well as affinity capture, differential extraction methods and centrifugation. Proteins and/or protein fractions can be further examined using intact protein methods such as top-down proteomics or gel chromatography such as SDS-PAGE.

[0079] Some embodiments include identifying, determining the presence or absence of a target molecule, and/or quantifying a target molecule, wherein the target molecule comprises a glycosylated protein and/or glycan. Glycosylated proteins and glycans can be analyzed by various methods well known in the art. Changes in glycosylation can be indicative of a disease or disease state. Thus, particular target molecules can include particular glycosylated proteins and/or glycans. As will be understood, a glycan may be a component of a glycoprotein, proteoglycan or other glycan containing compounds.

[0080] Some embodiments include identifying, determining the presence or absence of a target molecule, and/or quantifying a target molecule, wherein the target molecule comprises a metabolite. Metabolites may be analyzed in a GLF or fecal sample using a variety of methods. For example, a GLF or fecal sample can be analyzed for metabolites using methods such as chromatography. Some components of the metabolome include bile acids and other small organic compounds. Metabolites can include peptides that are present in a GLF or fecal sample.

Methods for identifying biomarkers

[0081] In some embodiments, the target molecules detected in GLF or a fecal sample can be evaluated to determine whether they are biomarkers associated with a particular condition, such as a disease, or physiological state. Such biomarkers can be indicative for a particular disease, predisposition to disease, prognosis, positive response

to a particular treatment regimen, or negative response to a particular treatment regimen. In some embodiments, the presence or absence, or level of a biomarker can be associated with a particular condition, such as a disease, or physiological state. In some embodiments, the presence or absence, or level of a biomarker can be statistically correlated to the particular condition, such as a disease, or physiological state. In some embodiments, a physiological state can include a disease. In some embodiments, a biomarker can be correlated to a particular condition, such as disease, or physiological state by comparing the level of expression of a biomarker in a subject having a condition, such as a disease, or physiological state with the level of expression of the biomarker in a subject not having a condition or physiological state.

[0082] In some embodiments, the differential expression of a biomarker in a subject having a condition compared to the expression of a biomarker in a subject not having a condition is indicative of a condition or physiological state. As used herein, “differential expression” refers to a difference in the level of expression of a biomarker in a subject having a condition, such as a disease, or physiological state and a subject not having the condition, such as a disease, or physiological state. For example, the term “differential expression” can refer to the presence or absence of a biomarker in a subject having a condition, such as a disease, or physiological state compared with a subject not having a condition or physiological state. In some embodiments, differential expression can refer to a difference in the level of expression of a biomarker in a subject having a condition, such as disease, or physiological state compared with the level of expression of a biomarker in a subject not having the condition, such as a disease, or physiological state.

[0083] Differences in the level of a biomarker can be determined by measuring the amount or level of expression of a biomarker using methods provided herein. In some embodiments, differential expression can be determined as the ratio of the levels of one or more biomarker products between reference subjects/populations having or not having a condition or physiological state, wherein the ratio is statistically significant. Differential expression between populations can be determined to be statistically significant as a function of p-value. When using p-value to determine statistical significance, a biomarker, the p-value is preferably less than 0.2. In another embodiment, the biomarker is identified as being differentially expressed when the p-value is less than 0.15, 0.1, 0.05, 0.01, 0.005, 0.0001 etc. When determining differential expression on the basis of the ratio, a biomarker product is differentially expressed if the

ratio of the level of expression in a first sample as compared with a second sample is greater than or less than 1.0. For example, a ratio of greater than 1.0 for example includes a ratio of greater than 1.1, 1.2, 1.5, 1.7, 2, 3, 4, 10, 20 and the like. A ratio of less than 1.0, for example, includes a ratio of less than 0.9, 0.8, 0.6, 0.4, 0.2, 0.1, 0.05 and the like. In another embodiment, a biomarker can be differentially expressed if the ratio of the mean of the level of expression of a first population as compared with the mean level of expression of the second population is greater than or less than 1.0. For example, a ratio of greater than 1.0 includes a ratio of greater than 1.1, 1.2, 1.5, 1.7, 2, 3, 4, 10, 20 and the like and a ratio less than 1.0, for example includes a ratio of less than 0.9, 0.8, 0.6, 0.4, 0.2, 0.1, 0.05 and the like. In another embodiment a biomarker is differentially expressed if the ratio of its level of expression in a first sample as compared with the mean of the second population is greater than or less than 1.0 and includes for example, a ratio of greater than 1.1, 1.2, 1.5, 1.7, 2, 3, 4, 10, 20, or a ratio less than 1, for example 0.9, 0.8, 0.6, 0.4, 0.2, 0.1, 0.05.

[0084] In some embodiments, a biomarker can be identified by measuring the level of at least 1 target molecule in a test GLF or test fecal sample from at least one test subject having a condition or physiological state and a control GLF or control fecal sample from at least 1 control subject not having the condition or physiological state; comparing the level of the at least 1 target molecule in the test GLF or test fecal sample with the level of the at least 1 target molecule in the control GLF or control fecal sample, wherein a significant difference in the level of the at least 1 target identifies a biomarker. Some embodiments include measuring and comparing a plurality of target molecules in a test GLF or test fecal sample from plurality of test subjects having a condition or physiological state and a control GLF or control fecal sample from plurality of control subjects not having a condition or physiological state. In some embodiments, at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 target molecules can be measured and compared. In some embodiments, a GLF or fecal sample can be obtained from at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 test subjects. In some embodiments, a GLF or fecal sample can be obtained from at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 control subjects. In some embodiments, a significant difference in the level of a target molecule in a test GLF or a test fecal sample compared to a control GLF or a control fecal sample can be a statistically significant.

Kits

[0085] Some embodiments of the methods and compositions provided herein relate to kits for detecting a target molecule in a GLF or fecal sample, determining the presence or absence of a target molecule in a GLF or fecal sample, quantifying a target molecule in a GLF or fecal sample, or identifying a target molecule in a GLF or fecal sample. Some such kits can include a lavage composition for oral administration to a subject. In some embodiments, the lavage fluid can include an ingredient such as polyethylene glycol, magnesium sulfate, sodium sulfate, potassium sulfate, magnesium citrate, and bisacodyl. In some embodiments, the lavage fluid can include polyethylene glycol with an electrolyte solution, optionally also including bisacodyl, or ascorbic acid (*e.g.*, GOLYTELY, HALFLYTELY, NULYTELY, MOVIPREP). In some embodiments, the lavage fluid can include phosphate salts (*e.g.* FLEET'S PHOSPHO-SODA). In some embodiments, the lavage fluid can include sulfate salts such as sodium sulfate, magnesium sulfate, or potassium sulfate (*e.g.*, SUPREP). In some embodiments, the lavage fluid can include magnesium citrate. In some embodiments, the lavage fluid can include sodium picosulfate.

[0086] In some embodiments, a kit can also include a vessel for collecting a GLF and/or fecal sample from a subject. A vessel for collecting a GLF can include an insert for a toilet which captures the GLF or fecal sample and the like. In some embodiments, the vessel can include a material to stabilize and/or preserve a target molecule, such as one or more isolated protease inhibitors. In some embodiments, the vessel can include an agent for detecting a target molecule, determining the presence or absence of a target molecule, quantifying a target molecule or identifying a target molecule.

Diseases

[0087] Some embodiments of the methods and compositions provided herein relate to the diagnosis, prognosis for a particular disease. Some embodiments include diseases and disorders related to the GI tract and organs associated therewith. Example diseases include cancers of the GI tract and organs associated therewith, *e.g.*, gastric cancer, liver cancer, pancreatic cancer. More examples of diseases include pancreatitis, pancreatic adenocarcinoma, gastrointestinal neuroendocrine tumors, gastric

adenocarcinoma, colon adenocarcinoma, hepatocellular carcinoma, cholangiocarcinoma, gallbladder adenocarcinoma, ulcerative colitis, and Crohn's disease. Some diseases relate to an inflammatory bowel disease (IBD). As used herein, the term “inflammatory bowel disease” can refer to a broad class of diseases characterized by inflammation of at least part of the gastrointestinal tract. IBD symptoms may include inflammation of the intestine and resulting in abdominal cramping and persistent diarrhea. Inflammatory bowel diseases include ulcerative colitis (UC), Crohn's disease (CD), indeterminate colitis, chronic colitis, discontinuous or patchy disease, ileal inflammation, extracolonic inflammation, granulomatous inflammation in response to ruptured crypts, aphthous ulcers, transmural inflammation, microscopic colitis, diverticulitis and diversion colitis. More examples of diseases include celiac sprue, malabsorption disorders, and other conditions of digestive tract, liver, pancreas, and biliary tree.

[0088] Some embodiments of the methods and compositions provided herein relate to determining the selection of a treatment (often referred to as personalized medicine), a subject's positive response to treatment, negative response to treatment, or lack of response to treatment. Some such embodiments include determining a patient's partial response to a treatment regimen. For example, the presence of a biomarker, absence of a biomarker, or level of a biomarker can be determined in a GLF or fecal sample from a subject at a first time point. At a second time point after treatment has begun and/or treatment has been completed, the presence of the biomarker, absence of the biomarker, or level of the biomarker can be determined in a GLF or fecal sample from the subject. The difference in the presence of the biomarker, absence of the biomarker, or level of the biomarker in the GLF or fecal sample at the second time point compared with the presence of the biomarker, absence of the biomarker, or level of the biomarker in the GLF or fecal sample from the first time point can be indicative of the subject's positive response to treatment, negative response to treatment, partial response to treatment, or lack of response to treatment. Alternatively, subjects can be given a treatment regimen and categorized as having a positive response, negative response, partial response, or no response. The presence, absence, or level of a target molecule in each group of subjects can be determined and those target molecules having a statistically significant association with each category of response can be identified. Some embodiments also include determining a future treatment regimen to be provided to a subject in view of determining the subject's positive response, negative response, partial response, or no response to a

former or current treatment regimen. Accordingly, a former or current treatment regimen can be modified based on determinations made by the methods provided herein.

[0089] More embodiments include methods for determining a subject's physiological status by evaluating a plurality of biomarkers. Some such methods include determining the presence, absence and/or levels of a plurality of biomarkers. The presence, absence and/or levels of a plurality of biomarkers can be correlated to the likelihood of the subject's physiological status, such as the subject's likelihood of developing a disease, and/or a subject's likely response to a treatment regimen to treat a particular disease. In some such methods, a subject's "clinical risk score" can be determined by correlating the presence, absence and/or levels of a plurality of biomarkers to determine the likelihood that a subject has a disease or will develop a disease (*see, e.g.,* Soonmyung P. *et al.*, (2004) *New Eng. J. of Medicine* 351:2817-2826; and Cho C.S. *et al.*, (2008) *J. Am. Coll. Surg.* 206:281-291, incorporated by reference herein in their entireties).

[0090] While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

EXAMPLES

Example 1—Proteomic analysis of sulfate-based GLF

[0091] In this analysis, the ability of a sulfate-based GLF to support proteomic analysis was assessed. To identify target molecules in a GLF obtaining using a sulfate-based lavage composition, SUPREP was administered to three human subjects, and proteins in the resultant GLF were analyzed by mass spectroscopy. In this example, the GLF was collected from subjects as part of a colonoscopy procedure.

