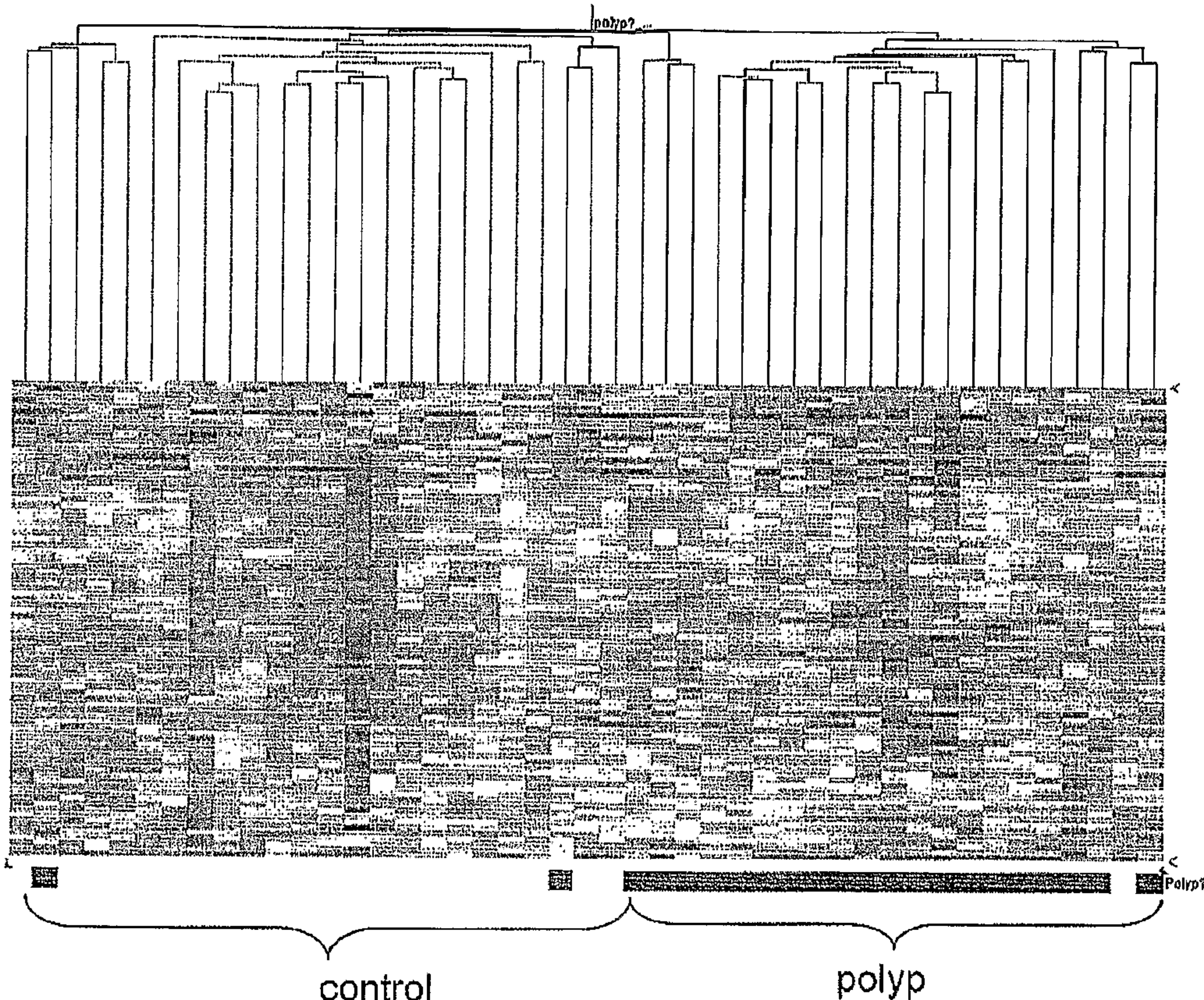




(86) Date de dépôt PCT/PCT Filing Date: 2006/10/23
(87) Date publication PCT/PCT Publication Date: 2007/04/26
(85) Entrée phase nationale/National Entry: 2008/04/18
(86) N° demande PCT/PCT Application No.: US 2006/041600
(87) N° publication PCT/PCT Publication No.: 2007/048074
(30) Priorités/Priorities: 2005/10/21 (US60/729,055);
2006/01/12 (US60/758,418)

(51) Cl.Int./Int.Cl. *C12Q 1/68* (2006.01)
(71) Demandeur/Applicant:
GENENEWS INC., CA
(72) Inventeurs/Inventors:
LIEW, CHOONG-CHIN, CA;
HAN, MARK, CA;
YAGER, THOMAS, CA;
CHAO, SAMUEL, CA;
ZHENG, RUN, CA;
ZHANG, HONGWEI, CA
(74) Agent: BORDEN LADNER GERVAIS LLP

(54) Titre : PROCEDE ET APPAREIL PERMETTANT DE METTRE DES NIVEAUX DE PRODUITS DE BIOMARQUEURS
EN CORRELATION AVEC UNE MALADIE
(54) Title: METHOD AND APPARATUS FOR CORRELATING LEVELS OF BIOMARKER PRODUCTS WITH DISEASE



(57) Abrégé/Abstract:

In one aspect the invention is a method of testing for one or more colorectal pathologies or one or more subtypes of colorectal pathology (in one embodiment colorectal cancer) in a test individual by providing data corresponding to a level of products of



(57) **Abrégé(suite)/Abstract(continued):**

selected biomarkers and applying the data to a formula to provide an indication of whether the test individual has one or more colorectal pathologies or one or more subtypes of colorectal pathology. In some aspects the method is computer based and a computer applies the data to the formula. In other aspects a computer system is configured with instructions that cause the processor to provide a user with the indication of whether the test individual has colorectal pathology. Also encompassed are kits for measuring data corresponding to the products of selected biomarkers which in some embodiments include a computer readable medium. Also encompassed are kits and methods of monitoring therapeutic efficacy of treatments for one or more colorectal pathologies.

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
26 April 2007 (26.04.2007)

PCT

(10) International Publication Number
WO 2007/048074 A1

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

Hongwei [CA/CA]; 3 Atwood Place, Toronto, Ontario, M9V 1L1 (CA).

(21) International Application Number:
PCT/US2006/041600

(74) Agent: FRASER, Janis, K.; FISH & RICHARDSON P.C., P.O. Box 1022, Minneapolis, MN 55440-1022 (US).

(22) International Filing Date: 23 October 2006 (23.10.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/729,055 21 October 2005 (21.10.2005) US
60/758,418 12 January 2006 (12.01.2006) US

(71) Applicant (for all designated States except US): GENE-NEWS INC. [CA/CA]; Unit 15-800 Petrolia Road, Toronto, Ontario, M3J 3K4 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): LIEW, Choong-chin [CA/CA]; 81 Millersgrove Drive, Toronto, Ontario, M2R 3S1 (CA). HAN, Mark [CN/CA]; 5-3 Four Winds Dr., North York, Ontario, M3J 1K7 (CA). YAGER, Thomas [US/CA]; 1268 Cermel Drive, Mississauga, Ontario, L5H 3V4 (CA). CHAO, Samuel [CA/CA]; 114 Glen Shield Avenue, Concord, Ontario, L4K 1T6 (CA). ZHENG, Run [CA/CA]; 20 Observatory Lane Unit 135, Richmond Hill, Ontario, L4C 0M7 (CA). ZHANG,

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, 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, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

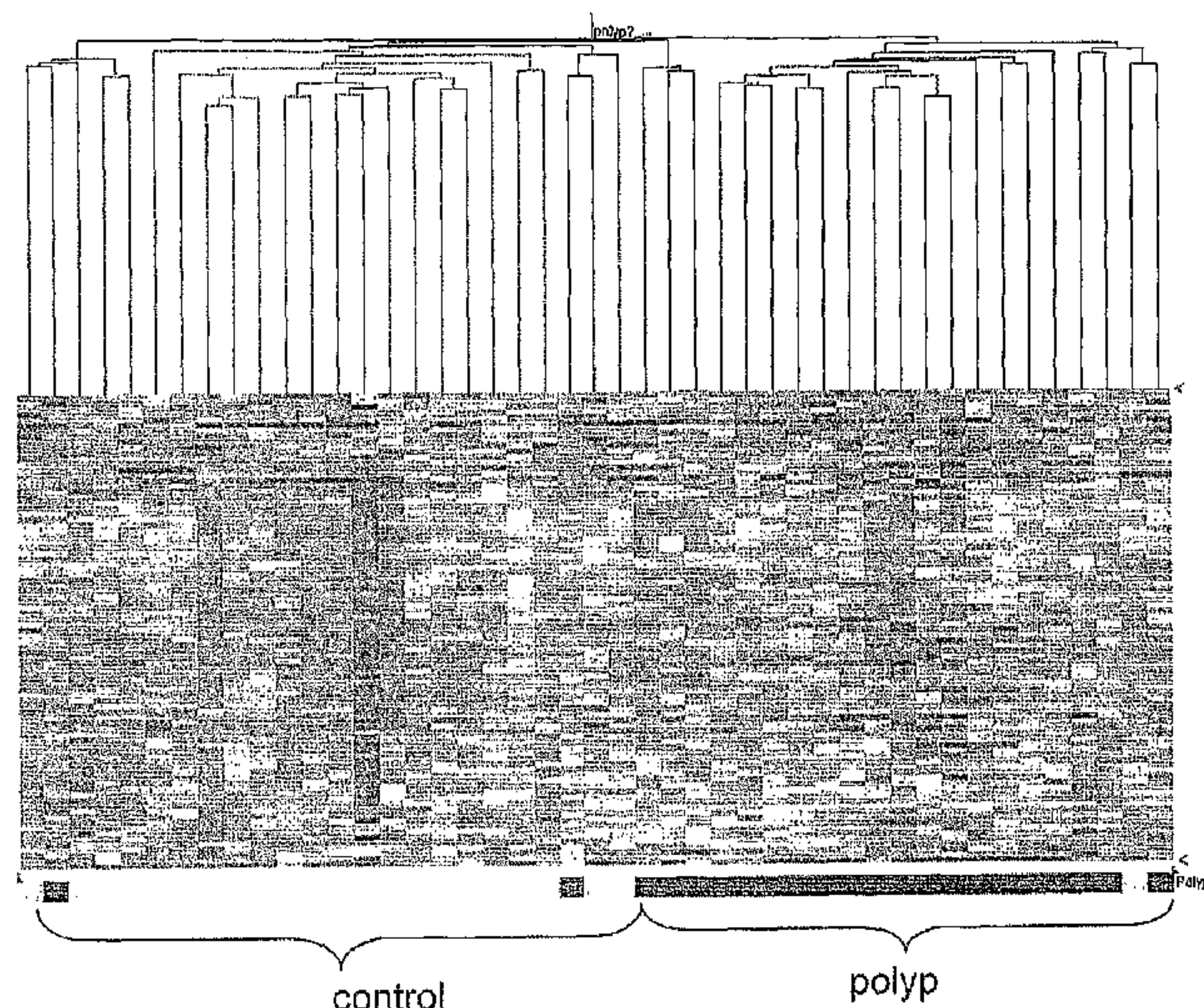
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR CORRELATING LEVELS OF BIOMARKER PRODUCTS WITH DISEASE



(57) Abstract: In one aspect the invention is a method of testing for one or more colorectal pathologies or one or more subtypes of colorectal pathology (in one embodiment colorectal cancer) in a test individual by providing data corresponding to a level of products of selected biomarkers and applying the data to a formula to provide an indication of whether the test individual has one or more colorectal pathologies or one or more subtypes of colorectal pathology. In some aspects the method is computer based and a computer applies the data to the formula. In other aspects a computer system is configured with instructions that cause the processor to provide a user with the indication of whether the test individual has colorectal pathology. Also encompassed are kits for measuring data corresponding to the products of selected biomarkers which in some embodiments include a computer readable medium. Also encompassed are kits and methods of monitoring therapeutic efficacy of treatments for one or more colorectal pathologies.

WO 2007/048074 A1

WO 2007/048074 A1



— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Date of publication of the amended claims: 28 June 2007

Published:

- *with international search report*
- *with amended claims*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 3
CONTENANT LES PAGES 1 À 230

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 3
CONTAINING PAGES 1 TO 230

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

METHOD AND APPARATUS FOR CORRELATING LEVELS OF BIOMARKER PRODUCTS WITH DISEASE

PRIORITY CLAIM

This application claims the benefit of priority under 35 U.S.C. § 119(e) to U.S. provisional
5 application serial no. 60/729,055, filed October 21, 2005, and to U.S. provisional application
serial no. 60/758,418, filed January 12, 2006, both of which are incorporated herein by reference
in their entirety.