[0092] Upon collection, a complete protease inhibitor tablet (ROCHE) was added and samples were spun at 1000 rpm for 30 minutes at 4 °C. Supernatants were spun again at 14,000 x g for 30 min to pellet bacteria and debris. 1.8 ml of supernatant was precipitated with 6 volumes of acetone followed by extraction with an equal volume of chloroform followed by separation on a C-2 reverse phase SPE column (Sep-Pak, Waters). The column was washed with 3 column volumes each of 0.1% trifluoroacetic acid (TFA), 10%, 20%, and 30% acetonitrile (ACN) in 0.1% TFA, and eluted with 3

column volumes of 60% ACN in 0.1% TFA. Samples were dried by centrifugal lyophilization, resuspended in 100 μ l of 50 mM ammonium bicarbonate / 10 mM tris (2-carboxyethyl) phosphine and digested with 2 μ l of 10 mM sequencing-grade trypsin (Promega, Madison, WI).

[0093] Data were acquired on an LTQ-Orbitrap mass spectrometer using input from an LC system. The A solvent contained 3% of B and 0.2% formic acid in water. The B solvent contained 3% of A and 0.2% formic acid in acetonitrile. Solvents were HPLC grade from Fisher. For a 120 min run, the starting solvent was 5% B and remains for 7 min. The gradient was changed to 10% by 13 min, 40% by 83 min, 90% by 103 min, then reduced from 90% to 5% at 111 min. It was then re-equilibrated for the next injection. Three injections were performed for each sample for repeatability determination.

[0094] The MS was scanned (Orbitrap) over the mass range from 400 m/z to 2000 m/z every second while the LTQ (Trap) acquired up to 5 MSMS (peptide sequence) spectra in parallel. Data were acquired using the standard Thermo Xcalibur software. MS data (Orbitrap) was stable to 2-3 ppm and a background ion was used for mass drift assessment. MSMS data (LTQ) was measured to approximately 0.6 Da but the parent mass was acquired from the low ppm Orbitrap data. Peptides were eluted from a C18 LC column using triplicate injections to ensure reliability and repeatability of the data. A search file was created from the triplicate injections from each lavage preparation (patient sample) and converted into a MGF (Mascot Generic Format) file using a combination of Xcalibur and Mascot software packages.

[0095] Database searching was done using the Mascot search engine (Matrix Science, UK) against the RefSeq database (<http://www.ncbi.nlm.nih.gov/RefSeq/>) with taxonomy specified as human (homo sapiens), a mass accuracy of 10 ppm for the parent ion (MS) and 0.6 Da for the fragment ions (MS/MS), and “no enzyme” selected. Searching without enzyme specificity was performed due to the presence of digestive enzymes in the sample that may modify or truncate peptides being examined. The RefSeq database was supplemented by the addition of antibody sequences that are included in the SwissProt protein database, as these antibody sequences are not part of the standard RefSeq listing.

[0096] Higher Mascot scores indicate better proteins hits and can be correlated to relative protein levels. A score threshold of “>40” was indicative of a p-value

significance of <0.05 as determined by the Mascot scoring system based on the search of this database with no enzyme specificity; a score of 40 is consistent with a $p < 0.01$. Standard Mascot scoring was used whereby only the highest score was added for each peptide detected, even if it was sampled during MS/MS multiple times. For all data included, scores were all > 40 in at least one sample per protein line. For additional confidence, the numbers of significant peptides were also reported and a minimum criterion of at least 2 peptides was selected. Very few had less than 3 peptides. All significant peptides counted represented different sequences (individual peptides) from their respective proteins. The score and numbers of significant peptides are reported in the format x/y where x is the score and y the number of significant peptides. If a protein was not detected in a particular sample it is listed as “ND”. Proteins are reported as protein name and the “gi” number defined by the protein database of the NCBI has been provided. The sequences contained in each of the “gi” numbers in the NCBI database listed throughout the present application are incorporated herein by reference. Where a protein is named in its preprotein or other non-mature form, the mature form of the protein is equally implied including such changes as removal of signal sequences and the addition of post-translational modifications. In all cases, the protein has been named by its gene derived sequence to provide consistency.

[0097] Table 1 lists examples of the most abundant proteins identified in GLF from three separate patients defined as patients 3, 4 and 6, presented in the format described above. As can be seen from Table 1, many proteins can be identified from GLF and a large number of these may be associated with the pancreas. Other proteins include DMBT1 (gi # 148539840) which may be associated with colon cancer and other cancers. Antibodies and putative glycosylated proteins were also identified.

TABLE 1

NCBI gi #	Protein	Mascot score / number of significant peptides		
		Sample 3	Sample 4	Sample 6
10835000	pancreatic lipase precursor	2224 / 24	1238 / 13	2926 / 34
4506147	protease serine 2 preproprotein	665 / 7	46 / 0	1189 / 11

NCBI gi #	Protein	Mascot score / number of significant peptides		
		Sample 3	Sample 4	Sample 6
4506145	protease serine 1 preproprotein	239 / 2	64 / 0	1002 / 11
29725633	regenerating islet-derived 1 alpha precursor	231 / 2	132 / 1	802 / 6
6679625	elastase 3B pancreatic preproprotein	852 / 12	1144 / 11	772 / 8
236460050	elastase 3A pancreatic preproprotein	1291 / 17	1306 / 14	769 / 7
118498350	chymotrypsin B2	945 / 10	244 / 2	724 / 6
15559207	elastase 2A preproprotein	752 / 10	952 / 8	593 / 6
54607080	pancreatic carboxypeptidase B1 preproprotein	702 / 9	84 / 1	499 / 4
50363217	serine proteinase inhibitor clade A member 1	655 / 9	406 / 3	490 / 2
10280622	amylase pancreatic alpha-2B precursor	ND	ND	388 / 2
4505847	phospholipase A2 group IB	258 / 3	639 / 8	384 / 5
4502085	pancreatic amylase alpha 2A precursor	95 / 1	193 / 3	365 / 2
4502997	carboxypeptidase A1 precursor	454 / 4	88 / 1	349 / 5
62526043	chymotrypsin C preproprotein	696 / 9	440 / 3	343 / 4
148539840	deleted in malignant brain tumors 1 isoform a precursor (DMBT1)	88 / 1	101 / 1	280 / 3
31377806	polymeric immunoglobulin receptor precursor	566 / 7	ND	279 / 3
41152086	serine (or cysteine) proteinase inhibitor clade B (ovalbumin) member 6	ND	269 / 3	276 / 2

NCBI gi #	Protein	Mascot score / number of significant peptides		
		Sample 3	Sample 4	Sample 6
148539842	deleted in malignant brain tumors 1 isoform b precursor	ND	ND	275 / 3
4507149	superoxide dismutase 1 soluble	ND	87 / 1	253 / 3
113584	RecName: Full=Ig alpha-1 chain C region	940 / 10	53 / 1	204 / 2
125145	RecName: Full=Ig kappa chain C region	659 / 9	106 / 1	180 / 1
98986445	carcinoembryonic antigen-related cell adhesion molecule 5 preproprotein	ND	219 / 3	135 / 0
218512088	RecName: Full=Ig alpha-2 chain C region	886 / 9	ND	ND
119395750	keratin 1	ND	499 / 7	ND
55956899	keratin 9	ND	395 / 3	ND

[0098] In another experiment, SUPREP was administered to a subject according to the manufacturer's guidelines and the resultant GLF was self-collected by the subject into a collection container placed in the toilet immediately prior to colonoscopy. The proteome of the GLF was analyzed by MS as described above. The results showing the Mascot scores for the most abundant species present are summarized in Table 2. The results indicated some urinary contamination. A similar proteomic profile was observed for a sample collected subsequently during colonoscopy. Table 2 shows that many different proteins were identified in GLF collected by a subject. Identified proteins included DMBT1, pancreatic proteins and antibodies, consistent with data in Table 1.

TABLE 2

NCBI gi #	Protein	Mascot score / number of significant peptides
		Sample 25
148539842	deleted in malignant brain tumors 1 isoform b precursor	1184 / 13
119395750	keratin 1	742 / 9
10835000	pancreatic lipase precursor	538 / 8
113584	RecName: Full=Ig alpha-1 chain C region	506 / 6
31377806	polymeric immunoglobulin receptor precursor	474 / 5
98986445	carcinoembryonic antigen-related cell adhesion molecule 5 preproprotein	424 / 5
125817	RecName: Full=Ig kappa chain V-III region HAH; Flags: Precursor	382 / 5
125797	RecName: Full=Ig kappa chain V-III region SIE	341 / 5
54607080	pancreatic carboxypeptidase B1 preproprotein	340 / 5
236460050	elastase 3A pancreatic preproprotein	328 / 4
125145	RecName: Full=Ig kappa chain C region	327 / 5
4502027	albumin preproprotein	319 / 3
33456	immunoglobulin M chain	239 / 3
118498350	chymotrypsin B2	238 / 3
125811	RecName: Full=Ig kappa chain V-III region VG; Flags: Precursor	237 / 3

NCBI gi #	Protein	Mascot score / number of significant peptides
		Sample 25
123843	RecName: Full=Ig heavy chain V-III region VH26; Flags: Precursor	219 / 3
157266300	membrane alanine aminopeptidase precursor	213 / 3
563454	Ig heavy chain (VH4) V region (VDJ)	206 / 3
125788	RecName: Full=Ig kappa chain V-II region TEW	202 / 3
4506147	protease serine 2 preproprotein	177 / 3
125809	RecName: Full=Ig kappa chain V-III region CLL; AltName: Full=Rheumatoid factor; Flags: Precursor	149 / 3

[0099] The foregoing analyses demonstrate that a large number of target molecules can be detected in samples obtained using a sulfate-based GLF.

Example 2—Proteomic analysis of polyethylene glycol based GLF

[0100] In this analysis, the ability of a polyethylene glycol based GLF to support proteomic analysis was assessed. To identify target molecules in a GLF obtaining using a polyethylene-based lavage composition, a polyethylene glycol-based lavage composition was administered to two human subjects, and proteins in the resultant GLF were analyzed by mass spectrometry as described in Example 1. Removal of the polyethylene glycol was largely achieved by chloroform extraction of the lavage fluid. Many different proteins were identified in the GLFs from these subjects administered a polyethylene glycol-based lavage composition. Examples of the most abundant identified proteins identified, which are consistent with those observed in previous tables, are presented in Table 3.

TABLE 3

NCBI gi#	Protein	Mascot score / number of significant peptides	
		Sample 1	Sample 5
98986445	carcinoembryonic antigen-related cell adhesion molecule 5 preproprotein	767 / 10	127 / 1
4502085	pancreatic amylase alpha 2A precursor	619 / 8	ND
1684927	immunoglobulin light chain	586 / 6	152 / 1
50363217	serine proteinase inhibitor clade A member 1	550 / 4	268 / 2
40254482	salivary amylase alpha 1A precursor	548 / 6	ND
298351713	RecName: Full=Ig lambda-1 chain C regions	501 / 6	139 / 1
4507725	transthyretin precursor	477 / 5	ND
236460050	elastase 3A pancreatic preproprotein	432 / 5	84 / 0
4502997	carboxypeptidase A1 precursor	412 / 5	106 / 1
113584	RecName: Full=Ig alpha-1 chain C region	404 / 4	214 / 3
4885165	cystatin A	352 / 4	66 / 1
40255013	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	349 / 3	ND
4506147	protease serine 2 preproprotein	326 / 5	434 / 6
125145	RecName: Full=Ig kappa chain C region	291 / 4	265 / 5
218512088	RecName: Full=Ig alpha-2 chain C region	272 / 3	ND
121039	RecName: Full=Ig gamma-1 chain C region	265 / 3	ND
54607080	pancreatic carboxypeptidase B1 preproprotein	263 / 2	ND

[0101] In another experiment, a PEG-based lavage composition was administered to a subject and the subject self-collected the resultant GLF into a collection container placed in the toilet immediately prior to colonoscopy. The proteome of the GLF was analyzed by MS as described herein. Many different proteins were identified in the self collected GLF sample. Examples of the most abundant identified proteins, and the corresponding Mascot scores and numbers of significant peptides for each protein are listed in Table 4. The more extensive protein list showed evidence of urinary

contamination. A similar proteomic profile was observed for a sample collected subsequently during colonoscopy.

TABLE 4

NCBI gi #	Protein	Mascot score / number of significant peptides
		Sample 26
50363217	serine proteinase inhibitor clade A member 1	406 / 3
4885165	cystatin A	294 / 3
4502027	albumin preproprotein	287 / 3
55956899	keratin 9	287 / 4
4506147	protease serine 2 preproprotein	227 / 3
4502085	pancreatic amylase alpha 2A precursor	217 / 3

[0102] The proteomes of GLFs resultant from the administration of sulfate-based lavage compositions and either collected during as part of a colonoscopy procedure or self-collected by a subject were compared. The Mascot scores and number of significant peptides for the most abundant proteins are summarized in Table 5. While there was a close correlation between the two proteomes observed, different isoforms were identified for at least two proteins. The selection of different isoforms may be a result of the collection of the sequence data during MS/MS and the search engine. There were fewer proteins detected in the self collected sample which was more dilute than that collected during the colonoscopy.

TABLE 5

NCBI gi #	Protein	Mascot score / number of significant peptides	
		Colonoscopy collected sample	Subject collected sample
148539842	deleted in malignant brain tumors 1 isoform b precursor	See isoform a	1184 / 13

NCBI gi #	Protein	Mascot score / number of significant peptides	
		Colonoscopy collected sample	Subject collected sample
148539840	deleted in malignant brain tumors 1 isoform a precursor	417 / 4	See isoform b
119395750	keratin 1	159 / 2	742 / 9
10835000	pancreatic lipase precursor	3719 / 37	538 / 8
113584	RecName: Full=Ig alpha-1 chain C region	958 / 9	506 / 6
31377806	polymeric immunoglobulin receptor precursor	469 / 4	474 / 5
98986445	carcinoembryonic antigen-related cell adhesion molecule 5 preproprotein	74 / 1	424 / 5
125817	RecName: Full=Ig kappa chain V- III region HAH; Flags: Precursor	ND	382 / 5
125797	RecName: Full=Ig kappa chain V- III region SIE	ND	341 / 5
54607080	pancreatic carboxypeptidase B1 preproprotein	1389 / 17	340 / 5
236460050	elastase 3A pancreatic preproprotein	2268 / 27	328 / 4
125145	RecName: Full=Ig kappa chain C region	734 / 7	327 / 5
118498350	chymotrypsin B2	See chymotrypsin B1	238 / 3
118498341	chymotrypsin B1	881 / 7	See chymotrypsin B2
62526043	chymotrypsin C preprotein	1002 / 10	ND
125811	RecName: Full=Ig kappa chain V- III region VG; Flags: Precursor	ND	237 / 3

NCBI gi #	Protein	Mascot score / number of significant peptides	
		Colonoscopy collected sample	Subject collected sample
1684927	immunoglobulin light chain	371 / 4	221 / 2
123843	RecName: Full=Ig heavy chain V-III region VH26; Flags: Precursor	181 / 2	219 / 3
157266300	membrane alanine aminopeptidase precursor	544 / 5	213 / 3
125788	RecName: Full=Ig kappa chain V-II region TEW	ND	202 / 3
4502085	pancreatic amylase alpha 2A precursor	4258 / 49	ND
10280622	amylase pancreatic alpha-2B precursor	3916 / 45	ND
6679625	elastase 3B pancreatic preproprotein	1955 / 22	ND
4506147	protease serine 2 preproprotein	1442 / 14	ND
4502997	carboxypeptidase A1 precursor	1168 / 14	ND
15559207	elastase 2A preproprotein	959 / 5	ND
217416390	carboxypeptidase A2 (pancreatic) precursor	811 / 10	ND
29725633	regenerating islet-derived 1 alpha precursor	714 / 8	ND
218512088	RecName: Full=Ig alpha-2 chain C region	663 / 6	ND
7669492	glyceraldehyde-3-phosphate dehydrogenase	593 / 5	ND
4506145	protease serine 1 preproprotein	551 / 6	ND
157266300	membrane alanine aminopeptidase precursor	544 / 5	ND

NCBI gi #	Protein	Mascot score / number of significant peptides	
		Colonoscopy collected sample	Subject collected sample
298351713	RecName: Full=Ig lambda-1 chain C regions	322 / 4	ND
119220569	zymogen granule membrane glycoprotein 2 isoform 1	304 / 3	ND
51593090	mucin 13 epithelial transmembrane	293 / 3	ND
125807	RecName: Full=Ig kappa chain V- III region GOL; AltName: Full=Rheumatoid factor	288 / 3	ND
10334859	creatine kinase mitochondrial 1B precursor	265 / 3	ND

[0103] The foregoing analyses demonstrate that a large number of target molecules can be detected in samples obtained using a polyethylene glycol based GLF.

Example 3—Proteomic analysis of magnesium citrate based GLF

[0104] In this analysis, the ability of a magnesium citrate based GLF to support proteomic analysis was assessed. To identify target molecules in a GLF from a human subject administered a magnesium citrate-based lavage composition, a magnesium citrate-based lavage composition was administered to a subject; the GLF was collected from subject as part of a colonoscopy procedure. The proteome of the GLF was analyzed by mass spectroscopy as described in Example 1. Many different proteins were identified in the GLF. Examples of the most abundant identified proteins are listed in Table 6. Many of the identified proteins were detected with different colonoscopy preparations suggesting that the proteome is not dependent on the bowel preparation used.

TABLE 6

NCBI gi #	Protein	Mascot score / number of significant peptides
		Sample 27
4502085	pancreatic amylase alpha 2A precursor	2977 / 31
10280622	amylase pancreatic alpha-2B precursor	2891 / 30
40254482	salivary amylase alpha 1A precursor	2472 / 26
50363217	serine proteinase inhibitor clade A member 1	1316 / 12
236460050	elastase 3A pancreatic preproprotein	1299 / 13
15559207	elastase 2A preproprotein	1109 / 11
6679625	elastase 3B pancreatic preproprotein	987 / 10
29725633	regenerating islet-derived 1 alpha precursor	577 / 6
4507725	transthyretin precursor	570 / 7
4506147	protease serine 2 preproprotein	521 / 6
58331211	elastase 2B preproprotein	498 / 4
157266292	intestinal alkaline phosphatase precursor	491 / 7
113584	RecName: Full=Ig alpha-1 chain C region	338 / 3
125145	RecName: Full=Ig kappa chain C region	337 / 4
98986445	carcinoembryonic antigen-related cell adhesion molecule 5 preproprotein	329 / 3
218512088	RecName: Full=Ig alpha-2 chain C region	293 / 3

NCBI gi #	Protein	Mascot score / number of significant peptides
		Sample 27
41152086	serine (or cysteine) proteinase inhibitor clade B (ovalbumin) member 6	293 / 4
4506145	protease serine 1 preproprotein	267 / 3
4502997	carboxypeptidase A1 precursor	146 / 3

[0105] The foregoing analysis demonstrates that a large number of target molecules can be detected in samples obtained using a magnesium citrate based GLF.

Example 4—Detection of IgA-1 and IgA-2 Antibodies in samples obtained using a GLF in combination with SSL-7 enrichment

[0106] In this analysis, the ability of a samples obtained using a GLF in combination with SSL-7 enrichment to detect IgA-1 and IgA-2 antibodies was assessed. To identify target molecules in a GLF with affinity to *Staphylococcus aureus* superantigen-like protein 7 (SSL-7), a sulfate-based lavage composition (SUPREP) was administered to human subjects, and proteins were enriched in each GLF using SSL-7 affinity beads. The GLF was collected from subjects as part of a colonoscopy procedure.

[0107] SSL-7 affinity beads were used to isolate IgA-1 and IgA-2 specifically. 20 µl of SSL-7 agarose (Invitrogen, San Diego, CA) was added to 1 ml of sample and incubated overnight on a roller at 4°C. Tubes were spun at 1,000xg for 2 minutes to pellet beads and supernatant discarded. Beads were washed 4x with 1X phosphate buffered saline and eluted with 20 µl of 100 mM glycine, pH 2.7 in a shaker for 1 hr at 600 rpm and 37°C. Eluted antibodies were diluted with 60 µl of 100 mM ammonium bicarbonate/ 10 mM tris (2-carboxyethyl) phosphine and digested with 2 µl of sequencing grade trypsin (Promega, Madison,WI). Mass spectrometry and database searches were performed as described above. The most abundant identified proteins present in GLF with affinity to SSL-7 and their corresponding Mascot scores are summarized in Table 7. As has been observed in prior examples, antibodies were present in the GLF and

enrichment and analysis of these is possible using the affinity reagents, thus allowing the specific analysis of this subproteome in the GLF. The most abundant antibodies were IgAs. IgAs are consistently reported to be present in the intestinal tract.

TABLE 7

NCBI gi #	Protein	Mascot score / number of significant peptides	
		Sample 12	Sample 13
113584	RecName: Full=Ig alpha-1 chain C region	3084 / 29	2477 / 24
31377806	polymeric immunoglobulin receptor precursor	3065 / 36	898 / 13
218512088	RecName: Full=Ig alpha-2 chain C region	2736 / 24	2225 / 18
125145	RecName: Full=Ig kappa chain C region	477 / 4	542 / 6
21489959	immunoglobulin J chain	403 / 4	308 / 2
298351715	RecName: Full=Ig lambda-3 chain C regions	352 / 3	ND
298351713	RecName: Full=Ig lambda-1 chain C regions	317 / 3	581 / 6
1684927	immunoglobulin light chain	ND	636 / 6

[0108] The foregoing analysis demonstrates that IgA antibodies can be detected in samples obtained using a GLF in combination with SSL-7 enrichment.

Example 5—Detection of IgA and IgM in samples obtained using a GLF in combination with ProteinL enrichment

[0109] In this analysis, the ability of samples obtained using a GLF in combination with Protein L enrichment to detect IgA and IgM antibodies was assessed. To identify target molecules in a GLF with affinity to Protein L, a sulfate-based lavage composition (SUPREP) was administered to human subjects, and proteins were enriched

in each GLF using Protein L affinity beads. The GLF was collected from subjects as part of a colonoscopy procedure.

[0110] Protein L affinity beads were used to isolate antibodies containing kappa light chains. 20 µl of Protein L agarose (Santa Cruz Biotechnology, Santa Cruz, CA) was added to 1 ml of sample and incubated overnight on a roller at 4 C. Tubes were spun at 1,000 x g for 2 minutes to pellet beads and supernatant discarded. Beads were washed 4 times with 1X phosphate buffered saline and eluted with 20 µl of 100 mM glycine, pH 2.7 in a shaker for 1 hr at 600 rpm and 37 °C. Eluted antibodies were diluted with 60 µl of 100 mM Ammonium bicarbonate/ 10 mM tris (2-carboxyethyl) phosphine and digested with 2 µl of sequencing grade trypsin (Promega, Madison, WI). The most abundant identified proteins present in GLF with affinity to Protein L and their corresponding Mascot scores and numbers of significant peptides are summarized in Table 8. As expected, IgA and associated chains from antibodies were again detected. As the Protein L is not totally specific for IgA antibodies, an IgM antibody (gi# 193806374) was also detected.