10

TECHNICAL FIELD

The disclosure relates to, apparatus and kits and computer based methods for correlating
15 data corresponding to levels of biomarker products with a disease state in a subject.

BACKGROUND

Colorectal cancer is the second-leading cause of cancer-related deaths in the United
States (11). Each year, approximately 150,000 people are diagnosed with colorectal cancer and
almost 60,000 people die from the disease. Of those diagnosed, nearly half are expected to die
20 within five years, since most cancers are detected when the cancer is less treatable. For those
whose cancer is detected at an earlier stage, the five-year survival rate can be greater than 90%.
The American Cancer Society recommends that all Americans age 50 and older be screened
regularly for colorectal cancer. Unfortunately, only a fraction of this population is screened for

the disease, as currently available screening technologies are considered as either too costly, and/or too invasive or in some cases insufficiently accurate.

Most colorectal cancers begin as small, noncancerous (benign) clumps of cells called polyps. Over time some of these polyps become cancerous. Incidence of polyps increase as individuals get older. It is estimated that 50% of the people over the age of 60 will have at least one polyp.

The significance of identifying one or more colorectal pathologies including polyps is that certain types of polyps are cancerous or indicative of an increased risk to develop cancer. It has been shown that the removal of certain subtypes of polyps reduces the risk of getting colorectal cancer significantly. Therefore, a test to screen for one or more colorectal pathologies including polyps and/or certain subtypes of polyps so as to allow early removal or to prevent unnecessary procedures should markedly reduce the incidence of colorectal cancer (12) and decrease the current costs to the medical system.

Currently utilized screening technologies to identify polyps include 1) a fecal occult blood test (FOBT); 2) a flexible sigmoidoscopy; 3) double contrast barium enema (DCBE); and 4) colonoscopy. Sometimes two or more of these tests are used in combination. The current recommended standards for screening for colorectal cancer in men over the age of 50 and who are considered part of an average risk population include: an FOBT yearly, a sigmoidoscopy every five years, a colonoscopy every ten years and a DCBE every five years. For a high risk population where one or more family members have had colorectal cancer, a colonoscopy is recommended every two years as a follow up to FOBT or sigmoidoscopy.

Each of these tests suffers significant disadvantages. FOBT testing, although a non-invasive procedure, requires significant dietary and other restrictions prior to testing and suffers from a low sensitivity. Sigmoidoscopy and colonoscopy are more sensitive since they involve direct visualization of the lumen, however, sigmoidoscopy only allows partial visualization, and the colonoscopy is known to miss about 12% of advanced adenomas. Both sigmoidoscopy and colonoscopy are highly invasive procedures which cause high levels of discomfort causing many individuals to opt not to undergo these recommended screening procedures. Also sigmoidoscopy and colonoscopy are costly, and may have complications which arise as a result of undergoing the procedure.

Thus, there is a need for an improved test which is minimally invasive so as to permit more widespread testing of the population to indicate the presence of one or more colorectal pathologies, and ensure greater adherence to recommended protocols. To date, despite this need, there have been very few advancements in identifying useful molecular biomarkers to test for colorectal pathology. Recent efforts have focused on DNA based biomarker methods (see for
5 example Shuber et al. U.S. Patent Application Publication No. 2005-0260638A1; or Lofton-Day et al. WO2005/001142).

Identification of biomarkers for use in a non-invasive test for colorectal pathology thus fulfills a longstanding need in the art.

SUMMARY

In contrast to technologies available in the art, the inventions described herein identify biomarkers not previously associated with colorectal pathology whose gene expression levels, measured alone or in combination, and optionally applied to formulas to convert the levels to a measure, give an indication of a likelihood of colorectal pathology.

15 The present invention discloses novel colorectal pathology-specific biomarkers, such as blood-specific biomarkers, and methods, compositions and kits for use in testing for colorectal pathologies, such as pre-cancerous and cancerous pathologies. This use can be effected in a variety of ways as further described and exemplified herein.

According to one aspect of the present invention there is provided a method of testing for
20 one or more colorectal pathologies in a test subject, the method comprising: (a) providing data representative of a level of one or more products of each of one or more biomarkers in a sample from the test subject; and (b) ascertaining whether the data characterizes either: (i) subjects having the one or more colorectal pathologies, or (ii) subjects not having the one or more colorectal pathologies; thus providing an indication of a probability that the test subject has the
25 one or more colorectal pathologies.

According to another aspect of the present invention there is provided a computer-based method for testing for one or more colorectal pathologies in a test subject, the method comprising: inputting, to a computer, data representing a level of products of each of one or more biomarkers in a sample isolated and/or derived from a test subject, wherein the biomarkers are
30 genes selected from the group consisting of BCNP1, CD163, CDA, MS4A1, BANK1 and

REPLACEMENT SHEET

MGC20553; and causing the computer to ascertain whether the data characterizes either: (i) subjects having the one or more colorectal pathologies, or (ii) subjects not having the one or more colorectal pathologies, thus providing an indication of a probability that the test subject has the one or more colorectal pathologies.

5 According to still another aspect of the present invention there is provided a computer-readable medium comprising instructions for ascertaining whether data characterizes either: (i) subjects having one or more colorectal pathologies, or (ii) subjects not having one or more colorectal pathologies, the data representing a level of one or more products of each of one or more biomarkers in a sample isolated and/or derived from a test subject, wherein the biomarkers
10 are genes selected from the group consisting of BCNP1, CD163, CDA, MS4A1, BANK1, and MGC20553, thus providing an indication of a probability that the test subject has the one or more colorectal pathologies.

According to a yet another aspect of the present invention there is provided a computer system for providing an indication of a probability that a test subject has one or more colorectal
15 pathologies, the computer system comprising a processor; and a memory configured with instructions that cause the processor to provide a user with the indication, wherein the instructions comprise ascertaining whether data characterizes either: (i) subjects having one or more colorectal pathologies, or (ii) subjects not having one or more colorectal pathologies, the data representing a level of one or more products of each of one or more biomarkers in a sample
20 isolated or derived from the test subject, wherein the biomarkers are genes selected from the group consisting of BCNP1, CD163, CDA, MS4A1, BANK1, and MGC20553; thus providing the indication of a probability that the test subject has the one or more colorectal pathologies.

According to further features in preferred embodiments of the invention described below, the products in the sample are RNA.

25 According to still further features in the described preferred embodiments, the products in the sample are RNA, whereas the data represent a level of cDNA, EST and/or PCR product derived from the RNA.

According to still another aspect of the present invention there is provided a kit comprising packaging and containing one or more primer sets, wherein each set of which is able to generate
30 an amplification product by selective amplification of at least a portion of a polynucleotide complementary to one or more RNA products of a biomarker, wherein the biomarker is a gene

REPLACEMENT SHEET

selected from the group consisting of: BCNP1, CD163, CDA, MS4A1, BANK1, and MGC20553; and wherein each set of the primer sets is selective for a different biomarker.

According to further features in preferred embodiments of the invention described below, the complementary polynucleotide is selected from the group consisting of total RNA, mRNA,
5 DNA, cDNA and EST.

According to still further features in the described preferred embodiments, the one or more biomarkers are at least two biomarkers.

According to still further features in the described preferred embodiments, each of the probes is capable of selectively hybridizing to either a sense or an antisense strand of the
10 amplification product.

According to still further features in the described preferred embodiments, the kit further comprises two or more components selected from the group consisting of: a thermostable polymerase, a reverse transcriptase, deoxynucleotide triphosphates, nucleotide triphosphates and enzyme buffer.

15 According to still further features in the described preferred embodiments, the kit further comprises a computer-readable medium encoded with instructions for ascertaining whether data characterizes either: (i) subjects having one or more colorectal pathologies, or (ii) subjects not having one or more colorectal pathologies, the data representing levels of the amplification product in a sample isolated and/or derived from a test subject, thus providing an indication of a
20 probability that the test subject has the one or more colorectal pathologies.

According to still further features in the described preferred embodiments, ascertaining whether the data characterizes either: (i) subjects having the one or more colorectal pathologies, or (ii) subjects not having the one or more colorectal pathologies, comprises applying to the data a formula based on (i) a dataset representing levels of one or more products of each of the
25 biomarkers in each subject of a reference population having the one or more pathologies, and (ii) a dataset representing levels of one or more products of each of the biomarkers in each subject of a reference population not having the one or more pathologies.

According to still further features in the described preferred embodiments, ascertaining whether the data characterizes either: (i) subjects having the one or more colorectal pathologies,
30 or (ii) subjects not having the one or more colorectal pathologies, comprises ascertaining whether the data correlates more closely with (i) a dataset representing levels of one or more products of

each of the biomarkers in each subject of a reference population of subjects who have the one or more colorectal pathologies, or (ii) a dataset representing levels of one or more products of each of the biomarkers in each subject of a reference population of subjects who do not have the colorectal pathology.

5 According to still further features in the described preferred embodiments, the formula has a form: $V = C + \sum \beta_i X_i$, wherein V is a value indicating a probability that the test subject has the colorectal pathology, X_i is a level of products of an i th biomarker of the biomarkers in the sample, β_i is a coefficient, and C is a constant.

10 According to still further features in the described preferred embodiments, the formula has a form: $V = C + \sum \beta_{ij}(X_i / X_j)$, wherein V is a value indicating a probability that the test subject has the colorectal pathology, X_i is a level of products of an i th biomarker of the biomarkers, and X_j is a level of products of a j th biomarker of the biomarkers in the sample, where the i th biomarker is not the j th biomarker, β_{ij} is a coefficient, and C is a constant.

15 According to still further features in the described preferred embodiments, the formula is derived by a method selected from the group consisting of logistic regression, linear regression, neural networks, and principle component analysis.

According to still further features in the described preferred embodiments, the sample is selected from the group consisting of blood, lymph and lymphoid tissue.

20 According to still further features in the described preferred embodiments, the sample is selected from the group consisting of a sample of serum-reduced blood, a sample of erythrocyte-reduced blood, a sample of serum-reduced and erythrocyte-reduced blood, a sample of unfractionated cells of lysed blood, and a sample of fractionated blood.

25 According to a further aspect of the present invention there is provided a composition comprising a collection of two or more isolated polynucleotides, wherein each isolated polynucleotide selectively hybridizes to a biomarker selected from the biomarkers set out in Table 2 and wherein the composition is used to measure the level of expression of at least two of the biomarkers.

30 According to further features in preferred embodiments of the invention described below, each isolated polynucleotide selectively hybridizes a biomarker selected from the group consisting of membrane-bound transcription factor protease site 1 (MBTPS1); MGC45871; muskellin 1 (MKLN1); nipped-B homolog (NIPBL); acylpeptide hydrolase (APEH); FLJ23091;

MGC40157; and protein phosphatase 1 regulatory subunit 2 (PPP1R2); and wherein the composition is used to measure the level of expression of at least two of the biomarkers.

According to yet a further aspect of the present invention there is provided a composition comprising a collection of two or more isolated polynucleotides, wherein each isolated
5 polynucleotide selectively hybridizes to (a) an RNA product of a biomarker selected from the biomarkers set out in Table 2, and/or (b) a polynucleotide sequence complementary to (a), wherein the composition is used to measure the level of RNA expression of at least two of the biomarkers.

According to further features in preferred embodiments of the invention described below,
10 each isolated polynucleotide selectively hybridizes to (a) an RNA product of a biomarker selected from the group consisting of membrane-bound transcription factor protease site 1 (MBTPS1); MGC45871; muskelin 1 (MKLN1); nipped-B homolog (NIPBL); acylpeptide hydrolase (APEH); FLJ23091; MGC40157; and protein phosphatase 1 regulatory subunit 2 (PPP1R2); and/or (b) a polynucleotide sequence complementary to (a), wherein the composition
15 is used to measure the level of RNA expression of at least two of the biomarkers.

According to still a further aspect of the present invention there is provided a composition comprising a collection of two or more isolated polynucleotides, wherein each isolated polynucleotide selectively hybridizes to (a) an RNA sequences set out in Table 3; and/or (b) a polynucleotide sequences complementary to (a).

20 According to an additional aspect of the present invention there is provided a composition comprising a collection of two or more sets of biomarker specific primers as set out in Table 4 and/or Table 6.

According to yet an additional aspect of the present invention there is provided a composition comprising two or more polynucleotide probes as set out in Table 4.

25 According to further features in preferred embodiments of the invention described below, the polynucleotides are useful in quantitative RT-PCR (QRT-PCR).

According to still further features in the described preferred embodiments, the isolated polynucleotides comprise single or double stranded RNA.

30 According to still further features in the described preferred embodiments, the isolated polynucleotides comprise single or double stranded DNA.