TABLE 8

NCBI gi #	Protein	Mascot score / number of significant peptides
		Sample 9
113584	RecName: Full=Ig alpha-1 chain C region	1183 / 14
31377806	polymeric immunoglobulin receptor precursor	1149 / 14
218512088	RecName: Full=Ig alpha-2 chain C region	985 / 11
125145	RecName: Full=Ig kappa chain C region	654 / 8
193806374	RecName: Full=Ig mu chain C region	407 / 5
187950123	immunoglobulin heavy chain variable region	289 / 3
21489959	immunoglobulin J chain	249 / 2

[0111] The foregoing analysis demonstrates that IgA and IgM antibodies can be detected in samples obtained using a GLF in combination with ProteinL enrichment.

Example 6—Detection of proteins of bacterial origin in samples obtained using a GLF

[0112] In this analysis, the ability of a GLF to facilitate detection of proteins of bacterial origin was assessed. To identify target molecules associated with bacteria in a GLF, a sulfate-based lavage composition (SUPREP) was administered to two human subjects; the resultant GLF was collected from the subjects as part of a colonoscopy procedure. Super Optimal Broth (SOB) media was inoculated with 100 µl from each GLF and incubated overnight at 37 °C and 220 rpm shaking. Pellets were lysed in a bead-beater in 8M urea and lysates were diluted to 2 M urea in 50 mM ammonium bicarbonate/10 mM tris (2-carboxyethyl) phosphine and digested with sequencing grade trypsin (Promega, Madison, WI). Data were acquired on the Orbitrap MS system using 120 mins runs as described earlier. A MGF search file was created and searched with the Mascot search engine (Matrix Science, UK) against the RefSeq database (<http://www.ncbi.nlm.nih.gov/RefSeq/>) with the taxonomy specified as Eubacteria with a mass accuracy of 10ppm for the parent ion (MS) and 0.6 Da for the fragment ions (MS/MS). The most abundant identified proteins present in GLF associated with bacteria and their corresponding Mascot scores and numbers of significant peptides are summarized in Table 9. In the sample shown, the bacterium that was cultured was *Escherichia coli*. Other samples show different bacteria showing that the lavage fluid still retains some of the gut bacteria.

TABLE 9

NCBI gi #	Protein	Mascot score / number of significant peptides
		Bacterial isolate
15834378	chaperonin GroEL [<i>Escherichia coli</i>]	1294 / 3
15803852	elongation factor Tu [<i>Escherichia coli</i>]	1244 / 7

NCBI gi #	Protein	Mascot score / number of significant peptides
		Bacterial isolate
157159481	Molecular chaperone DnaK [Escherichia coli]	683 / 3
123444102	elongation factor Tu [Yersinia enterocolitica subsp. enterocolitica 8081]	518 / 3

[0113] The foregoing analysis demonstrates that proteins of bacterial origin can be detected in samples obtained using a GLF.

Example 7—Proteomic analysis of combined samples from several subjects

[0114] In order to further facilitate the identification of a large number of target proteins detectable in samples obtained using a GLF, the search files generated from the data acquired individually from 12 subjects were concatenated into a single search file and searched using the previously specified parameters for the Orbitrap data. Many proteins were analyzed in various GLF samples and proteins selected that (predominantly) had at least 3 unique significant peptides detected with thresholds of $p < 0.05$ (Mascot score approximately 41). Only 3 listed proteins (gi's 5031863, 6466801 and 115430223) had less than 3 significant peptides but these had Mascot scores of approximately 400, well above the 95% confidence level for protein identification. Proteins identified in this combined analysis are listed in Table 10 along with reported origins of particular proteins and reported associated cancers. Table 10 also lists SEQ ID NO.s. for identified peptides that had Mascot scores of 40 or greater for each unique identified protein. Many identified proteins have been reported to be present in pancreatic juice. References listed in Table 10 are provided in this application.

TABLE 10

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
10835000	pancreatic lipase precursor	6919/75	1-77	pancreas	Yes	pancreatic	Friess (2003)
4502085	pancreatic amylase alpha 2A precursor	5766/60	78-137	pancreas	Yes	gastric	Kang (2010)
10280622	amylase, pancreatic, alpha-2B precursor	5332/55	78-83, 85-100, 102, 103, 105-108, 110-130, 133, 136-141	pancreas	Yes	liver	Koyama (2001)
40254482	salivary amylase alpha 1A precursor	4712/50	78-83, 86-90, 93-99, 102, 105-107, 110-116, 118-126, 128, 129, 131, 133-135, 139, 140, 142, 143	pancreas/salivary gland	Yes	lung	Tomita (1989)
236460050	elastase 3A, pancreatic preproprotein	4267/49	144-193	pancreas	Yes	lung	Shimada (2002)

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
6679625	elastase 3B, pancreatic preproprotein	4123/44	145-147, 151-153, 155, 157, 161-163, 165, 166, 168, 170, 172-176, 178, 179, 181, 183, 185, 186, 188, 191, 194-211	pancreas	Yes	pancreatic	Gao (2010)
4506147	protease, serine, 2 preproprotein	3427/33	212-247	pancreas	Yes	pancreatic	Gao (2010)
118498350	chymotrypsin B2	2621/27	248-274	pancreas	Yes	pancreatic	Miao (2008)
118498341	chymotrypsin B1	2527/26	248-256, 258-274	pancreas	Yes	general	Miao (2008)
50363217	serine proteinase inhibitor, clade A, member 1	2443/27	275-304	pancreas/liver	Yes	general	Normandin (2010), Sato (2004), Zhou (2000)
15559207	elastase 2A preproprotein	2351/25	305-331	pancreas	Yes	pancreatic	Akakura (2001) Yamamura (1989)

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
54607080	pancreatic carboxypeptidase B1 preproprotein	2143/24	332-355	pancreas	Yes	liver	Matsugi (2007)
4506145	protease, serine, 1 preproprotein	2135/21	215, 217, 219, 221, 222, 225, 228, 229, 231, 232, 236, 238-240, 242, 245, 356-361	pancreas	Yes	pancreatic	
148539844	deleted in malignant brain tumors 1 isoform c precursor	2129/23	362-385	epithelial, pancreas	Yes	pancreatic, brain, lung, colon, gastric	Sasaki (2002), Kuramitsu (2006)
113584	RecName: Full=Ig alpha-1 chain C region	2065/19	386-405	antibody - heavy chain IgA	Yes		
4502997	carboxypeptidase A1 precursor	2026/25	334, 406-429	pancreas	Yes	pancreatic	Matsugi (2007)

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
29725633	regenerating islet-derived 1 alpha precursor	1799/17	430-444	pancreas	Yes	liver	Cavard (2006)
31377806	polymeric immunoglobulin receptor precursor	1783/19	445-465	epithelial	Yes	endometria	DeSouza (2005)
218512088	RecName: Full=Ig alpha-2 chain C region	1779/16	386-388, 390-392, 397, 399-404, 466-469	antibody-heavy chain	Yes		
157266300	membrane alanine aminopeptidase precursor	1581/16	470-486	small intestine	Yes	breast	Liang (2006)
119395750	keratin 1	1491/15	487-501	epithelial			
62526043	chymotrypsin C preproprotein	1407/16	502-517	pancreas	Yes	liver	Wang (2011)
98986445	carcinoembryonic antigen-related cell adhesion molecule 5 preproprotein	1140/11	518-528	epithelial	Yes	pancreatic, colon	Sato (2004), Van Gisbergen (2005)

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
217416390	carboxypeptidase A2 (pancreatic) precursor	1129/11	529-539	pancreas	Yes	pancreatic	Matsugi (2007)
4505847	phospholipase A2 group IB	1124/13	540-552	pancreas	Yes	colon, prostate	Belinsky (2007), Sved (2004)
125145	RecName: Full=Ig kappa chain C region	1121/14	553-566	antibody – light chain	Yes		
7669492	glyceraldehyde-3-phosphate dehydrogenase	1106/11	567-578	epithelial/bacterial		colon	Egea (2007), Shin (2009)
58331211	elastase 2B preproprotein	1083/11	305, 307-310, 312-314, 318, 324, 328, 331	pancreas			
105990514	filamin B, beta (actin binding protein 278)	1041/7	579-586	multiple cell types		prostate	Harding (2006)
4507725	transthyretin precursor	952/11	587-597	Liver/ serum protein	Yes	colon	Fentz (2007)

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
55956899	keratin 9	945/9	598-607	epithelial			
170296790	mesotrypsin isoform 1 preproprotein	907/8	214, 215, 218, 219, 221, 224, 226, 233, 608	pancreas		breast	Hockla (2010)
10835248	regenerating islet-derived 1 beta precursor	903/9	430, 431, 433-437, 441	pancreas	Yes	pancreatic	Cui (2010)
41152086	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 6	902/9	609-617	keratinocytes, muscle, lung, liver, pancreas	Yes	colon	Krasnov (2009)
1684927	immunoglobulin light chain	889/8	618-625				
4505605	pancreatitis-associated protein precursor	848/9	626-634	pancreas	Yes	pancreatic	Rosty (2002)

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
13489087	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 1	842/8	635-642	keratinocytes, muscle, lung, liver, pancreas	Yes	pancreatic	Sato (2004)
226529917	triosephosphate isomerase 1 isoform 2	760/7	643-649	multiple cell types	Yes	breast	Tamesa (2009)
298351713	RecName: Full=Ig lambda-1 chain C regions	760/8	618-623, 625, 650	antibody – light chain			
47132620	keratin 2	713/7	487, 495, 651-655	epithelial			
5080756	Human Fc gamma BP [AA 1-2843]	706/4	656-659				
195972866	keratin 10	676/5	660-664	epithelial		Liver, pancreatic	Yang (2008), Xiao (2010)
4502027	albumin preproprotein	642/4	665-668				

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
40255013	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	614/5	522, 524, 669-671	epithelial	Yes	colon	Van Gisbergen (2005)
125819	RecName: Full=Ig kappa chain V-III region HIC; Flags: Precursor	611/6	672-677	antibody – light chain		leukemia	Kipps (1988)
154146262	Fc fragment of IgG binding protein	607/4	656-659			prostate	Gazi (2008)
157266292	intestinal alkaline phosphatase precursor	605/7	678-684	small intestine		liver	Yamamoto (1984)
50845388	annexin A2 isoform 1	604/5	685-689	multiple cell types		liver	Mohammad (2008)
50659080	serpin peptidase inhibitor, clade A, member 3 precursor	599/4	690-693	liver		melanoma	Wang (2010)

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
51593090	mucin 13, epithelial transmembrane	575/5	694-698	colon		GI cancer	Maier (2011)
119220569	zymogen granule membrane glycoprotein 2 isoform 1	567/4	699-702	pancreas	Yes		
125817	RecName: Full=Ig kappa chain V-III region HAH; Flags: Precursor	554/6	672-677	antibody – light chain		leukemia	Kipps (1988)
151301154	mucin 6, gastric	544/3	703-705				
10334859	creatine kinase, mitochondrial 1B precursor	533/5	706-711	mitochondria		tissue damage	Bark (1980)
223099	Ig Aalpha1 Bur	520/4	388, 391, 467-469				
4504919	keratin 8	506/6	712-717	epithelial		skin	Yamashiro (2010)