According to still an additional aspect of the present invention there is provided a composition comprising a collection of two or more isolated proteins, wherein each isolated protein binds selectively to a protein product of a biomarker selected from the biomarkers set out in Table 2 and wherein the composition is used to measure the level of expression of at least two
5 of the biomarkers.

According to further features in preferred embodiments of the invention described below, each isolated protein binds selectively to a protein product of a biomarker selected from the group consisting of membrane-bound transcription factor protease site 1 (MBTPS1); MGC45871; muskelin 1 (MKLN1); nipped-B homolog (NIPBL); acylpeptide hydrolase
10 (APEH); FLJ23091; MGC40157; and protein phosphatase 1 regulatory subunit 2 (PPP1R2); and wherein the composition is used to measure the level of expression of at least two of the biomarkers.

According to still further features in the described preferred embodiments, the isolated proteins are selected from the proteins set out in Table 5.

15 According to still further features in the described preferred embodiments, the isolated proteins are ligands.

According to still further features in the described preferred embodiments, the ligands are antibodies or fragments thereof.

20 According to still further features in the described preferred embodiments, the antibodies are monoclonal antibodies.

According to yet still an additional aspect of the present invention there is provided a method of diagnosing or detecting one or more colon pathologies in an individual comprising: (a) determining the level of RNA product of one or more biomarker selected from the group consisting of the biomarkers set out in Table 2 in a sample of an individual; and (b) comparing
25 the level of RNA products in the sample with a control, wherein detecting differential expression of the RNA products between the individual and the control is indicative of a one or more colon pathologies in the individual.

According to further features in preferred embodiments of the invention described below, the method of diagnosing or detecting one or more colon pathologies in an individual
30 comprising: (a) determining the level of RNA product of one or more biomarker selected from the group consisting of the biomarkers set out in Table 2 in a sample of an individual; and (b)

comparing the level of RNA products in the sample with a control further comprises (a) determining the level of RNA product of one or more biomarker selected from the group consisting of membrane-bound transcription factor protease site 1 (MBTPS1); MGC45871; muskelin 1 (MKLN1); nipped-B homolog (NIPBL); acylpeptide hydrolase (APEH); FLJ23091; 5 MGC40157; and protein phosphatase 1 regulatory subunit 2 (PPP1R2); in a sample from an individual; and (b) comparing the level of RNA products in the sample with a control, wherein detecting differential expression of the RNA products between the individual and the control is indicative of a one or more colon pathologies in the individual.

10 According to still further features in the described preferred embodiments, the sample comprises whole blood.

According to still further features in the described preferred embodiments, the sample comprises a drop of whole blood.

According to still further features in the described preferred embodiments, the sample comprises blood that has been lysed.

15 According to still further features in the described preferred embodiments, prior to the determining step, the method comprises isolating RNA from the sample.

According to still further features in the described preferred embodiments, the step of determining the level of the RNA products comprises using quantitative RT-PCR (QRT-PCR).

20 According to still further features in the described preferred embodiments, the QRT-PCR comprises hybridizing primers which hybridize to the one or more RNA products or the complement thereof.

According to still further features in the described preferred embodiments, the primers are 15-25 nucleotides in length.

25 According to still further features in the described preferred embodiments, the step of determining the level of each of the one or more RNA products comprises hybridizing a first plurality of isolated polynucleotides that correspond to the one or more transcripts, to an array comprising a second plurality of isolated polynucleotides.

According to still further features in the described preferred embodiments, the first plurality of isolated polynucleotides comprises RNA, DNA, cDNA, PCR products or ESTs.

According to still further features in the described preferred embodiments, the array comprises a plurality of isolated polynucleotides comprising RNA, DNA, cDNA, PCR products or ESTs.

5 According to still further features in the described preferred embodiments, the second plurality of isolated polynucleotides on the array comprises polynucleotides corresponding to one or more of the biomarkers of Table 2.

According to still further features in the described preferred embodiments, the control is derived from an individual that does not have one or more colon pathologies.

10 According to another aspect of the present invention there is provided a kit for diagnosing or detecting one or more colon pathologies comprising any one of the compositions and instructions for use.

According to yet another aspect of the present invention there is provided a kit for diagnosing or detecting one or more colon pathologies comprising: (a) at least two sets of biomarker specific primers wherein each set of biomarker specific primers produces double
15 stranded DNA complementary to a unique biomarker selected from Table 2; wherein each first primers of the sets contains a sequence which can selectively hybridize to RNA, cDNA or an EST complementary to one of the biomarkers to create an extension product and each the second primers of the sets is capable of selectively hybridizing to the extension product; (b) an enzyme with reverse transcriptase activity; (c) an enzyme with thermostable DNA polymerase activity,
20 and (d) a labeling means; wherein each of the primer sets is used to detect the quantitative expression levels of the biomarker in a test subject.

According to further features in preferred embodiments of the invention described below, the kit for diagnosing or detecting one or more colon pathologies comprising: (a) at least two sets of biomarker specific primers wherein each set of biomarker specific primers produces double
25 stranded DNA complementary to a unique biomarker selected from Table 2; and an enzyme further comprises (a) at least two sets of biomarker specific primers wherein each set of biomarker specific primers produces double stranded DNA complementary to a unique biomarker selected from the group consisting of membrane-bound transcription factor protease site 1 (MBTPS1); MGC45871; muskelin 1 (MKLN1); nipped-B homolog (NIPBL); acylpeptide
30 hydrolase (APEH); FLJ23091; MGC40157; and protein phosphatase 1 regulatory subunit 2 (PPP1R2); wherein each first primers of the sets contains a sequence which can selectively

hybridize to RNA, cDNA or an EST complementary to one of the biomarkers to create an extension product and each the second primers of the sets is capable of selectively hybridizing to the extension product; (b) an enzyme with reverse transcriptase activity; (c) an enzyme with thermostable DNA polymerase activity, and (d) a labeling means; wherein each of the primer
5 sets is used to detect the quantitative expression levels of the biomarker in a test subject.

According to still another aspect of the present invention there is provided a method for diagnosing or detecting one or more colon pathologies in an individual comprising: (a) determining the level of protein product of one or more biomarker selected from the group consisting of the biomarkers set out in Table 2 in a sample from an individual; and (b) comparing
10 the level of protein products in the sample with a control, wherein detecting differential expression of the protein products between the individual and the control is indicative of a one or more colon pathologies in the individual.

According to further features in preferred embodiments of the invention described below, the method for diagnosing or detecting one or more colon pathologies in an individual comprising determining the level of protein product of one or more biomarker selected from the group consisting of the biomarkers set out in Table 2 in a sample from an individual further
15 comprises determining the level of protein product of one or more biomarker selected from the group consisting of membrane-bound transcription factor protease site 1 (MBTPS1); MGC45871; muskelin 1 (MKLN1); nipped-B homolog (NIPBL); acylpeptide hydrolase (APEH); FLJ23091; MGC40157; and protein phosphatase 1 regulatory subunit 2 (PPP1R2); in a
20 sample from an individual; and (b) comparing the level of protein products in the sample with a control, wherein detecting differential expression of the protein products between the individual and the control is indicative of a one or more colon pathologies in the individual.

According to still further features in the described preferred embodiments, the level of
25 protein product is determined using antibodies or fragments thereof.

According to still further features in the described preferred embodiments, the antibodies are selected from the group of antibodies set out in Table 5.

According to still further features in the described preferred embodiments, the antibodies are monoclonal antibodies.

30 According to a further aspect of the present invention there is provided a composition comprising a collection of two or more isolated polynucleotides, wherein each isolated

polynucleotide selectively hybridizes to a biomarker selected from the biomarkers set out in Table 12 and wherein the composition is used to measure the level of expression in blood of at least two of the biomarkers.

According to yet a further aspect of the present invention there is provided a kit for
5 diagnosing or detecting one or more colon pathologies comprising any one of the compositions comprising a collection of two or more isolated proteins and instructions for use.

The present invention successfully addresses the shortcomings of the presently known configurations, in particular by providing an effective and non-invasive method of testing for colorectal pathologies, such as pre-cancerous and cancerous pathologies, via biomarker analysis
10 in surrogate tissues such as blood.

Other features and advantages of the invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples while indicating preferred embodiments of the invention are given by way of
15 illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in relation to the drawings in which:

FIG. 1 shows an exemplary computer system for practicing certain of the methods
20 described herein.

FIG. 2 is a comparison of gene expression profiles obtained from hybridization of RNA isolated from serum-reduced, erythrocyte-reduced blood from 23 controls having no identified colorectal polyps and 22 subjects with colorectal polyps as described in **Example 2**. Patients with colorectal polyps could have one or more of the subtypes of polyps including Hyperplastic;
25 Tubular Adenoma; Villous Adenoma; Tubulovillous Adenoma; Hyperplasia; High Grade Dysplasia and Colorectal cancer; Gene expression profiles were clustered according to the expression of 86 significantly ($P < 0.001$) differentially expressed genes. As noted in grayscale, some of the individuals were misclassified (ie are shown in different grey scale under the appropriate bracket) and are considered outliers. Each column indicates a gene expression
30 profile for a single sample and each row represents the expression level of a single gene in each

of the samples. The color of each band within a row indicates the relative level of gene expression (grayscale represents level of expression, from low to high in expression). The resulting gene list from FIG. 2 is shown in **Table 1**.

FIG. 3 shows blood mRNA levels as tested by QRT-PCR for four upregulated genes (A) and four downregulated genes (B) selected from the genes demonstrating statistically significant differentially expression using microarray analysis as described in **Example 5**. QRT-PCR results were tested as between 50 patients diagnosed as having colorectal pathology (ie one or more subtypes of colorectal polyps) (n=50) as compared with 78 control individuals where the control individuals were diagnosed as not having colorectal pathology (n=78). Comparative (Ct) method based fold change was used. Mann-Whitney test was used for statistical analysis between two groups. Results which demonstrated a p value of less than 0.05 were considered statistically significant suggesting that the gene corresponding to the mRNA level tested is differentially expressed as between the two patient populations (patients with colorectal pathology and patients without colorectal pathology). The lines inside the boxes denote the medians. The boxes mark the interval between the 25th and 75th percentiles. The whiskers denote the interval between the 10th and 90th percentiles. • indicates data points outside the 10th and 90th percentiles.

FIG. 4 depicts an ROC curve for a selected biomarker combination from combinations of pairs of the eight biomarkers tested and shown in **FIG. 3**. Experimental details are as described in **Example 5**. Combinations of a selection of the biomarkers identified in **Table 1** were tested to determine the ability of the combinations to screen test subjects for one or more colorectal pathologies more effectively than can be achieved using the biomarkers of **Table 1** individually. QRT-PCR as described in **Example 3** was performed to measure the level of the RNA products of a select group of individual biomarkers from **Table 1**. Selected combinations of biomarkers were tested by applying logistic regression analysis to the QRT-PCR results of the selected combination and the ROC curve for the resulting logistic regression equation (logit function) determined. Panel (A) - ROC curve (ROC Area 0.72) for logit function for one of the datasets tested (AJ36h). This function returned by Simple Logistic operator in WEKA (ROC Area 0.66).

FIG. 5 depicts the graphical output results of the analysis of all possible combinations of 9 genes selected from the genes depicted in **Table 12** and further described in **Example 8**. Shown is a graphical depiction of ROC area, sensitivity (specificity is set at the 90% threshold) and

specificity (sensitivity is set at the 90% threshold) for each possible combination of 1, 2, 3, 4, 5, 6, 7, 8 or all 9 genes. Further details are described in **Example 8**.

FIG. 6 depicts the graphical output results of the analysis of all possible combinations of 6 genes as further described in **Example 11**. Shown is a graphical depiction of the ROC area, sensitivity (specificity is set at the 90% threshold) and specificity (sensitivity is set at the 90% threshold) for each possible combination of 1, 2, 3, 4, 5 and all 6 of the genes noted in **Example 11**.

BRIEF DESCRIPTION OF THE TABLES

Table 1 (provided herewith on CD-R) shows genes identified as differentially expressed in samples from individuals having or not having one or more of any type of colorectal polyps where the polyps can include be one or more of the following subtypes of polyps: Hyperplastic; Tubular Adenoma; Villous Adenoma; Tubulovillous Adenoma; Hyperplasia; High Grade Dysplasia and Cancer. The table provides the Hugo Gene name (second column), symbol and locus link ID; the human RNA and protein accession number; the p value (which represents the statistical significance of the observed differential expression as determined by measuring RNA encoded by the biomarkers noted) and a measure of the fold change as between the average measured level of individuals having one or more types of colorectal pathology and the average measured level of individuals not having colorectal pathology. Column 1 is AffySpotID, column 2 is GeneSymbol, column 3 is GeneID, Column 4 is pvalue, column 5 is the HumanRNA Accession Number, column 6 is the Human Protein Accession Number, column 7 is the Fold Change, and column 8 is the Gene Description.

Table 2 is a selection of genes listed in **Table 1**. The table provides the gene symbol, locus link ID, and gene description. The table also includes the p value (which represents the statistical significance of the observed differential expression), the Mann-Whitney value (which is another measure to represent the statistical significance of differentiating samples with and without colorectal pathology), the measure of the fold change as between the average measured level of individuals having polyps and the average measured level of individuals not having polyps, and the direction of the differential expression between individuals having colorectal pathology and not having colorectal pathology.

Table 3 provides the human RNA accession number and human protein accession number of various species of the biomarkers identified as differentially expressed in samples from individuals having or not having colorectal polyps. The table provides the gene symbol and a description of the gene.

5 **Table 4** provides a selection of examples of primers and TaqMan® probes useful in the invention for measuring the RNA products of the biomarkers disclosed in **Table 2**.

Table 5 provides reference to antibodies which are commercially available for protein products of the genes identified in **Table 2**.

10 **Table 6** shows primer sequences utilized to perform RT-PCR to measure the differential expression of RNA products of the genes (biomarkers) from **Table 2** as noted in **Example 3** in samples from individuals having or not having one or more polyps. The table also provides the gene symbol and RNA accession number corresponding to the biomarkers tested.

15 **Table 7** provides a summary of the phenotypic information of the patients used as noted in **Example 5** to test selected biomarkers for the ability to test for the presence of colorectal polyps. Parameters including the sample size, gender, age and polyp subtype, (as determined by pathology reports), are listed.

20 **Table 8** lists selected classifiers for use with data corresponding to the eight selected biomarkers MBTPS1, MGC45871, MKLN1, NIPBL APEH, MGC40157, PPP1R2, FLJ23091 which resulted in a ROC area of 0.72 so as to test for the presence of a colorectal polyps as described in **Example 5**. Results of QRT-PCR of the eight selected genes are shown in **FIG. 3**.

Table 9 provides the results of the blind test described in **Example 5** when applying the formula comprised of classifiers noted in **Table 8**.

25 **Table 10** is a list of reporter genes and the properties of the reporter gene products which may be used to identify compounds for use in preventing or treating one or more forms of colorectal pathology.

30 **Table 11** (provided herewith on CD-R) shows genes identified as differentially expressed in samples from individuals having “high risk polyps” as compared with individuals not having high risk polyps (ie having low risk polyps or having no pathology at all) using microarray as described in **Example 2**. The table provides the gene name, gene ID; a representative human RNA accession number, and also provides the p value, the fold change (as between the average of individuals classified as having high risk polyps as compared with the average of individuals

having low risk polyps), along with the coefficient of variation for both the high risk polyp individuals and the low risk polyp individuals (the standard deviation of the normalized intensity divided by the mean normalized intensity). Column 1 is AffySpotID, column 2 is Fold Change, column 3 is p value, Column 4 is CV (Coefficient of Variation) (High RiskPolyp), column 5 is CV(Low Risk Polyp), column 6 is Gene ID, column 7 is the HUGO Gene Symbol, column 8 is the Human RNA Accession Number and column 9 is the Gene Description.

Table 12 shows 48 biomarkers tested for differential expression by QRT-PCR in samples from individuals having colorectal cancer and individuals not having colorectal cancer. The 48 biomarkers were tested using QRT-PCR. The table provides the gene symbol, locus link ID, and gene description for each biomarker. The table also includes the p value (which represents the statistical significance of the observed differential expression), the measure of the fold change as between the average measured level of individuals having colorectal cancer and the average measured level of individuals not having colorectal cancer and the direction of the differential expression between individuals having colorectal cancer and not having colorectal cancer.

Table 13 provides the human RNA accession number and human protein accession number of various species of the biomarkers identified as differentially expressed in samples from individuals having or not having colorectal cancer in **Table 12**. The table provides the gene symbol and a description of the gene.

Table 14 provides a selection of examples of primers and TaqMan® probes useful for measuring one or more RNA products of the biomarkers disclosed in **Table 12** as described in **Examples 6, 7, 8 or 9**.

Table 15 provides reference to antibodies which are commercially available to measure protein products of the biomarkers identified in **Table 12**.

Table 16 shows a selection of primers which have been used for the genes described in **Table 12** to quantify one or more RNA products of the biomarkers.

Table 17 shows primers and TaqMan® probes used in **Example 11**.

DETAILED DESCRIPTION

(A) Overview

In one aspect the invention discloses methods of generating formulas/classifiers which can be applied to data corresponding to levels of one or more products of selected biomarker

combinations to classify test subjects as having one or more colorectal pathologies or one or more subtypes of colorectal pathologies. Also disclosed are biomarkers whose product levels are useful for testing subjects for one or more colorectal pathologies or one or more subtypes of colorectal pathologies. Also disclosed are computer-readable media comprising instructions for
5 applying a formula to data representing a level of products of biomarkers so as to test subjects for one or more colorectal pathologies. Also disclosed is a computer system which is configured with instructions to provide the user with an indication of the probability of a test subject having one or more colorectal pathologies by applying a formula to biomarker product data.

The present invention provides biomarker product ligands capable of specific
10 hybridization with RNA biomarker products so as to enable quantitation of the biomarker products. The biomarker product ligands may enable quantitation, either directly and/or indirectly in any one of various ways known to the skilled artisan. The biomarker product ligands capable of specifically hybridizing with biomarker RNA products or polynucleotides derived therefrom may have any one of various compositions. For example, specific ligands of
15 biomarker products, such as biomarker RNA products, may be either polynucleotides (e.g. polynucleotides complementary to at least a portion of the RNA products or polynucleotides derived therefrom) or polypeptides (e.g. antibodies or affinity-selected polypeptides specific for at least a portion of the RNA products or polynucleotides derived therefrom). In one embodiment polynucleotide and/or polypeptide ligands are disclosed which are probes capable
20 of specifically and/or selectively hybridizing with so as to quantitate biomarker RNA products and/or polynucleotides products. Such probes include those useful in techniques such as quantitative real-time PCR (QRT-PCR), and may be used, for example, with SYBR®Green, or using TaqMan® or Molecular Beacon techniques. In one embodiment, the polynucleotides useful as nucleic acid probes which can be spotted onto an array to measure levels of biomarker
25 RNA products, or of polynucleotides derived therefrom, which are isolated or derived from test subjects. In another embodiment, arrays for use in measuring the expression of the RNA products are contemplated.

In another embodiment, polynucleotide ligands are disclosed which are biomarker specific primer sets capable of specifically amplifying biomarker RNA products and/or polynucleotides
30 corresponding to biomarker RNA products.

Further disclosed are methods of screening of products of the identified biomarkers to screen for therapeutic targets for treating or preventing one or more colorectal pathologies.

Kits of polynucleotides and/or polypeptide ligands which can be used to detect and monitor differential gene expression of the products of the identified biomarkers and biomarker combinations are also provided as are kits which include a computer readable medium to allow an indication of a probability that a test subject has one or more colorectal pathologies. Methods of generating the formulas for testing for one or more colorectal pathologies are also provided.

Further disclosed is the measurement/monitoring of biomarker product levels to screen for therapeutic targets for treating or preventing one or more colorectal pathologies. Methods of generating the formulas for testing a test subject for a one or more colorectal pathologies are also provided.

Also disclosed are methods of testing combinations of biomarkers by generating classifiers. Classifiers are generated by applying one or more mathematical models to data representative of the level of expression of the RNA and/or protein products of the biomarkers across a reference population encompassing subjects who have one or more colorectal pathologies, or one or more subtypes thereof, and subjects who do not have the one or more colorectal pathologies. Classifiers can be used alone or in combination to create a formula which is useful in testing a subject for the probability of having one or more colon pathologies subtypes. Also disclosed are methods of further selecting classifiers on the basis of area under the curve (AUC), sensitivity and/or specificity. One or more selected classifiers can be used to generate a formula and subsequently classifiers can be selected for inclusion into the formulas. Classifiers are generated by measuring levels of one or more RNA products and/or one or more protein products of the biomarkers in a sample and using the data resulting from said measurement for input into the mathematical model. Note that it is not necessary that the same method used to generate the data for creating the formulas is the method used to generate data from the test subject for inclusion within the formula for diagnostic purposes.

Other aspects of the invention are disclosed herein.

(B) Definitions

The practice of the present invention will employ, unless otherwise indicated, techniques of molecular biology, microbiology and recombinant DNA techniques, which are familiar to one

of the skill in the art. Such techniques are explained fully in the literature. See, e.g., Sambrook, Fritsch & Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition; *Oligonucleotide Synthesis* (M.J. Gait, ed., 1984); *Nucleic Acid Hybridization* (B.D. Harnes & S.J. Higgins, eds., 1984); *A Practical Guide to Molecular Cloning* (B. Perbal, 1984); and a series, 5 *Methods in Enzymology* (Academic Press, Inc.); *Short Protocols In Molecular Biology*, (Ausubel et al., ed., 1995). All patents, patent applications, and publications mentioned herein, both supra and infra, are hereby incorporated by reference in their entireties.

The following definitions are provided for specific terms which are used in the following written description.

10 As used herein, the “5' end” refers to the end of an mRNA up to the first 1000 nucleotides or 1/3 of the mRNA (where the full length of the mRNA does not include the poly A tail) starting at the first nucleotide of the mRNA. The “5' region” of a gene refers to a polynucleotide (double-stranded or single-stranded) located within or at the 5' end of a gene, and includes, but is not limited to, the 5' untranslated region, if that is present, and the 5' protein coding region of a 15 gene. The 5' region is not shorter than 8 nucleotides in length and not longer than 1000 nucleotides in length. Other possible lengths of the 5' region include but are not limited to 10, 20, 25, 50, 100, 200, 400, and 500 nucleotides.

As used herein, the “3' end” refers to the end of an mRNA up to the last 1000 nucleotides or 1/3 of the mRNA, where the 3' terminal nucleotide is that terminal nucleotide of the coding or 20 untranslated region that adjoins the poly-A tail, if one is present. That is, the 3' end of an mRNA does not include the poly-A tail, if one is present. The “3' region” of a gene refers to a polynucleotide (double-stranded or single-stranded) located within or at the 3' end of a gene, and includes, but is not limited to, the 3' untranslated region, if that is present, and the 3' protein coding region of a gene. The 3' region is not shorter than 8 nucleotides in length and not longer 25 than 1000 nucleotides in length. Other possible lengths of the 3' region include but are not limited to 10, 20, 25, 50, 100, 200, 400, and 500 nucleotides. As used herein, the “internal coding region” of a gene refers to a polynucleotide (double-stranded or single-stranded) located between the 5' region and the 3' region of a gene as defined herein. The “internal coding region” is not shorter than 8 nucleotides in length and not longer than 1000 nucleotides in length. Other 30 possible lengths of the “internal coding region” include but are not limited to 10, 20, 25, 50, 100, 200, 400, and 500 nucleotides. The 5', 3' and internal regions are non-overlapping and may, but

need not be contiguous, and may, but need not, add up to the full length of the corresponding gene.

As used herein, the “amino terminal” region of a polypeptide refers to the polypeptide sequences encoded by polynucleotide sequences (double-stranded or single-stranded) located within or at the 5' end of an mRNA molecule. As used herein, the “amino terminal” region refers to the amino terminal end of a polypeptide up to the first 300 amino acids or 1/3 of the polypeptide, starting at the first amino acid of the polypeptide. The “amino terminal” region of a polypeptide is not shorter than 3 amino acids in length and not longer than 350 amino acids in length. Other possible lengths of the “amino terminal” region of a polypeptide include but are not limited to 5, 10, 20, 25, 50, 100 and 200 amino acids.

As used herein, the “carboxy terminal” region of a polypeptide refers to the polypeptide sequences encoded by polynucleotide sequences (double-stranded or single-stranded) located within or at the 3' end of an mRNA molecule. As used herein, the “carboxy terminal” region refers to the carboxy terminal end of a polypeptide up to 300 amino acids or 1/3 of the polypeptide from the last amino acid of the polypeptide. The “3' end” does not include the polyA tail, if one is present. The “carboxy terminal” region of a polypeptide is not shorter than 3 amino acids in length and not longer than 350 amino acids in length. Other possible lengths of the “carboxy terminal” region of a polypeptide include, but are not limited to, 5, 10, 20, 25, 50, 100 and 200 amino acids.