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
119703753	keratin 6B	504/4	488, 653, 654, 718	epithelial		breast	Millar (2009)
121039	RecName: Full=Ig gamma-1 chain C region	501/4	719-722	antibody - heavy chain			
4507149	superoxide dismutase 1, soluble	498/5	723-727	epithelial/ mitochondria	Yes	multiple	Pham (2009)
223942069	enterokinase precursor	497/3	728-731	small intestine		multiple	Vilen (2008)
153070262	meprin A alpha	497/4	732-735	small intestine		colon	Lottaz (1999)
157364974	sucrase-isomaltase	480/3	736-738	small intestine		colon	Gu (2006)
125797	RecName: Full=Ig kappa chain V-III region SIE	478/5	672-676, 739	antibody - light chain			
38455402	lipocalin 2	463/4	740-743	Epithelial, many cell types	Yes	pancreatic, breast, endometria l, prostate	Sato (2004), Lin (2011), Mahadevin (2011)

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
167857790	orosomucoid 1 precursor	446/6	744-750	serum/ acute phase protein			
5031839	keratin 6A	441/3	653, 654, 718	epithelial		breast	Millar (2009)
32313593	olfactomedin 4 precursor	437/4	751-754	small intestine , colon , pancreas		pancreatic, colon	Kobayashi (2007), Koshida (2007)
125803	RecName: Full=Ig kappa chain V-III region WOL	432/5	672-676	antibody - light chain			
10835063	nucleophosmin 1 isoform 1	414/5	755-759	multiple cell types		liver, others	Kuramitsu (2006), Grisendi (2006)
75707587	immunoglobulin light chain variable region	414/4	673, 674, 676, 760	antibody - light chain			
5031863	galectin 3 binding protein	414/2	761, 762	multiple cell types	Yes	colon	Bresalier (2004), Kim (2011)

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
187960098	medium-chain acyl-CoA dehydrogenase isoform b precursor	407/3	763-765	mitochondrial/bacterial			
4503143	cathepsin D preproprotein	407/4	766-769	multiple cell types		breast	Wolf (2003)
106507261	pancreatic lipase-related protein 2	393/2	1, 5, 770	pancreas	Yes		
223718246	plastin 1	393/3	771-774	small intestine, colon, kidney		pancreatic	Terris (2002)
4885165	cystatin A	384/4	775-778	macrophages		colon	Kupio (1998), Kos (2000)
6466801	intestinal mucin 3	379/1	779	epithelial		pancreatic	Park (2003)
115430223	galectin 3	377/2	780, 781	multiple cell types		pancreatic	Jiang (2008)

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
19923195	carcinoembryonic antigen-related cell adhesion molecule 1 isoform 1 precursor	375/4	518, 670, 782, 783	epithelial	Yes	multiple	Gerstel (2011)
19923748	dihydrolipoamide S-succinyltransferase (E2 component of 2-oxo-glutarate complex)	362/3	784-786	mitochondria			
193806374	RecName: Full=Ig mu chain C region	361/4	787-790	antibody - heavy chain			
16306550	selenium binding protein 1	361/3	791-793			ovarian, uterine, gastric, esophageal	Huang (2006), Zhang (2010), Zhang (2011), Silvers (2010)
33456	immunoglobulin M chain	353/4	794-797	antibody – heavy chain			

NCBI gi#	Detected protein	Mascot score / number of significant peptides	SEQ ID NO.s of peptides with Mascot scores >40	Origin of detected protein	Presence in pancreatic juice	Associated cancer	References
125811	RecName: Full=Ig kappa chain V-III region VG; Flags: Precursor	319/3	760, 798, 799	antibody – light chain			
123843	RecName: Full=Ig heavy chain V-III region VH26; Flags: Precursor	219/3	796, 800, 801	antibody – heavy chain			
563454	Ig heavy chain (VH4) V region (VDJ)	206/3	802-804	antibody – heavy chain			

[0115] The foregoing analysis demonstrates that a large number of proteins can be detected in samples obtained using a GLF.

Example 8—Proteomic analysis of fecal samples

[0116] In this analysis, the ability of fecal samples to support proteomic analysis was assessed. To identify target molecules in a fecal sample, no lavage composition was administered to a human subject. A fecal sample was collected from the subject during normal defecation using a collection container placed in the toilet. A small amount of the fecal (stool) sample was homogenized in 0.1% TFA and then centrifuged at 13000 x g. The protein was precipitated with 6 volumes of acetone, resuspended in 0.1% TFA, extracted with an equal volume of chloroform, and then processed in a SPE column as described in Example 1. The most abundant identified proteins and their corresponding Mascot scores and number of significant peptides are summarized in Table 11. Proteins, largely with likely pancreatic origin, were detected in the sample indicating that, after discovery in GLF, stool is a source for the detection of these biomarker proteins. However, the sample did also contain a number of other non-human materials that make the analysis much more limiting, especially for discovery of biomarkers.

TABLE 11

NCBI gi #	Protein	Mascot score / number of significant peptides
		Sample Fecal 1
119395750	keratin 1	549 / 7
15559207	elastase 2A preproprotein	546 / 5
236460050	elastase 3A pancreatic preproprotein	532 / 4
55956899	keratin 9	444 / 5
125145	RecName: Full=Ig kappa chain C region	396 / 5
4506147	protease serine 2 preproprotein	344 / 4

4506145	protease serine 1 preproprotein	320 / 5
118498350	chymotrypsin B2	298 / 3
6679625	elastase 3B pancreatic preproprotein	286 / 3

[0117] The foregoing analysis demonstrates that a large number of target molecules can be detected in fecal samples.

Example 9—Detection of glycans in samples obtained using a GLF

[0118] In this analysis, the ability of glycans to be detected in samples obtained in samples using a GLF was assessed. To identify and analyze target molecules, including glycans, in a GLF, GLF was collected from a human subject. 1.8 mL GLF was added to 12 mL ice cold acetone and incubated for one hour to pellet the protein. The sample was centrifuged at 12,000 x g for 15 minutes and the acetone removed. After washing with ice cold acetone, the pellet was resuspended in 0.1% TFA and passed through a 5 mL, syringe style, SepPak C2 column. Proteins were eluted with 60% acetonitrile / 40% 0.1% TFA. After removal of the solvent under vacuum, the protein fraction was redissolved in 100 µL 50 mM ammonium bicarbonate and deglycosylated with 2 µL PNGaseF overnight at 37 °C in a shaker. After quenching with 1 mL 0.1% TFA, the glycans were collected as the flow through fraction from a 1 mL, syringe style, SepPak C18 column using a vacuum manifold. The dried glycans were labeled with 4-ABEE (ethyl 4-aminobenzoate) by reductive amination by adding 25 µL derivatizing solution (90:10 MeOH:HAc containing 35mM ABEE and 100 mM 2-PB) to the dried glycans, and incubating at 65 °C for 2 hours. Excess ABEE was removed by adding 1 mL ethyl ether, vortexing and discarding the ether. After a second ether extraction, the sample was briefly put in the SpeedVac to remove any residual ether. The labeled glycans were then run on the HPLC and eluted between 20-25% acetonitrile from an Agilent C8 reverse phase column. This fraction was vacuum-dried, redissolved in 50 µL of 0.1% TFA and run on the Waters Q-TOF LC-ESI-MS for glycan analysis. The derivatized glycans eluted from the C18 reverse phase column on the Q-TOF MS at about 20-25% acetonitrile in 0.2% formic acid. The mass spectrometer was scanned in MS-only mode

from m/z 150-2000 every second to acquire the derivatized glycan profile data. FIG. 1 summarizes these results, and depicts a graph of the relative abundance of various glycoprotein derived glycan structures present in a fraction of a gastrointestinal lavage fluid. As shown in FIG. 1, some glycoprotein derived glycan structures include particular modifications that are associated with truncation of the chains. These modifications may be due to bacterial activity present in the GLF sample as it is known that bacteria can digest and consume glycans from proteins. However, such modifications can also be associated with disease, especially cancer where aberrant glycosylation has been linked to the disease.

[0119] The foregoing analysis demonstrates that a large number of glycans can be detected in samples obtained using a GLF.

Example 10—Detection of metabolites in samples obtained using a GLF

[0120] In this analysis, the ability of metabolites to be detected in samples obtained using a GLF was assessed. To identify and analyze target molecules, including metabolites, in a GLF, a magnesium citrate-based lavage composition was administered to human subjects, and the resultant GLF was analyzed for metabolites such as cholic acid and other bile salts. The resultant GLF was collected from the subjects as part of a colonoscopy procedure.

[0121] 3 ml GLF was centrifuged at max speed for 20 minutes and the supernatant acidified with 0.1% TFA. The supernatant was applied to a C18 SPE column (Waters Sep-Pak), washed with 3 volumes of 0.1% TFA, and eluted with 50% ACN in 0.1%TFA. The elutant was dried by centrifugal lyophilization and re-dissolved in 500 μ l 0.1% TFA.

[0122] Data were acquired on a Waters Q-TOF mass spectrometer using input from an LC system. The A solvent contained 3% of B and 0.2% formic acid in water. The B solvent contained 3% of A and 0.2% formic acid in acetonitrile. Solvents were HPLC grade from Fisher. The starting solvent was 5% B and remained for 5 min and then changed to 40% by 25 min, 90% by 30 min, and then reset to 5% at 36. The MS scanned over the mass range from m/z 100 m/z to 2000 every second. Data were acquired using the standard MassLynx software. The eluting compounds with the cholic acid peak marked are summarized in FIG. 2. A similar profile of peaks was observed on the

Orbitrap instrument where the cholic acid peak was identified using a standard and MS/MS data. Metabolites including cholic acid were identified.

[0123] The foregoing analysis demonstrates that metabolites can be detected in samples obtained using a GLF.

[0124] The term “comprising” as used herein is synonymous with “including,” “containing,” or “characterized by,” and is inclusive or open-ended and does not exclude additional, unrecited elements or method steps.

[0125] All numbers expressing quantities of ingredients, reaction conditions, and so forth used in the specification are to be understood as being modified in all instances by the term “about.” Accordingly, unless indicated to the contrary, the numerical parameters set forth herein are approximations that may vary depending upon the desired properties sought to be obtained. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of any claims in any application claiming priority to the present application, each numerical parameter should be construed in light of the number of significant digits and ordinary rounding approaches.

[0126] The above description discloses several methods and materials of the present invention. This invention is susceptible to modifications in the methods and materials, as well as alterations in the fabrication methods and equipment. Such modifications will become apparent to those skilled in the art from a consideration of this disclosure or practice of the invention disclosed herein. Consequently, it is not intended that this invention be limited to the specific embodiments disclosed herein, but that it cover all modifications and alternatives coming within the true scope and spirit of the invention.

References

Each of the following references is incorporated by reference herein in its entirety.

Pancreatic Juice References

- Chen, R., S. Pan, et al. (2007). "Comparison of pancreas juice proteins from cancer versus pancreatitis using quantitative proteomic analysis." *Pancreas* 34(1): 70-9.
- Chen, R., S. Pan, et al. (2010). "Elevated level of anterior gradient-2 in pancreatic juice from patients with pre-malignant pancreatic neoplasia." *Mol Cancer* 9: 149.
- Chen, R., S. Pan, et al. (2006). "Quantitative proteomic profiling of pancreatic cancer juice." *Proteomics* 6(13): 3871-9.

- Gao, J., F. Zhu, et al. (2010). "Identification of pancreatic juice proteins as biomarkers of pancreatic cancer." *Oncol Rep* 23(6): 1683-92.
- Gomez-Lazaro, M., C. Rinn, et al. (2010). "Proteomic analysis of zymogen granules." *Expert Rev Proteomics* 7(5): 735-47.
- Gronborg, M., J. Bunkenborg, et al. (2004). "Comprehensive proteomic analysis of human pancreatic juice." *J Proteome Res* 3(5): 1042-55.
- Paulo, J. A., L. S. Lee, et al. (2010). "Identification of pancreas-specific proteins in endoscopically (endoscopic pancreatic function test) collected pancreatic fluid with liquid chromatography--tandem mass spectrometry." *Pancreas* 39(6): 889-96.
- Zhou, L., Z. Lu, et al. (2007). "Comparative proteomic analysis of human pancreatic juice: methodological study." *Proteomics* 7(8): 1345-55.

Cancer References

- Akakura, N., M. Kobayashi, et al. (2001). "Constitutive expression of hypoxia-inducible factor-1alpha renders pancreatic cancer cells resistant to apoptosis induced by hypoxia and nutrient deprivation." *Cancer Res* 61(17): 6548-54.
- Bark, C. J. (1980). "Mitochondrial Creatine Kinase." *Journal of the American Medical Association* 243(20): 2058-2060.
- Belinsky, G. S., T. V. Rajan, et al. (2007). "Expression of secretory phospholipase A2 in colon tumor cells potentiates tumor growth." *Mol Carcinog* 46(2): 106-16.
- Bresalier, R. S., J. C. Byrd, et al. (2004). "A circulating ligand for galectin-3 is a haptoglobin-related glycoprotein elevated in individuals with colon cancer." *Gastroenterology* 127(3): 741-8.
- Cavard, C., B. Terris, et al. (2006). "Overexpression of regenerating islet-derived 1 alpha and 3 alpha genes in human primary liver tumors with beta-catenin mutations." *Oncogene* 25(4): 599-608.
- Cui, L., F. Li, et al. (2010). "Screening and Verification of Differentially Expressed Proteins from Pancreatic Cancer Tissue." *Chinese Journal of Chemistry* 28(6): 884-890.
- DeSouza, L., G. Diehl, et al. (2005). "Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry." *J Proteome Res* 4(2): 377-86.
- Egea, L., L. Aguilera, et al. (2007). "Role of secreted glyceraldehyde-3-phosphate dehydrogenase in the infection mechanism of enterohemorrhagic and enteropathogenic *Escherichia coli*: interaction of the extracellular enzyme with human plasminogen and fibrinogen." *Int J Biochem Cell Biol* 39(6): 1190-203.
- Fentz, A. K., M. Sporn, et al. (2007). "Detection of colorectal adenoma and cancer based on transthyretin and C3a-desArg serum levels." *Proteomics Clin Appl* 1(6): 536-44.
- Friess, H., J. Ding, et al. (2003). "Microarray-based identification of differentially expressed growth- and metastasis-associated genes in pancreatic cancer." *Cell Mol Life Sci* 60(6): 1180-99.
- Gao, J., F. Zhu, et al. (2010). "Identification of pancreatic juice proteins as biomarkers of pancreatic cancer." *Oncol Rep* 23(6): 1683-92.
- Gazi, M. H., M. He, et al. (2008). "Downregulation of IgG Fc binding protein (Fc gammaBP) in prostate cancer." *Cancer Biol Ther* 7(1): 70-5.
- Gerstel, D., F. Wegwitz, et al. (2011). "CEACAM1 creates a pro-angiogenic tumor microenvironment that supports tumor vessel maturation." *Oncogene*.
- Grisendi, S., C. Mecucci, et al. (2006). "Nucleophosmin and cancer." *Nat Rev Cancer* 6(7): 493-505.

- Gu, N., T. Adachi, et al. (2006). "Sucrase-isomaltase gene expression is inhibited by mutant hepatocyte nuclear factor (HNF)-1alpha and mutant HNF-1beta in Caco-2 cells." *J Nutr Sci Vitaminol (Tokyo)* 52(2): 105-12.
- Harding, T. C., M. Nguyen, et al. (2006). Humoral immune response induced to filamin B in patients with metastatic hormone-refractory prostate cancer (HRPC) treated with a GM-CSF-transduced allogeneic prostate cancer vaccine (GVAX(R)). AACR.
- Hockla, A., D. C. Radisky, et al. (2010). "Mesotrypsin promotes malignant growth of breast cancer cells through shedding of CD109." *Breast Cancer Res Treat* 124(1): 27-38.
- Huang, K. C., D. C. Park, et al. (2006). "Selenium binding protein 1 in ovarian cancer." *Int J Cancer* 118(10): 2433-40.
- Jiang, H. B., M. Xu, et al. (2008). "Pancreatic stellate cells promote proliferation and invasiveness of human pancreatic cancer cells via galectin-3." *World J Gastroenterol* 14(13): 2023-8.
- Kang, J. U., S. H. Koo, et al. (2010). "AMY2A: a possible tumor-suppressor gene of 1p21.1 loss in gastric carcinoma." *Int J Oncol* 36(6): 1429-35.
- Kim, G. E., H. I. Bae, et al. (2002). "Aberrant expression of MUC5AC and MUC6 gastric mucins and sialyl Tn antigen in intraepithelial neoplasms of the pancreas." *Gastroenterology* 123(4): 1052-60.
- Kim, Y. S., J. A. Jung, et al. (2011). "Galectin-3 binding protein promotes cell motility in colon cancer by stimulating the shedding of protein tyrosine phosphatase kappa by proprotein convertase 5." *Biochem Biophys Res Commun* 404(1): 96-102.
- Kipps, T. J., E. Tomhave, et al. (1988). "Autoantibody-associated kappa light chain variable region gene expressed in chronic lymphocytic leukemia with little or no somatic mutation. Implications for etiology and immunotherapy." *J Exp Med* 167(3): 840-52.
- Kobayashi, D., S. Koshida, et al. (2007). "Olfactomedin 4 promotes S-phase transition in proliferation of pancreatic cancer cells." *Cancer Sci* 98(3): 334-40.
- Kos, J., M. Krasovec, et al. (2000). "Cysteine proteinase inhibitors stefin A, stefin B, and cystatin C in sera from patients with colorectal cancer: relation to prognosis." *Clin Cancer Res* 6(2): 505-11.
- Koshida, S., D. Kobayashi, et al. (2007). "Specific overexpression of OLFM4(GW112/HGC-1) mRNA in colon, breast and lung cancer tissues detected using quantitative analysis." *Cancer Sci* 98(3): 315-20.
- Koyama, I., S. Komine, et al. (2001). "alpha-Amylase expressed in human liver is encoded by the AMY-2B gene identified in tumorous tissues." *Clin Chim Acta* 309(1): 73-83.
- Krasnov, G. S., N. Oparina, et al. (2009). "[Colorectal cancer 2D-proteomics: identification of altered protein expression]." *Mol Biol (Mosk)* 43(2): 348-56.
- Kuopio, T., A. Kankaanranta, et al. (1998). "Cysteine proteinase inhibitor cystatin A in breast cancer." *Cancer Res* 58(3): 432-6.
- Kuramitsu, Y. and K. Nakamura (2006). "Proteomic analysis of cancer tissues: shedding light on carcinogenesis and possible biomarkers." *Proteomics* 6(20): 5650-61.
- Liang, X., J. Zhao, et al. (2006). "Quantification of membrane and membrane-bound proteins in normal and malignant breast cancer cells isolated from the same patient with primary breast carcinoma." *J Proteome Res* 5(10): 2632-41.
- Lin, H. H., C. J. Liao, et al. (2011). "Lipocalin-2-induced cytokine production enhances endometrial carcinoma cell survival and migration." *Int J Biol Sci* 7(1): 74-86.
- Lottaz, D., C. A. Maurer, et al. (1999). "Nonpolarized secretion of human meprin alpha in colorectal cancer generates an increased proteolytic potential in the stroma." *Cancer Res* 59(5): 1127-33.

- Mahadevan, N. R., J. Rodvold, et al. (2011). "ER stress drives Lipocalin 2 upregulation in prostate cancer cells in an NF-kappaB-dependent manner." *BMC Cancer* 11: 229.
- Maher, D. M., B. K. Gupta, et al. (2011). "Mucin 13: structure, function, and potential roles in cancer pathogenesis." *Mol Cancer Res* 9(5): 531-7.
- Matsugi, S., T. Hamada, et al. (2007). "Serum carboxypeptidase A activity as a biomarker for early-stage pancreatic carcinoma." *Clin Chim Acta* 378(1-2): 147-53.
- Miao, Q., Y. Sun, et al. (2008). "Chymotrypsin B cached in rat liver lysosomes and involved in apoptotic regulation through a mitochondrial pathway." *J Biol Chem* 283(13): 8218-28.
- Millar, E. K., P. H. Graham, et al. (2009). "Prediction of local recurrence, distant metastases, and death after breast-conserving therapy in early-stage invasive breast cancer using a five-biomarker panel." *J Clin Oncol* 27(28): 4701-8.
- Mohammad, H. S., K. Kurokohchi, et al. (2008). "Annexin A2 expression and phosphorylation are up-regulated in hepatocellular carcinoma." *Int J Oncol* 33(6): 1157-63.
- Normandin, K., B. Peant, et al. (2010). "Protease inhibitor SERPINA1 expression in epithelial ovarian cancer." *Clin Exp Metastasis* 27(1): 55-69.
- Park, H. U., J. W. Kim, et al. (2003). "Aberrant expression of MUC3 and MUC4 membrane-associated mucins and sialyl Le(x) antigen in pancreatic intraepithelial neoplasia." *Pancreas* 26(3): e48-54.
- Pham, T. M., Y. Fujino, et al. (2009). "Relationship between serum levels of superoxide dismutase activity and subsequent risk of cancer mortality: Findings from a nested case-control study within the Japan Collaborative Cohort Study." *Asian Pac J Cancer Prev* 10 Suppl: 69-73.
- Rosty, C., L. Christa, et al. (2002). "Identification of hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein I as a biomarker for pancreatic ductal adenocarcinoma by protein biochip technology." *Cancer Res* 62(6): 1868-75.
- Sasaki, K., K. Sato, et al. (2002). "Peptidomics-based approach reveals the secretion of the 29-residue COOH-terminal fragment of the putative tumor suppressor protein DMBT1 from pancreatic adenocarcinoma cell lines." *Cancer Res* 62(17): 4894-8.
- Sato, N., N. Fukushima, et al. (2004). "Gene expression profiling identifies genes associated with invasive intraductal papillary mucinous neoplasms of the pancreas." *Am J Pathol* 164(3): 903-14.
- Shimada, S., K. Yamaguchi, et al. (2002). "Pancreatic elastase IIIA and its variants are expressed in pancreatic carcinoma cells." *Int J Mol Med* 10(5): 599-603.
- Shin, Y. K., B. C. Yoo, et al. (2009). "Upregulation of glycolytic enzymes in proteins secreted from human colon cancer cells with 5-fluorouracil resistance." *Electrophoresis* 30(12): 2182-92.
- Silvers, A. L., L. Lin, et al. (2010). "Decreased selenium-binding protein 1 in esophageal adenocarcinoma results from posttranscriptional and epigenetic regulation and affects chemosensitivity." *Clin Cancer Res* 16(7): 2009-21.
- Sved, P., K. F. Scott, et al. (2004). "Oncogenic action of secreted phospholipase A2 in prostate cancer." *Cancer Res* 64(19): 6934-40.
- Tamesa, M. S., Y. Kuramitsu, et al. (2009). "Detection of autoantibodies against cyclophilin A and triosephosphate isomerase in sera from breast cancer patients by proteomic analysis." *Electrophoresis* 30(12): 2168-81.
- Terris, B., E. Blaveri, et al. (2002). "Characterization of gene expression profiles in intraductal papillary-mucinous tumors of the pancreas." *Am J Pathol* 160(5): 1745-54.