As used herein, the “internal polypeptide region” of a polypeptide refers to the polypeptide sequences located between the amino terminal region and the carboxy terminal region of a polypeptide, as defined herein. The “internal polypeptide region” of a polypeptide is not shorter than 3 amino acids in length and not longer than 350 amino acids in length. Other possible lengths of the “internal polypeptide region” of a polypeptide include, but are not limited to, 5, 10, 20, 25, 50, 100 and 200 amino acids. The amino terminal, carboxy terminal and internal polypeptide regions of a polypeptide are non-overlapping and may, but need not be contiguous, and may, but need not, add up to the full length of the corresponding polypeptide.

As used herein, the term “amplified”, when applied to a nucleic acid sequence, refers to a process whereby one or more copies of a particular nucleic acid sequence is generated from a template nucleic acid, in some embodiments by the method of polymerase chain reaction (Mullis and Faloona, 1987, Methods Enzymol., 155:335). “Polymerase chain reaction” or “PCR” refers

to a method for amplifying a specific template nucleic acid sequence. In some embodiments, the PCR reaction involves a repetitive series of temperature cycles and is typically performed in a volume of 50-100 μ l. The number of cycles performed in the PCR reaction can include 15, 20, 25, 30, 35, 40, 45, 50, 55 or 60 cycles. The reaction mix comprises dNTPs (each of the four deoxynucleotides dATP, dCTP, dGTP, and dTTP), primers, buffers, DNA polymerase, and nucleic acid template. The PCR reaction can comprise providing a set of polynucleotide primers wherein a first primer contains a sequence complementary to a region in one strand of the nucleic acid template sequence and primes the synthesis of a complementary strand, and a second primer contains a sequence complementary to a region in a second strand of the target nucleic acid sequence and primes the synthesis of a complementary strand, and amplifying the nucleic acid template sequence employing a nucleic acid polymerase as a template-dependent polymerizing agent under conditions which are permissive for PCR cycling steps of (i) annealing of primers required for amplification to a target nucleic acid sequence contained within the template sequence, (ii) extending the primers wherein the nucleic acid polymerase synthesizes a primer extension product. "A set of polynucleotide primers" or "a set of PCR primers" or "a set of primers" can comprise two, three, four or more primers. In some embodiments, nested PCR can occur using a primer set wherein a first subset of primers is utilized to amplify a single product and then a second subset of primers is utilized which hybridize to the product of the first subset of primers to amplify a smaller version of the. In one embodiment, an *exo*- Pfu DNA polymerase is used to amplify a nucleic acid template in PCR reaction. Other methods of amplification include, but are not limited to, ligase chain reaction (LCR), polynucleotide-specific based amplification (NSBA), or any other method known in the art.

In one aspect an "array" includes a specific set of probes, such as oligonucleotides and/or cDNA's (e.g. ESTs) corresponding in whole or in part, and/or continuously or discontinuously, to regions of expressed genomic DNA; wherein the probes are localized onto a support. In one embodiment, the probes can correspond to the 5' ends or 3' ends of the internal coding regions of a biomarker RNA product of the invention. Of course, mixtures of a 5' end of one gene may be used as a target or a probe in combination with a 3' end of another gene to achieve the same or similar biomarker RNA product level measurements.

As used herein, an "analog" of a reference proteinaceous agent includes any proteinaceous agent that possesses a similar or identical function as the reference proteinaceous agent but does

not comprise a similar or identical amino acid sequence as reference proteinaceous agent, and/or possess a similar or identical structure as the reference proteinaceous agent. A proteinaceous agent that has a similar amino acid sequence as a second proteinaceous agent is at least one of the following: (a) a proteinaceous agent having an amino acid sequence that is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 99% identical to the amino acid sequence of the second proteinaceous agent; (b) a proteinaceous agent encoded by a nucleotide sequence that hybridizes under stringent conditions to at least a segment of the nucleotide sequence encoding the second proteinaceous agent, where the segment has a length of at least 5 contiguous amino acid residues, at least 10 contiguous amino acid residues, at least 15 contiguous amino acid residues, at least 20 contiguous amino acid residues, at least 25 contiguous amino acid residues, at least 40 contiguous amino acid residues, at least 50 contiguous amino acid residues, at least 60 contiguous amino residues, at least 70 contiguous amino acid residues, at least 80 contiguous amino acid residues, at least 90 contiguous amino acid residues, at least 100 contiguous amino acid residues, at least 125 contiguous amino acid residues, or at least 150 contiguous amino acid residues; and (c) a proteinaceous agent encoded by a nucleotide sequence that is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 99% identical to the nucleotide sequence encoding the second proteinaceous agent. A proteinaceous agent with similar structure to a second proteinaceous agent refers to a proteinaceous agent that has a similar secondary, tertiary or quaternary structure as the second proteinaceous agent. The structure of a proteinaceous agent can be determined by methods known to those skilled in the art, including but not limited to, peptide sequencing, X-ray crystallography, nuclear magnetic resonance, circular dichroism, and crystallographic electron microscopy.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino acid or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the

corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity=number of identical overlapping positions/total number of positions.times.100%). In one embodiment, the two sequences are the same length.

5 The determination of percent identity between two sequences can also be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. U.S.A. 87:2264-2268, modified as in Karlin and Altschul, 1993, Proc. Natl. Acad. Sci. U.S.A. 90:5873-5877. Such an algorithm is incorporated into the NBLAST and
10 XBLAST programs of Altschul et al., 1990, J. Mol. Biol. 215:403. BLAST nucleotide searches can be performed with the NBLAST nucleotide program parameters set, e.g., for score=100, wordlength=12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the present invention. BLAST protein searches can be performed with the XBLAST program parameters set, e.g., to score=50, wordlength=3 to obtain amino acid sequences homologous to a
15 protein molecule of the present invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., 1997, Nucleic Acids Res. 25:3389-3402. Alternatively, PSI-BLAST can be used to perform an iterated search which detects distant relationships between molecules (Id.). When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (e.g., of XBLAST
20 and NBLAST) can be used (see, e.g., the NCBI website). Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, 1988, CABIOS 4:11-17. Such an algorithm is incorporated in the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue
25 table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically only exact matches are counted.

As used herein, the term "analog" in the context of a non-proteinaceous molecules refers
30 to a second organic or inorganic molecule which possess a similar or identical function as a first organic or inorganic molecule and is structurally similar to the first organic or inorganic

molecule. The term “analog” includes a molecule whose core structure is the same as, or closely resembles that of the first molecule, but which has a chemical or physical modification. The term “analog” includes copolymers of the first molecule that can be linked to other atoms or molecules. A “biologically active analog” and “analog” are used interchangeably herein to cover
5 an organic or inorganic molecule that exhibits substantially the same agonist or antagonist effect of the first organic or inorganic molecule.

A “nucleotide analog”, as used herein, refers to a nucleotide in which the pentose sugar and/or one or more of the phosphate esters is replaced with its respective analog. Exemplary phosphate ester analogs include, but are not limited to, alkylphosphonates, methylphosphonates,
10 phosphoramidates, phosphotriesters, phosphorothioates, phosphorodithioates, phosphoroselenoates, phosphorodiselenoates, phosphoroanilothioates, phosphoroanilidates, phosphoroamidates, boronophosphates, etc., including any associated counterions, if present. Also included within the definition of “nucleotide analog” are nucleobase monomers which can be polymerized into polynucleotide analogs in which the DNA/RNA phosphate ester and/or
15 sugar phosphate ester backbone is replaced with a different type of linkage. Further included within “nucleotide analogs” are nucleotides in which the nucleobase moiety is non-conventional, i.e., differs from one of G, A, T, U or C. Generally a non-conventional nucleobase will have the capacity to form hydrogen bonds with at least one nucleobase moiety present on an adjacent counter-directional polynucleotide strand or provide a non-interacting, non-interfering base.

20 The term “antibody” also encompasses antigen-binding fragments of an antibody. The term “antigen-binding fragment” of an antibody (or simply “antibody portion,” or “fragment”), as used herein, refers to one or more fragments of a full-length antibody that retain the ability to specifically bind to a polypeptide encoded by one of the genes of a biomarker of the invention. Examples of binding fragments encompassed within the term “antigen-binding fragment” of an
25 antibody include (i) a Fab fragment, a monovalent fragment consisting of the VL, VH, CL and CH1 domains; (ii) a F(ab')₂ fragment, a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; (iii) a Fd fragment consisting of the VH and CH1 domains; (iv) a Fv fragment consisting of the VL and VH domains of a single arm of an antibody, (v) a dAb fragment (Ward *et al.*, (1989) *Nature* 341:544-546), which consists of a VH
30 domain; and (vi) an isolated complementarity determining region (CDR). Furthermore, although the two domains of the Fv fragment, VL and VH, are coded for by separate genes, they can be

joined, using recombinant methods, by a synthetic linker that enables them to be made as a single protein chain in which the VL and VH regions pair to form monovalent molecules (known as single chain Fv (scFv); see *e.g.*, Bird *et al.* (1988) *Science* 242:423-426; and Huston *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:5879-5883). Such single chain antibodies are also intended to be
5 encompassed within the term “antigen-binding fragment” of an antibody. These antibody fragments are obtained using conventional techniques known to those with skill in the art, and the fragments are screened for utility in the same manner as are intact antibodies. The antibody is in some embodiments monospecific, *e.g.*, a monoclonal antibody, or antigen-binding fragment thereof. The term “monospecific antibody” refers to an antibody that displays a single binding
10 specificity and affinity for a particular target, *e.g.*, epitope. This term includes a “monoclonal antibody” or “monoclonal antibody composition,” which as used herein refer to a preparation of antibodies or fragments thereof of single molecular composition.

As used herein, the terms “attaching” and “spotting” in relation to an array can include a process of depositing or localizing a nucleic acid or proteinaceous agent onto a substrate to form
15 a nucleic acid or protein array. In one embodiment, the substance spotted is attached or localized onto the array via covalent bonds, hydrogen bonds or ionic interactions. .

As used herein, the term “biomarker” refers to a gene whose products can be measured and correlated with disease. A biomarker refers to a gene which encodes one or products (*e.g.* unspliced RNA, mRNA and/or polypeptide) present at measurably different levels in
20 corresponding samples isolated and/or derived from subjects having the pathology and subjects not having the pathology. A biomarker may be a DNA molecule which is transcribed into RNA product. Alternately, the biomarker may be an RNA molecule which is translated into protein product, or reverse-transcribed into DNA product.

As used herein, a “blood nucleic acid sample” or “blood polynucleotide sample”, refers
25 to polynucleotides derived from blood and can include polynucleotides isolated and/or derived from whole blood, serum-reduced whole blood, lysed blood (erythrocyte depleted blood), centrifuged lysed blood (serum-depleted, erythrocyte depleted blood), serum depleted whole blood or peripheral blood leukocytes (PBLs), globin reduced RNA from blood, or any possible fraction of blood as would be understood by a person skilled in the art. A blood polynucleotide
30 sample can refer to RNA, mRNA or a nucleic acid corresponding to mRNA, for example, cDNA

or EST derived from RNA isolated from said sample. A blood polynucleotide sample can also include a PCR product derived from RNA, mRNA or cDNA.

As used herein, the term "formula" includes one or more classifiers, or combination of classifiers where the term classifier is used to describe the output of a mathematical model.

5 As used herein the term "colorectal pathology" comprises any of one or more types or subtypes of pathology of the rectum and/or colon. "Colorectal pathologies" include pre-cancerous polyps, cancerous polyps, polyps at risk of becoming cancerous, and polyps of unknown cancer-related status. As would be understood, in some cases a subject according to embodiments of the invention can have at any one time one or more colorectal pathologies, each
10 of which being of the same or a different type or subtype of polyp. Colorectal pathologies may be classified in any one of various ways, for example as is known in the art. In one embodiment "Polyp" or "colorectal polyp" as would be understood in the art, includes an abnormal growth of cells and/or tissue, and/or a growth of cells and/or tissue that may project into the colon or rectum. A polyp can be further defined according to various factors including the morphology of
15 the polyp; the risk of the polyp developing into a cancerous polyp and the like as would be understood by a person ordinarily skilled in the art. In one embodiment, polyps can be classified into various subtypes including: Hyperplastic; Tubular Adenoma; Villous Adenoma; Tubulovillous Adenoma; Hyperplasia; High Grade Dysplasia; and Cancer. For any one individual and or polyp, one or more polyp subtype description could apply. In another
20 embodiment, (7) colorectal cancer can be subclassified into various categories as well. In yet another embodiment, one or more of the listed subtypes can be grouped together according to any one of various parameters in one category. Alternately, one or more of the listed subtypes can be further subclassified according to any one of various parameters. In yet another
25 embodiment, one or more of the listed subtypes can be grouped together according to any one of various parameters in one category. Alternately, one or more of the listed subtypes can be further subclassified according to any one of various parameters. For example, in one embodiment, Tubular Adenoma polyps can be further classified in accordance with the diameter of the Adenoma polyp. For example Adenoma polyps with a diameter greater than 1mm, 2mm, 3mm, 4mm, 5mm, 6mm, 7mm, 8mm, 9mm, 10mm, 11mm, 12mm, 13mm, 14mm or 15mm are
30 possible. In yet another example, colorectal cancer can be further subclassified in accordance with disease progression as would be understood. For example, colorectal cancer can be

subclassified using the Duke or Modified Duke Staging System. The Modified Duke Staging System groups colorectal cancer into four different stages A-D. Stage A indicates the tumor penetrating the mucosa of the colon and/or bowel but no further. Stage B1 indicates the tumor penetrating into, but not through the muscularis propria (the muscular layer) of the colon and/or
5 bowel wall. Stage B2 indicates the tumor has penetrated into and through the muscularis propria of the colon and/or bowel wall. A Stage C1 tumor penetrates into, but not through the muscularis propria of the colon and/or bowel wall; there is pathologic evidence of colorectal cancer in the lymph nodes. Stage C2 tumors penetrates into and through the muscularis propria of the bowel wall; but there is pathologic evidence of colorectal cancer in the lymph nodes.
10 Finally Stage D indicates the tumor has spread beyond the lymph nodes to other organs. In yet another embodiment, colorectal cancer can be subclassified using the TNM staging system. According to the TNM staging system there are four stages, stages I through IV, each reflecting status regarding Tumor, Node, and Metastasis. Tumor is subdivided as follows T1: Tumor invades submucosa, T2: Tumor invades muscularis propria, T3: Tumor invades through the
15 muscularis propria into the subserosa, or into the pericolic or perirectal tissues and T4: Tumor directly invades other organs or structures, and/or perforates. Node is subdivided as follows: N0 indicates no regional lymph node metastasis. N1 indicates metastasis in 1 to 3 regional lymph nodes. N2 indicates metastasis in 4 or more regional lymph nodes. Metastasis is divided as follows: M0 indicates no distant metastasis and M1 indicates distant metastasis present. Thus
20 for Stage I, in accordance with the TNM system, the tumor can either be categorized as T1N0M0 or T2N0M0; the cancer has begun to spread but is still in the inner lining. Stage II is T3N0M0 or T4 N0M0; the cancer has spread to other organs near the colon or rectum but has not yet reached the inner lining. Stage III includes all T's, N1-2 and M0; cancer has spread to lymph nodes but has not been carried to distant parts of the body. Stage IV includes any T, any N and
25 M1; cancer is metastatic and has been carried to other organs, likely the lung or liver.

As used herein, the term "high risk polyps" indicates those subtypes of polyps which are considered at higher risk for developing into cancer or are already cancerous as would be understood by a person skilled in the art, and includes cancer-prone polyps or cancer-disposed polyps and cancerous polyps, whereas a "low risk polyp" includes all other types of polyps. For
30 example, 70 to 90 percent of colorectal cancers arise from adenomatous polyps, and thus are considered high risk polyps. Adenomatous polyps can be further categorized into subtypes

including: Tubular adenoma, which has been suggested to have approximately a 4% potential for malignancy; Tubulovillous adenoma, which has been suggested to have approximately a 16% potential for malignancy and Villous adenoma, which has been suggested to have approximately a 21% potential for malignancy. In addition, high grade dysplasia has increased malignant potential. In one embodiment, polyps which are “high risk polyps” are Tubulovillous Adenoma, Villous Adenoma, High Grade Dysplasia and Tubular Adenoma and also includes polyps which are cancerous including those which are cancerous and localized and those which have already led to dissemination in the peripheral blood. In this embodiment, a “low risk polyp” includes any other polyp morphology. In another embodiment, polyps which are “high risk polyps” are Tubulovillous Adenoma, Villous Adenoma, High Grade Dysplasia and Tubular Adenoma and does not include polyps which are already cancerous. The size of the polyp also correlates with the risk for developing into cancer. For example, polyps greater than 10mm in diameter are considered large polyps and have a greater potential for malignancy. Polyps larger than 2 cm in diameter have a 50 percent chance of becoming malignant. See Zauber (2004) Gastroenterology; 126(5): 1474. In another embodiment “high risk polyps” comprise Tubulovillous Adenoma, Villous Adenoma, High Grade Dysplasia and Tubular Adenoma where the diameter of the Tubular Adenoma polyp is greater than 10mm and the remaining polyp morphologies are considered “low risk polyps”.

As used herein, the terms “compound” and “agent” are used interchangeably.

As used herein, the term “control” or “control sample” can include one or more samples isolated and/or derived from a subject or group of subjects who have been diagnosed as having one or more colorectal pathologies, including having one or more polyps or having one or more subtypes of polyps; not having colorectal pathologies; not having polyps; or not having one or more subtypes of polyps. The term control or control sample can also refer to the compilation of data derived from one or more samples of one or more subjects.

A “coding region” in reference to a DNA refers to DNA which encodes RNA.

A “coding region” in reference to RNA refers to RNA which encodes protein.

As used herein, the term “data” in relation to one or more biomarkers, or the term “biomarker data” generally refers to data reflective of the absolute and/or relative abundance (level) of a product of a biomarker in a sample. As used herein, the term “dataset” in relation to one or more biomarkers refers to a set of data representing levels of each of one or more

biomarker products of a panel of biomarkers in a reference population of subjects. A dataset can be used to generate a formula/classifier of the invention. According to one embodiment the dataset need not comprise data for each biomarker product of the panel for each individual of the reference population. For example, the “dataset” when used in the context of a dataset to
5 be applied to a formula can refer to data representing levels of products of each biomarker for each individual in one or more reference populations, but as would be understood can also refer to data representing levels of products of each biomarker for 99%, 95%, 90%, 85%, 80%, 75%, 70% or less of the individuals in each of said one or more reference populations and can still be useful for purposes of applying to a formula.

10 As used herein, the term “derivative” in the context of proteinaceous agent (e.g., proteins, polypeptides, peptides, and antibodies) refers to a proteinaceous agent that comprises an amino acid sequence which has been altered by the introduction of amino acid residue substitutions, deletions, and/or additions. The term “derivative” as used herein also refers to a proteinaceous agent which has been modified, i.e., by the covalent attachment of any type of molecule to the
15 proteinaceous agent. For example, but not by way of limitation, an antibody may be modified, e.g., by glycosylation, acetylation, pegylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to a cellular ligand or other protein, etc. A derivative of a proteinaceous agent may be produced by chemical modifications using techniques known to those of skill in the art, including, but not limited to specific chemical
20 cleavage, acetylation, formylation, metabolic synthesis of tunicamycin, etc. Further, a derivative of a proteinaceous agent may contain one or more non-classical amino acids. A derivative of a proteinaceous agent possesses a similar or identical function as the proteinaceous agent from which it was derived.

As used herein, the term “derivative” in the context of a non-proteinaceous derivative
25 refers to a second organic or inorganic molecule that is formed based upon the structure of a first organic or inorganic molecule. A derivative of an organic molecule includes, but is not limited to, a molecule modified, e.g., by the addition or deletion of a hydroxyl, methyl, ethyl, carboxyl or amine group. An organic molecule may also be esterified, alkylated and/or phosphorylated.

As used herein the terms “testing”, “diagnosis” and “screening”, in relation to colorectal
30 pathologies refer to a process of determining a likelihood (probability) of a test subject having one or more colorectal pathologies and includes both traditional medical diagnostic techniques as

well as testing methods as encompassed by one or more aspects of the current invention. Traditional medical diagnostic techniques for testing for colorectal pathology includes physical exam and history, medical evaluation, and appropriate laboratory tests which can include FOBT, sigmoidoscopy and colonoscopy. In one embodiment, “diagnosis of colorectal pathology” refers to a determination as between two options: e.g., (i) that an individual has colorectal pathology or one or more subtypes of colorectal pathology, or one or more polyps and (ii) that an individual does not have the colorectal pathology or the one or more polyps or the one or more subtypes of polyps. In another embodiment diagnosis can include an option that it cannot be determined with sufficient degree of certainty as to whether an individual can be characterized as having colorectal pathology or not. In one context, a “sufficient degree of certainty” takes into account any limitations – such as limitations in the technology, equipment or measuring where as a result of the limitations, the result is within a range which suggests that the test is indeterminate. The range which suggests the test is indeterminate will depend upon the specific limitations of the equipment, reagents and technology used. In another context, “sufficient degree of certainty” depends upon the medical requirements for the sensitivity and/or specificity of the test. More particularly the sufficient degree of certainty includes greater than 50% sensitivity and/or specificity, greater than 60% sensitivity and/or specificity, greater than 70% sensitivity and/or specificity, greater than 80% sensitivity and/or specificity, greater than 90% sensitivity and/or specificity and 100% sensitivity and/or specificity.

As used herein, “normal” refers to an individual or group of individuals who do not have colorectal pathology. In some embodiments, the diagnosis of said individual or group of individuals not having colorectal pathology is determined using conventional diagnostic methods. In some embodiments said individual or group of individuals have not been diagnosed with any other disease. “Normal”, according to the invention, also refers to samples isolated from normal individuals and includes blood, total RNA or mRNA isolated from normal individuals. A sample taken from a normal individual can include a sample taken from an individual who does not have colorectal pathology at the time the sample is taken.

As used herein, the term “differential expression” refers to a difference in the level of expression of the products of one or more biomarkers. For instance, the term “differential expression” can refer to the difference in the level of RNA of one or more biomarkers between samples from subjects having and subjects not having one or more colorectal pathologies.

Differences in biomarker RNA product levels can be determined by directly or indirectly measuring the amount or level of Differentially expressed” can also include different levels of protein encoded by the biomarker of the invention between samples or reference populations. “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 one or more colorectal pathologies, wherein the ratio is not equal to 1.0. 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 of the invention a biomarker product is 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 of the invention a biomarker product 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.

“Differentially increased expression” or “up regulation” refers to biomarker product levels which are at least 10% or more, for example, 20%, 30%, 40%, or 50%, 60%, 70%, 80%, 90% higher or more, and/or 1.1 fold, 1.2 fold, 1.4 fold, 1.6 fold, 1.8 fold higher or more, than a control.

“Differentially decreased expression” or “down regulation” refers to biomarker product levels which are at least 10% or more, for example, 20%, 30%, 40%, or 50%, 60%, 70%, 80%,

90% lower or less, and/or 0.9 fold, 0.8 fold, 0.6 fold, 0.4 fold, 0.2 fold, 0.1 fold or less lower than a control.

For example, up regulated or down regulated genes include genes having an increased or decreased level, respectively, of expression of product (e.g. mRNA or protein) in blood isolated
5 from individuals characterized as having one or more colorectal pathologies as compared with normal individuals. In another example, up regulated or down regulated genes include genes having an increased or decreased level, respectively, of expression of product (e.g. mRNA or protein) in blood isolated from individuals having one type of colorectal pathology or collection of colorectal pathologies as compared with individuals having a different type of colorectal
10 pathology or collection of colorectal pathologies, respectively.

For example, up regulated genes include genes having an increased level of biomarker products in a test sample as compared with a control sample.

As used herein, the term “differential hybridization” refers to a difference in a quantitative level of hybridization of a nucleic acid or derivative thereof isolated and/or derived from a
15 sample from a first individual or individuals with a trait to a complementary nucleic acid target as compared with the hybridization of a nucleic acid or derivative thereof isolated and/or derived from a second individual or individuals not having said trait to a complementary nucleic acid target. A “differential hybridization” means that the ratio of the level of hybridization of the first sample as compared with the second sample is not equal to 1.0. For example, the ratio of the
20 level of hybridization of the first sample to the target as compared to the second sample is greater than 1.1, 1.2, 1.5, 1.7, 2, 3, 4, 10, 20, or less than 1, for example 0.9, 0.8, 0.6, 0.4, 0.2, 0.1, 0.05. A differential hybridization also exists if the hybridization is detectable in one sample but not another sample.

As used herein, the term “drug efficacy” refers to the effectiveness of a drug. “Drug
25 efficacy” is usually measured by the clinical response of the patient who has been or is being treated with a drug. A drug is considered to have a high degree of efficacy, if it achieves desired clinical results, for example, the alteration of gene expression and the gene expression pattern reflective of one or more colorectal pathologies as described herein. The amount of drug absorbed may be used to predict a patient's response. A general rule is that as the dose of a drug
30 is increased, a greater effect is seen in the patient until a maximum desired effect is reached. If

more drug is administered after the maximum point is reached, the side effects will normally increase.