- Tomita, N., A. Horii, et al. (1989). "A novel type of human alpha-amylase produced in lung carcinoid tumor." *Gene* 76(1): 11-8.
- van Gisbergen, K. P., C. A. Aarnoudse, et al. (2005). "Dendritic cells recognize tumor-specific glycosylation of carcinoembryonic antigen on colorectal cancer cells through dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin." *Cancer Res* 65(13): 5935-44.
- Vilen, S. T., P. Nyberg, et al. (2008). "Intracellular co-localization of trypsin-2 and matrix metalloprotease-9: possible proteolytic cascade of trypsin-2, MMP-9 and enterokinase in carcinoma." *Exp Cell Res* 314(4): 914-26.
- Wang, H., W. Sha, et al. (2011). "Effect of chymotrypsin C and related proteins on pancreatic cancer cell migration." *Acta Biochim Biophys Sin (Shanghai)* 43(5): 362-71.
- Wang, Y., H. Jiang, et al. (2010). "Alpha 1 antichymotrypsin is aberrantly expressed during melanoma progression and predicts poor survival for patients with metastatic melanoma." *Pigment Cell Melanoma Res* 23(4): 575-8.
- Wolf, M., I. Clark-Lewis, et al. (2003). "Cathepsin D specifically cleaves the chemokines macrophage inflammatory protein-1 alpha, macrophage inflammatory protein-1 beta, and SLC that are expressed in human breast cancer." *Am J Pathol* 162(4): 1183-90.
- Xiao, J., W. N. Lee, et al. (2010). "Profiling pancreatic cancer-secreted proteome using ¹⁵N amino acids and serum-free media." *Pancreas* 39(1): e17-23.
- Yamamoto, H., M. Tanaka, et al. (1984). "Intestinal-type alkaline phosphatase produced by human hepatoblastoma cell line HUH-6 clone 5." *Cancer Res* 44(1): 339-44.
- Yamamura, H., M. Tatsuta, et al. (1989). "Effectiveness of discriminant analysis of serum CA 19-9 and elastase 1 in diagnosis of pancreatic carcinoma." *Pancreas* 4(4): 401-5.
- Yamashiro, Y., K. Takei, et al. (2010). "Ectopic coexpression of keratin 8 and 18 promotes invasion of transformed keratinocytes and is induced in patients with cutaneous squamous cell carcinoma." *Biochem Biophys Res Commun* 399(3): 365-72.
- Yang, X. R., Y. Xu, et al. (2008). "Cytokeratin 10 and cytokeratin 19: predictive markers for poor prognosis in hepatocellular carcinoma patients after curative resection." *Clin Cancer Res* 14(12): 3850-9.
- Zhang, J., W. G. Dong, et al. (2011). "Reduced selenium-binding protein 1 is associated with poor survival rate in gastric carcinoma." *Med Oncol* 28(2): 481-7.
- Zhang, P., C. Zhang, et al. (2010). "The expression of selenium-binding protein 1 is decreased in uterine leiomyoma." *Diagn Pathol* 5: 80.
- Zheng, H., H. Takahashi, et al. (2006). "MUC6 down-regulation correlates with gastric carcinoma progression and a poor prognosis: an immunohistochemical study with tissue microarrays." *J Cancer Res Clin Oncol* 132(12): 817-23.
- Zhou, H., M. E. Ortiz-Pallardo, et al. (2000). "Is heterozygous alpha-1-antitrypsin deficiency type PIZ a risk factor for primary liver carcinoma?" *Cancer* 88(12): 2668-76.

[0127] All references cited herein, including but not limited to published and unpublished applications, patents, and literature references, are incorporated herein by reference in their entirety and are hereby made a part of this specification. To the extent publications and patents or patent applications incorporated by reference contradict the disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

WHAT IS CLAIMED IS:

1. A method for assessing the physiological state of a subject comprising:
obtaining a gastrointestinal lavage fluid from the subject; and
detecting a target molecule which originated from outside the gastrointestinal system in the gastrointestinal lavage fluid.
2. A method for assessing the physiological state of a subject comprising:
obtaining a fecal sample from the subject; and
detecting a target molecule which originated from outside the gastrointestinal system in the fecal sample.
3. The method of claim 1, wherein the gastrointestinal lavage fluid is obtained from the subject by partially purging the subject's gastrointestinal system.
4. The method of claim 1, wherein the gastrointestinal lavage fluid comprises fecal matter.
5. The method of claim 2, wherein the fecal sample comprises a gastrointestinal lavage fluid.
6. The method of any one of claims 1-5, wherein the target molecule comprises a polypeptide, antibody, bile acid, metabolite, or glycan.
7. The method of any one of claims 1-6, wherein the target molecule comprises a biomarker.
8. The method of claim 7, wherein the biomarker is associated with a disease, a positive response to treatment, a partial response to treatment, a negative response to treatment, or no response to treatment.
9. The method of any one of claims 1-8, wherein the target molecule is associated with presence of a cancer or predisposition to a cancer.
10. The method of claim 8, wherein the cancer is pancreatic cancer, colorectal cancer, liver cancer, or gastric cancer.
11. The method of any one of claims 1-10, wherein the target molecule originated from an accessory digestive gland.
12. The method of claim 11, wherein the accessory digestive gland is salivary glands, pancreas, gallbladder, or liver.
13. The method of any one of claims 1-12, further comprising administering a lavage fluid to the subject.

14. The method of claim 13, wherein the lavage fluid is administered orally.
15. The method of claim 13, wherein the lavage fluid comprises an ingredient selected from the group consisting of polyethylene glycol, magnesium sulfate, sodium sulfate, potassium sulfate, magnesium citrate, ascorbic acid, sodium picosulfate, and bisacodyl.
16. The method of claim 13, wherein the lavage fluid is selected from the group consisting of GOLYTELY, HALFLYTELY, NULYTELY, SUPREP, FLEET'S PHOSPHO-SODA, magnesium citrate, and their generic equivalents.
17. The method of any one of claims 1-16, further comprising performing a colonoscopy on the subject.
18. The method of any one of claims 1-17, wherein the subject is mammalian.
19. The method of any one of claims 1-18, wherein the subject is human.
20. A method for identifying a biomarker comprising:
 - obtaining a test gastrointestinal lavage fluid from a plurality of test subjects having a condition or physiological state of interest and a control gastrointestinal lavage fluid from a plurality of control subjects who do not have said condition or physiological state;
 - determining the level of at least 5 target molecules in the test gastrointestinal lavage fluid and the control gastrointestinal lavage fluid, and
 - identifying a target molecule which is present at significantly different levels in the test gastrointestinal lavage fluid relative to the levels in the control gastrointestinal lavage fluid, thereby identifying a biomarker.
21. The method of claim 20, wherein the gastrointestinal lavage fluid comprises fecal matter.
22. The method of any one of claims 20-21, wherein the target molecules are selected from the group consisting of polypeptides, bile acids, antibodies, metabolites, glycans, and a combination thereof.
23. The method of any one of claims 20-22, comprising determining the level of at least 10 target molecules in the test gastrointestinal lavage fluid and the control gastrointestinal lavage fluid.
24. The method of any one of claims 20-22, comprising determining the level of at least 20 target molecules in the test gastrointestinal lavage fluid and the control gastrointestinal lavage fluid.

25. The method of any one of claims 20-22, comprising determining the level of at least 30 target molecules in the test gastrointestinal lavage fluid and the control gastrointestinal lavage fluid.

26. The method of any one of claims 20-22, comprising determining the level of at least 50 target molecules in the test gastrointestinal lavage fluid and the control gastrointestinal lavage fluid.

27. The method of any one of claims 20-22, comprising determining the level of at least 100 target molecules in the test gastrointestinal lavage fluid and the control gastrointestinal lavage fluid.

28. The method of any one of claims 20-27, wherein the biomarker is associated with a disease, a positive response to treatment, a partial response to treatment, a negative response to treatment or no response to treatment.

29. The method of any one of claims 20-27, wherein the biomarker is associated with the presence of a cancer or predisposition to a cancer.

30. The method of claim 28, wherein the cancer is pancreatic cancer, liver cancer, or gastric cancer.

31. The method of any one of claims 20-30, wherein at least one target molecule originated from an accessory digestive gland.

32. The method of claim 31, wherein the accessory digestive gland is salivary glands, pancreas, gallbladder, or liver.

33. The method of any one of claims 20-32, further comprising administering a lavage fluid to the test subjects and the control subjects.

34. The method of claim 33, wherein the lavage fluid is administered orally.

35. The method of claim 33, wherein the lavage fluid comprises an ingredient selected from the group consisting of polyethylene glycol, magnesium sulfate, sodium sulfate, potassium sulfate, magnesium citrate, ascorbic acid, sodium picosulfate, and bisacodyl.

36. The method of claim 33, wherein the lavage fluid is selected from the group consisting of GOLYTELY, HALFLYTELY, NULYTELY, SUPREP, FLEET'S PHOSPHO-SODA, magnesium citrate, and their generic equivalents.

37. The method of any one of claims 20-36, further comprising performing a colonoscopy on the test subjects and control subjects.

38. The method of any one of claims 20-37, wherein the test subjects and control subjects are mammalian.

39. The method of any one of claims 20-38, wherein the test subjects and control subjects are human.

40. A method for identifying a biomarker comprising:

obtaining a test fecal sample from a plurality of test subjects having a condition of interest and a control fecal sample from a plurality of control subjects and;

determining the level of at least 5 target molecules in the test fecal sample and the control fecal sample, identifying a target molecule which is present at significantly different levels in the test fecal sample relative to the levels in the control fecal sample, thereby identifying a biomarker.

41. The method of claim 40, wherein the fecal sample comprises a gastrointestinal lavage fluid.

42. The method of any one of claims 40-41, wherein the target molecules are selected from the group consisting of polypeptides, nucleic acids, bile acids, antibodies, metabolites, glycans, and a combination thereof.

43. The method of any one of claims 40-42, comprising determining the level of at least 10 target molecules in the test fecal sample and the control fecal sample.

44. The method of any one of claims 40-42, comprising determining the level of at least 20 target molecules in the fecal sample and the control fecal sample.

45. The method of any one of claims 40-42, comprising determining the level of at least 30 target molecules in the fecal sample and the control fecal sample.

46. The method of any one of claims 40-42, comprising determining the level of at least 50 target molecules in the fecal sample and the control fecal sample.

47. The method of any one of claims 40-42, comprising determining the level of at least 100 target molecules in in the fecal sample and the control fecal sample.

48. The method of any one of claims 40-47, wherein the biomarker is associated with a disease, a positive response to treatment, or a negative response to treatment.

49. The method of any one of claims 40-48, wherein the biomarker is associated with the presence of a cancer or predisposition to a cancer.

50. The method of claim 49, wherein the cancer is pancreatic cancer, colorectal cancer, liver cancer, or gastric cancer.

51. The method of any one of claims 40-50, wherein at least one target molecule originated from an accessory digestive gland.

52. The method of claim 51, wherein the accessory digestive gland is salivary glands, pancreas, gallbladder, or liver.

53. The method of any one of claims 40-52, wherein the test subjects and control subjects are mammalian.

54. The method of any one of claims 40-53, wherein the test subjects and control subjects are human.

55. A kit for detecting a target molecule in a gastrointestinal lavage fluid comprising:

a lavage fluid for oral administration to a subject;

a vessel for collecting the gastrointestinal lavage fluid from the subject;

and

an agent for detecting a target molecule which originated from outside the gastrointestinal system.