As used herein, the term “effective amount” refers to the amount of a compound which is sufficient to reduce or prevent the progression and/or severity of one or more colorectal pathologies; prevent the development, recurrence of onset of one or more colorectal pathologies; or enhance or improve the prophylactic or therapeutic effect(s) of another therapy.

As used herein, the term “fragment” in the context of a proteinaceous agent refers to a peptide or polypeptide comprising an amino acid sequence of at least 5 contiguous amino acid residues, at least 10 contiguous amino acid residues, at least 15 contiguous amino acid residues, at least 20 contiguous amino acid residues, at least 25 contiguous amino acid residues, at least 40 contiguous amino acid residues, at least 50 contiguous amino acid residues, at least 60 contiguous amino acid residues, at least 70 contiguous amino acid residues, at least contiguous 80 amino acid residues, at least contiguous 90 amino acid residues, at least contiguous 100 amino acid residues, at least contiguous 125 amino acid residues, at least 150 contiguous amino acid residues, at least contiguous 175 amino acid residues, at least contiguous 200 amino acid residues, or at least contiguous 250 amino acid residues of the amino acid sequence of another polypeptide or a protein. In a specific embodiment, a fragment of a protein or polypeptide retains at least one function of the protein or polypeptide. In another embodiment, a fragment of a protein or polypeptide retains at least two, three, four, or five functions of the protein or polypeptide. In some embodiments, a fragment of an antibody retains the ability to immunospecifically bind to an antigen.

As used herein, the term “fusion protein” refers to a polypeptide that comprises an amino acid sequence of a first protein or polypeptide or functional fragment, analog or derivative thereof, and an amino acid sequence of a heterologous protein, polypeptide, or peptide (i.e., a second protein or polypeptide or fragment, analog or derivative thereof different than the first protein or fragment, analog or derivative thereof). In one embodiment, a fusion protein comprises a prophylactic or therapeutic agent fused to a heterologous protein, polypeptide or peptide. In accordance with this embodiment, the heterologous protein, polypeptide or peptide may or may not be a different type of prophylactic or therapeutic agent.

As used herein, a “gene” of the invention can include a gene expressed in blood, a gene expressed in blood and in a non-blood tissue, a gene differentially expressed in blood, a gene

expressed in a non-blood cell, a gene expressed in a cell which is not of haematopoietic origin, a gene expressed in a specific subtype of cell found in blood including lymphocytes, granulocytes, leukocytes, basophils and the like. A gene can be an immune response gene or a gene not involved in an immune response. In particular an immune response gene is a gene in the major
5 histocompatibility complex that controls a cells response to a foreign antigen. A gene of the invention can also include a gene which is differentially regulated in response to a foreign antigen introduced into peripheral blood.

As used herein, a “gene expression pattern” or “gene expression profile” indicates the pattern of the level of expression of two or more biomarkers of the invention including 2, 3, 4, 5,
10 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 or more or all of the biomarkers of the invention. A gene expression pattern or gene expression profile can be determined from the measurement of expression levels of the products of the biomarkers of the invention using any known technique. For example techniques to measure expression of the RNA products of the biomarkers of the invention include, PCR based methods (including reverse transcription-PCR, PCR, QRT-PCR)
15 and non PCR based method, as well as microarray analysis. To measure levels of protein products of the biomarkers of the invention, techniques include densitometric western blotting and ELISA analysis.

As used herein, the term “hybridizing to” or “hybridization” refers to the sequence specific non-covalent binding interactions with a complementary nucleic acid, for example
20 interactions between a target nucleic acid sequence and a nucleic acid member on an array.

As used herein, the term “immunoglobulin” refers to a protein consisting of one or more polypeptides substantially encoded by immunoglobulin genes. The recognized human immunoglobulin genes include the kappa, lambda, alpha (IgA1 and IgA2), gamma (IgG1, IgG2, IgG3, IgG4), delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin
25 variable region genes. Full-length immunoglobulin “light chains” (about 25 Kd or 214 amino acids) are encoded by a variable region gene at the NH₂-terminus (about 110 amino acids) and a kappa or lambda constant region gene at the COOH-terminus. Full-length immunoglobulin “heavy chains” (about 50 Kd or 446 amino acids), are similarly encoded by a variable region gene (about 116 amino acids) and one of the other aforementioned constant region genes, *e.g.*,
30 gamma (encoding about 330 amino acids).

As used herein, the term “in combination” when referring to therapeutic treatments refers to the use of more than one type of therapy (*e.g.*, more than one prophylactic agent and/or therapeutic agent). The use of the term “in combination” does not restrict the order in which therapies (*e.g.*, prophylactic and/or therapeutic agents) are administered to a subject. A first therapy (*e.g.*, a first prophylactic or therapeutic agent) can be administered prior to (*e.g.*, 5 minutes, 15 minutes, 30 minutes, 45 minutes, 1 hour, 2 hours, 4 hours, 6 hours, 12 hours, 24 hours, 48 hours, 72 hours, 96 hours, 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 8 weeks, or 12 weeks before), concomitantly with, or subsequent to (*e.g.*, 5 minutes, 15 minutes, 30 minutes, 45 minutes, 1 hour, 2 hours, 4 hours, 6 hours, 12 hours, 24 hours, 48 hours, 72 hours, 96 hours, 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 8 weeks, or 12 weeks after) the administration of a second therapy (*e.g.*, a second prophylactic or therapeutic agent) to a subject.

As used herein, “indicative of one or more colorectal pathologies” refers to a determination of a probability that a subject has or will have the one or more colorectal pathologies. In one aspect the application of a formula to data corresponding to biomarker products of a test subject can result in determination of the probability of whether the test subject has one or more colorectal pathologies as compared with not having said one or more colorectal pathologies. In another embodiment, an expression pattern can be indicative of one or more colorectal pathologies including one or more polyps or one or more subtypes of polyps if the expression pattern is found significantly more often in patients with said colorectal pathology than in patients without said colorectal pathology (as determined using routine statistical methods setting confidence levels at a minimum of 70%, 75%, 80%, 85%, 90%, 95% and the like). In some embodiments, an expression pattern which is indicative of disease is found in at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or more in patients who have the disease and is found in less than 10%, less than 8%, less than 5%, less than 2.5%, or less than 1% of patients who do not have the disease. “Indicative of colorectal pathology” can also indicate an expression pattern which more properly categorizes with control expression patterns of individuals with the one or more colorectal pathologies as compared with control expression patterns of individuals without the one or more colorectal pathologies using statistical algorithms for class prediction as would be understood by a person skilled in the art and see for

example commercially available programs such as those provided by Silicon Genetics (e.g. GeneSpring™).

As used herein, “isolated” or “purified” when used in reference to a nucleic acid means that a naturally occurring sequence has been removed from its normal cellular (e.g.,
5 chromosomal) environment or is synthesized in a non-natural environment (e.g., artificially synthesized). Thus, an “isolated” or “purified” sequence may be in a cell-free solution or placed in a different cellular environment. The term “purified” does not imply that the sequence is the only nucleotide present, but that it is essentially free (about 90-95% pure) of non-nucleotide material naturally associated with it, and thus is distinguished from isolated chromosomes.

10 As used herein, the terms “isolated” and “purified” in the context of a proteinaceous agent (e.g., a peptide, polypeptide, protein or antibody) refer to a proteinaceous agent which is substantially free of cellular material and in some embodiments, substantially free of heterologous proteinaceous agents (i.e., contaminating proteins) from the cell or tissue source from which it is derived, or substantially free of chemical precursors or other chemicals when
15 chemically synthesized. The language “substantially free of cellular material” includes preparations of a proteinaceous agent in which the proteinaceous agent is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, a proteinaceous agent that is substantially free of cellular material includes preparations of a proteinaceous agent having less than about 40%, 30%, 20%, 10%, or 5% (by dry weight) of
20 heterologous proteinaceous agent (e.g., protein, polypeptide, peptide, or antibody; also referred to as a “contaminating protein”). When the proteinaceous agent is recombinantly produced, it is also in some embodiments substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When the proteinaceous agent is produced by chemical synthesis, it is in some embodiments substantially
25 free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the proteinaceous agent. Accordingly, such preparations of a proteinaceous agent have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the proteinaceous agent of interest. In some embodiments, proteinaceous agents disclosed herein are isolated.

30 As used herein, a sample which is “isolated and/or derived” includes a sample which has been removed it from its natural environment in a subject and also includes samples which are

further modified or altered. For example samples can include tissue, lymph, bodily fluid, blood, RNA, protein, mRNA, serum reduced blood, erythrocyte reduced blood, serum reduced and erythrocyte reduced blood, unfractionated cells of a lysed blood, globin reduced mRNA, cDNA, PCR products and the like.

5 As used herein, the term “level” or “level of expression” when referring to RNA refers to a measurable quantity (either absolute or relative quantity) of a given nucleic acid as determined by hybridization or measurements such as QRT-PCR and includes use of both SYBR® green and TaqMan® technology and which corresponds in direct proportion with the amount of product of the gene in a sample. Level of expression when referring to RNA can also refer to a
10 measurable quantity of a given nucleic acid as determined by PCR wherein the number of cycles of PCR is limited to 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, or 60 cycles. The level of expression when referring to RNA can also refer to a measurable quantity of a given nucleic acid as determined relative to the amount of total RNA, or cDNA used in QRT-PCR wherein the amount of total RNA used is 100ng; 50ng, 25ng; 10ng; 5ng; 1.25ng; 0.05ng; 0.3ng; 0.1ng; 0.09 ng;
15 0.08ng; 0.07ng; 0.06ng; or 0.05ng. The level of expression of a nucleic acid can be determined by any methods known in the art. For microarray analysis, the level of expression is measured by hybridization analysis using nucleic acids corresponding to RNA isolated from one or more individuals according to methods well known in the art. The label can either be incorporated into the RNA or used in another manner as would be understood so as to monitor hybridization.
20 The label used can be a luminescent label, an enzymatic label, a radioactive label, a chemical label or a physical label. In some embodiments, target and/or probe nucleic acids are labeled with a fluorescent molecule. Preferred fluorescent labels include, but are not limited to: fluorescein, amino coumarin acetic acid, tetramethylrhodamine isothiocyanate (TRITC), Texas Red, Cyanine 3 (Cy3) and Cyanine 5 (Cy5). The level of expression when referring to RNA can
25 also refer to a measurable quantity of a given nucleic acid as determined relative to the amount of total RNA or cDNA used in microarray hybridizations wherein the amount of total RNA is 10 µg, 5 µg, 2.5 µg; 2 µg; 1 µg; 0.5 µg; 0.1 µg; 0.05 µg; 0.01 µg; 0.005 µg; 0.001 µg and the like.

As used herein, a “ligand” is a molecule that binds to another. “Polynucleotide Ligands” are those that specifically and/or selectively hybridize to products of the biomarkers and/or
30 derivatives thereof. Polynucleotide ligands can specifically and/or selectively hybridize to RNA and/or protein products of the biomarkers, allowing measurement of the levels of the biomarker

products. The polynucleotide ligands may be any of various types of molecule, including but not limited to, any of various combinations of oligonucleotides, cDNA, DNA, RNA, PCR products, synthetic DNA, synthetic RNA, and/or modified nucleotides.

A ligand of the invention can also include a "polypeptide ligand" that specifically or
5 selectively binds to the biomarker products, for example, allowing detection or measurement of the level of biomarker products including either RNA products and/or protein products. A polypeptide ligand may include a scaffold peptide, a linear peptide, or a cyclic peptide. In a preferred embodiment, the polypeptide ligand is an antibody. The antibody can be a human
10 antibody, a chimeric antibody, a recombinant antibody, a humanized antibody, a monoclonal antibody, or a polyclonal antibody. The antibody can be an intact immunoglobulin, *e.g.*, an IgA, IgG, IgE, IgD, IgM or subtypes thereof. The antibody can be conjugated to a functional moiety (*e.g.*, a compound which has a biological or chemical function (which may be a second different polypeptide, a therapeutic drug, a cytotoxic agent, a detectable moiety, or a support. A polypeptide ligand *e.g.* antibody of the invention interacts with a polypeptide, encoded by one of
15 the genes of a biomarker, with high affinity and specificity. For example, the polypeptide ligand binds to a polypeptide, encoded by one of the genes of a biomarker, with an affinity constant of at least 10^7 M^{-1} , preferably, at least 10^8 M^{-1} , 10^9 M^{-1} , or 10^{10} M^{-1} . The polynucleotide ligands and protein ligands may be used, according to standard art knowledge, to practice techniques such as Western blotting, immunoprecipitation, enzyme-linked immunosorbent assay (ELISA),
20 protein microarray analysis and the like to measure the level of disclosed biomarker protein products.