56. A kit for detecting a target molecule in a fecal sample comprising:

a lavage fluid for oral administration to a subject;

a vessel for collecting the fecal sample from the subject; and

an agent for detecting a target molecule which originated from outside the gastrointestinal system.

57. The kit of claim any one of claims 55-56, further comprising a protease inhibitor.

58. The kit of any one of claims 55-57, wherein the target molecule comprises a polypeptide, antibody, bile acid, metabolite, or glycan.

59. The kit of any one of claims 55-58, wherein the target molecule comprises a biomarker.

60. The kit of claim 59, wherein the biomarker is associated with a disease, a positive response to treatment, or a negative response to treatment.

61. The kit of any one of claims 55-60, wherein the target molecule is associated with presence of a cancer or predisposition to a cancer.

62. The kit of claim 61, wherein the cancer is pancreatic cancer, liver cancer, colorectal cancer, or gastric cancer.

63. The kit of any one of claims 55-62, wherein the target molecule originated from an accessory digestive gland.

64. The kit of claim 63, wherein the accessory digestive gland is salivary glands, pancreas, gallbladder, or liver.

65. The kit of any one of claims 55-64, wherein the lavage fluid comprises an ingredient selected from the group consisting of polyethylene glycol, magnesium sulfate, sodium sulfate, potassium sulfate, magnesium citrate, ascorbic acid, sodium picosulfate, and bisacodyl.

66. The kit of any one of claims 55-64, wherein the lavage fluid is selected from the group consisting of GOLYTELY, HALFLYTELY, NULYTELY, SUPREP, FLEET'S PHOSPHO-SODA, magnesium citrate, and their generic equivalents.

colonic lavage glycans labeled with 4abee, HPLC eluted at 13.5min 0.1% TFA 6.4ul inj
jl_06211_glycans_1 516 (23.605) M3 [Ev-733016, It50, En1] (0.050,200.00,0.200,1400.00,4,Cmp); Cm(454:537)

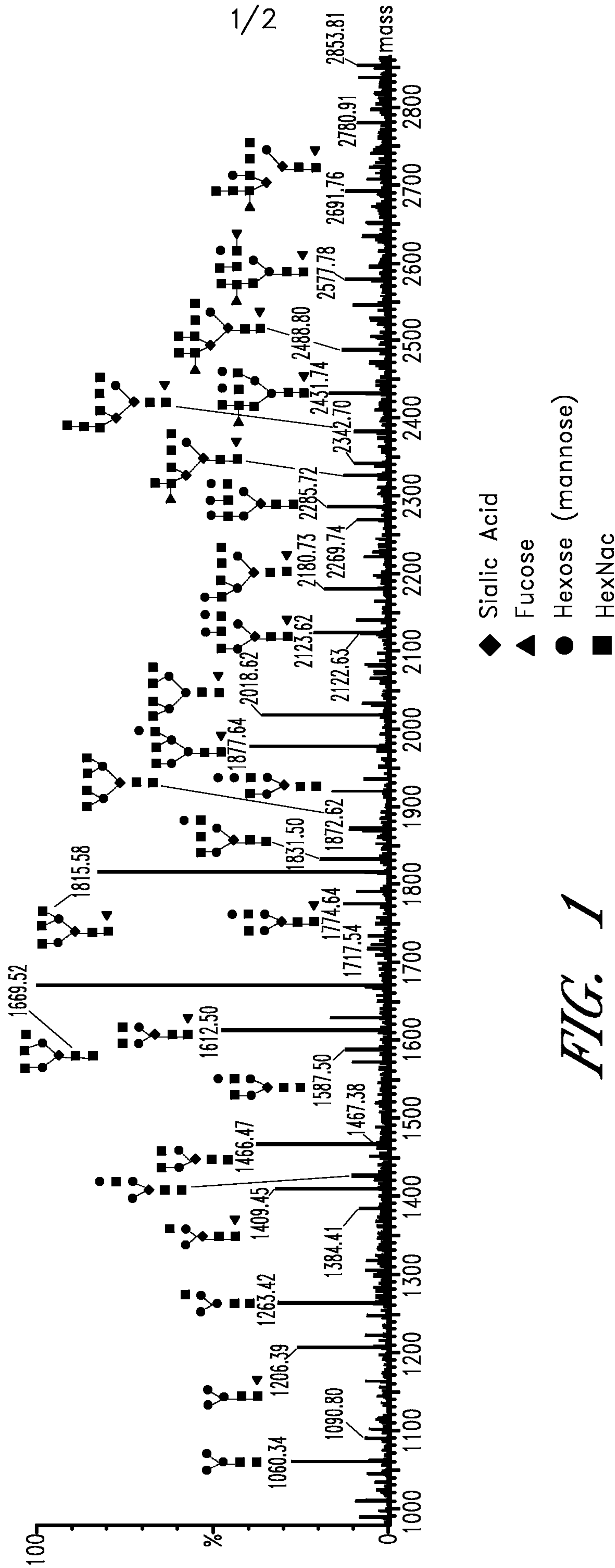
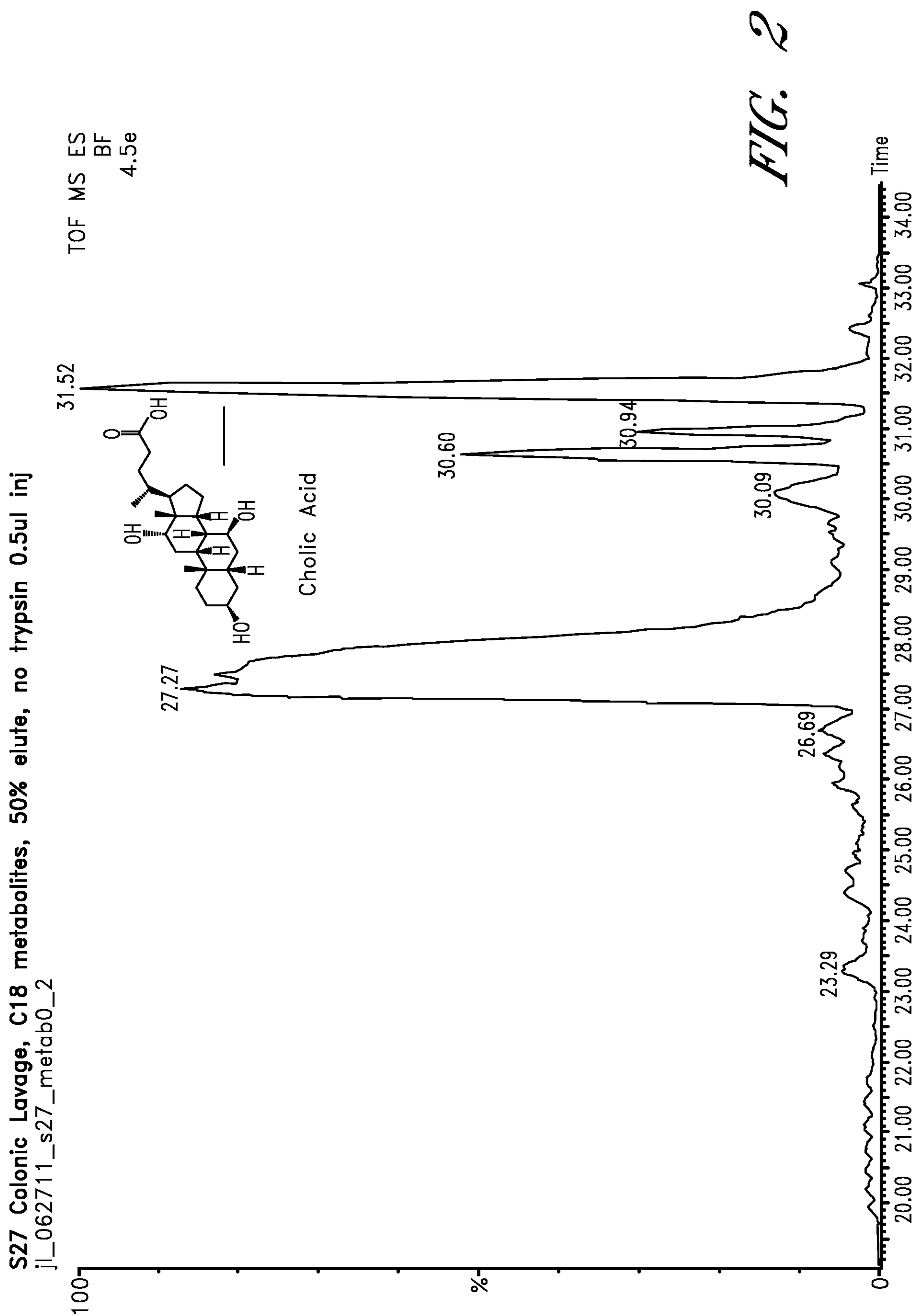


FIG. 1

2/2



colonic lavage glycans labeled with 4abee, HPLC eluted at 13.5min 0.1% TFA 6.4ul inj
 jl_06211_glycans_1 516 (23.605) M3 [Ev-733016, It50, En1] (0.050,200.00,0.200,1400.00,4,Cmp); Cm(454:537)

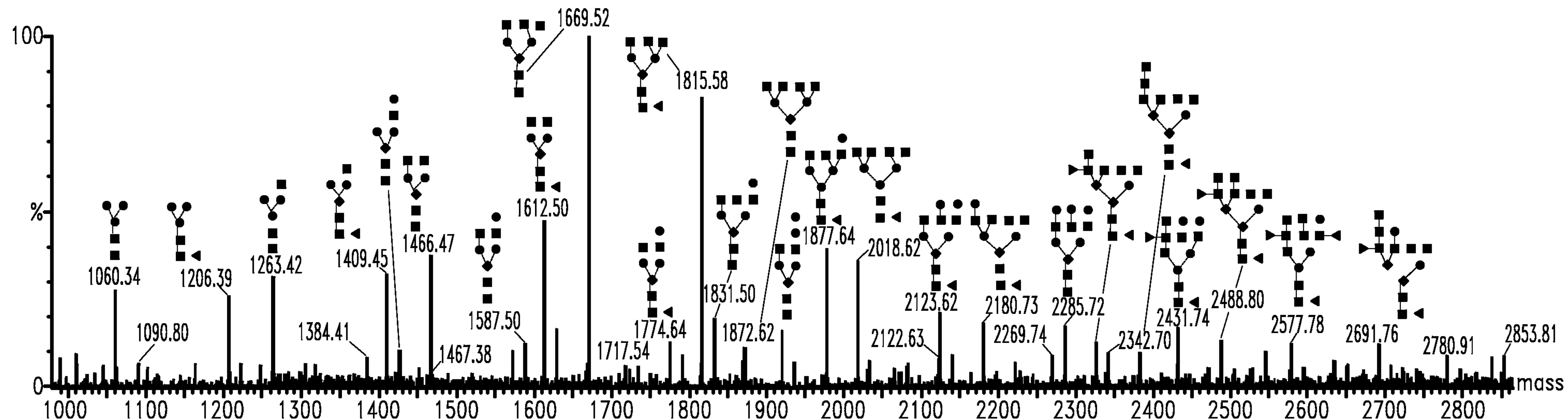


FIG. 1

- ◆ Sialic Acid
- ▲ Fucose
- Hexose (mannose)
- HexNac