An "mRNA" means an RNA complementary to a gene; an mRNA includes a protein coding region, and also may include 5' end and 3' untranslated regions (UTR).

As used herein, the term "majority" refers to a number representing more than 50% (*e.g.*,
25 51%, 60%, or 70%, or 80% or 90% or up to 100%) of the total members of a composition. The term "majority", when referring to an array, it means more than 50% (*e.g.*, 51%, 60%, or 70%, or 80% or 90% or up to 100%) of the total nucleic acid members that are stably associated with the solid substrate of the array.

Treatment of one or more colorectal pathologies or one or more subtypes of colorectal
30 pathology is defined herein to provide medical aid to counteract the disease itself, the symptoms and/or the progression of the disease. Treatments also include removing the one or more

colorectal pathologies and include palliative therapy to help relieve symptoms and improve quality of life. Treatments also include reducing or preventing polyp formation, reducing or preventing polyp differentiation or morphology changes, and can also include development, recurrence and onset.

5 As used herein, "mRNA integrity" refers to the quality of mRNA extracts from either tissue samples or samples. In one embodiment, mRNA extracts with good integrity do not appear to be degraded when examined by methods well known in the art, for example, by RNA agarose gel electrophoresis (e.g., Ausubel et al., John Wiley & Sons, Inc., 1997, *Current Protocols in Molecular Biology*). Preferably, the mRNA samples have good integrity (e.g., less
10 than 10%, in some embodiments less than 5%, and more in some embodiments less than 1% of the mRNA is degraded) to truly represent the gene expression levels of sample from which they are extracted.

As used herein, "nucleic acid(s)" and "nucleic acid molecule(s)" are interchangeable with the term "polynucleotide(s)" and it generally refers to any polyribonucleotide or poly-
15 deoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA or any combination thereof. "Nucleic acids" include, without limitation, single- and double-stranded nucleic acids. As used herein, the term "nucleic acid(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "nucleic acids". The term "nucleic acids" as it is
20 used herein embraces such chemically, enzymatically or metabolically modified forms of nucleic acids, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including for example, simple and complex cells. A "nucleic acid" or "nucleic acid sequence" may also include regions of single- or double-stranded RNA or DNA or any combinations thereof and can include expressed sequence tags (ESTs) according to some embodiments of the
25 invention. An EST is a portion of the expressed sequence of a gene (i.e., the "tag" of a sequence), made by reverse transcribing a region of mRNA so as to make cDNA.

As defined herein, a "nucleic acid array" refers a plurality of nucleic acids (or "nucleic acid members") localized on a support where each of the nucleic acid members is localized to a unique pre-selected region of a support. In one embodiment, a nucleic acid member is attached
30 to the surface of the support and the nucleic acid member is DNA. In another embodiment, the nucleic acid member is either cDNA or and oligonucleotide. In another embodiment, the nucleic

acid member localized on the support is cDNA synthesized by polymerase chain reaction (PCR). The term “nucleic acid”, as used herein, is interchangeable with the term “polynucleotide”. In another preferred embodiment, a “nucleic acid array” refers to a plurality of unique nucleic acids attached to nitrocellulose or other membranes used in Southern and/or Northern blotting techniques.

As used herein “nucleic acid sample for hybridization to an array” is defined as a nucleic acid isolated and/or derived from a sample capable of binding to a nucleic acid bound to an array of complementary sequence through sets of non-covalent bonding interactions including complementary base pairing interactions. The nucleic acid sample for hybridization to an array can either be an isolated nucleic acid sequence corresponding to a gene or portion thereof, total RNA or mRNA isolated from a sample. In one embodiment, the nucleic acid sample for hybridization to an array is a blood nucleic acid sample (including whole blood, lysed blood, serum reduced, erythrocyte reduced blood, or peripheral blood leukocytes (PBLs)). In some embodiments, the nucleic acid sample is single- or double-stranded DNA, RNA, or DNA-RNA hybrids, from human blood and in some embodiments from RNA or mRNA extracts.

As used herein, a “nucleic acid member on an array” or a “nucleic acid member” includes nucleic acid immobilized on an array and capable of binding to a nucleic acid probes or samples of complementary sequence through sets of non-covalent bonding interactions, including complementary base pairing interactions. As used herein, a nucleic acid member or target may include natural (i.e., A, G, C, or T) or modified bases (7-deazaguanosine, inosine, etc.). In addition, the bases in nucleic acids may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization (i.e., the nucleic acid target still specifically binds to its complementary sequence under standard stringent or selective hybridization conditions). Thus, nucleic acid members may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. In one embodiment, a conventional nucleic acid array of ‘target’ sequences bound to the array can be representative of the entire human genome, e.g. Affymetrix chip, and the biomarker or isolated biomarker consisting of or comprising one or more of the genes set out in **Table 1**, **Table 2**, or **Table 11**, or **Table 12** or gene probes (e.g. **Table 4**) is applied to the conventional array. In another embodiment, sequences bound to the array can be the biomarker or isolated biomarker according to the invention and total cellular RNA is applied to the array.

As used herein, the term "oligonucleotide" is defined as a molecule comprised of two or more deoxyribonucleotides and/ or ribonucleotides, and preferably more than three. Its exact size will depend upon many factors which, in turn, depend upon the ultimate function and use of the oligonucleotide. The oligonucleotides may be from about 8 to about 1,000 nucleotides long.

5 Although oligonucleotides of 8 to 100 nucleotides are useful in the invention, preferred oligonucleotides range from about 8 to about 15 bases in length, from about 8 to about 20 bases in length, from about 8 to about 25 bases in length, from about 8 to about 30 bases in length, from about 8 to about 40 bases in length or from about 8 to about 50 bases in length.

10 As used herein, "patient" or "individual" or "subject" refers to a mammal and is in some embodiments human.

As used herein the term "peptide" refers to a polypeptide which is 50 amino acids in length or less.

As used herein, the phrase "pharmaceutically acceptable salt(s)" includes, but is not limited to, salts of acidic or basic groups that may be present in compounds identified using the
15 methods of the present invention. Compounds that are basic in nature are capable of forming a wide variety of salts with various inorganic and organic acids. The acids that can be used to prepare pharmaceutically acceptable acid addition salts of such basic compounds are those that form non-toxic acid addition salts, i.e., salts containing pharmacologically acceptable anions, including but not limited to sulfuric, citric, maleic, acetic, oxalic, hydrochloride, hydrobromide,
20 hydroiodide, nitrate, sulfate, bisulfate, phosphate, acid phosphate, isonicotinate, acetate, lactate, salicylate, citrate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucaronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate and pamoate (i.e., 1,1'-methylene-bis-(2-hydroxy-3-naphthoate)) salts. Compounds that include an amino
25 moiety may form pharmaceutically acceptable salts with various amino acids, in addition to the acids mentioned above. Compounds that are acidic in nature are capable of forming base salts with various pharmacologically acceptable cations. Examples of such salts include alkali metal or alkaline earth metal salts and, particularly, calcium, magnesium, sodium lithium, zinc, potassium, and iron salts.

30 As used herein, "polynucleotide" encompasses single and double stranded polynucleotides, such as double-stranded DNA, single-stranded DNA, double-stranded RNA,

single-stranded RNA or DNA-RNA double stranded hybrids, and the like, of more than 8 nucleotides in length. The term “polynucleotide” includes a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides, that comprise purine and pyrimidine bases, or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. The backbone of the polynucleotide can comprise sugars and phosphate groups, as may typically be found in RNA or DNA, or modified or substituted sugar or phosphate groups. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. The sequence of nucleotides may be interrupted by non-nucleotide components.

10 As used herein a “polynucleotide ligand that specifically and/or selectively hybridizes to RNA products of the biomarkers” (“biomarker RNA products”) and/or to polynucleotides corresponding to biomarker RNA products, allowing measurement of levels of the RNA products are disclosed.

15 The polynucleotide ligands may be any of various types of molecule, including but not limited to, any of various combinations of oligonucleotides, cDNA, DNA, RNA, PCR products, synthetic DNA, synthetic RNA, and/or modified nucleotides.

As used herein, the term “proteinaceous agent” refers to polypeptides, proteins, peptides, and the like.

As used herein, “polypeptide sequences encoded by or protein products of” refers to the amino acid sequences obtained after translation of the protein coding region of an mRNA transcribed from a gene. As would be understood, one or more mRNA nucleotide sequences for each of the genes (biomarkers) of the invention can be identified using public databases such as the NCBI database found at <http://www.ncbi.nlm.nih.gov>. For example, representative mRNA species of those biomarkers identified in **Table 2** and **Table 12** are provided by their Human Genbank Accession number (see **Table 3** and **Table 13** respectively) and the corresponding polypeptide sequence is identified by a Protein Accession number (see **Table 3** and **Table 13** respectively). These Genbank Accession numbers provide the sequence of products of the biomarkers. When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as epitopes or antigenic determinants. As used herein, “antigenic fragments” refers

portions of a polypeptide that contains one or more epitopes. Epitopes can be linear, comprising essentially a linear sequence from the antigen, or conformational, comprising sequences which are genetically separated by other sequences but come together structurally at the binding site for the polypeptide ligand. "Antigenic fragments" may be 5000, 1000, 500, 400, 300, 200, 100, 50
5 or 25 or 20 or 10 or 5 amino acids in length.

As used herein, the terms "prevent", "preventing" and "prevention" refer to the prevention of the development, recurrence or formation or growth or transformation of colorectal pathology including polyps or subtypes of polyps resulting from the administration of one or more compounds identified in accordance the methods of the invention or the administration of a
10 combination of such a compound and another therapy.

The term, "primer", as used herein refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product, which is complementary to a nucleic acid strand, is induced, i.e., in
15 the presence of nucleotides and an inducing agent such as a DNA polymerase and at a suitable temperature and pH. The primer may be either single-stranded or double-stranded and must be sufficiently long to prime the synthesis of the desired extension product in the presence of the inducing agent. The exact length of the primer will depend upon many factors, including temperature, source of primer and the method used and also the specificity or selectivity of the
20 desired priming (ie so as to act as a point of initiation of synthesis which is specific or selective for a given sequence of polynucleotide). For example, for testing applications, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15-25, but may contain additional nucleotides as well as fewer nucleotides. In addition, in some cases the primer can be selected so as to have a high GC content, can be selected so as to bind to
25 regions which do not contain SNPs, can be selected so as to span an intron/ exon junction of RNA and the like. Other factors involved in determining the appropriate length and/or features of a primer are readily known to one of ordinary skill in the art.

The term "biomarker specific set of primers" or "primer sets" as used herein refers to a set of polynucleotide primers wherein one primer primes the synthesis of a sense strand, and
30 the second primer primes the synthesis of an antisense strand so as to produce double stranded DNA complementary to a portion of one or more RNA products of the biomarker of the

invention. For example, the primers can include a first primer which is a sequence that can selectively hybridize to RNA, cDNA or EST complementary to a region of the biomarker of the invention to create an extension product and a second primer capable of selectively hybridizing to the extension product, which are used to produce double stranded DNA
5 complementary to a region of the biomarker of the invention or products of the biomarker of the invention. The invention includes primers useful for measuring the level of RNA products of a biomarker. **Table 4, Table 6, Table 14, Table 16 and Table 17** provide representative species of primers of the invention. A biomarker specific set of primers can be selected so that they will selectively amplify only portions of a polynucleotide complementary to one or more
10 RNA products of a biomarker and do not amplify portions of polynucleotides complementary to other biomarkers.

As used herein, the term "probe" means oligonucleotides and analogs thereof and refers to a range of chemical species that recognize polynucleotide target sequences through hydrogen bonding interactions with the nucleotide bases of the target sequences. The probe or
15 the target sequences may be single- or double-stranded RNA or single- or double-stranded DNA or a combination of DNA and RNA bases. A probe is at least 8 nucleotides in length and less than the length of a complete gene. A probe may be 10, 20, 30, 50, 75, 100, 150, 200, 250, 400, 500 and up to 2000 nucleotides in length as long as it is less than the full length of the target gene. In some embodiments, probes can be used as target sequences bound on a
20 microarray. In some embodiments, probes can be used for quantitative real-time PCR (QRT-PCR) and include modifications so as to incorporate a fluorophore, a quencher, a minor groove binding reagent or other substances which allow detection of the probe during PCR amplification. The probe can also be modified so as to have both a detectable tag and a quencher molecule, for example Taqman® and Molecular Beacon® probes. The invention
25 includes probes useful for measuring the expression of RNA products of biomarkers of the invention. For example, **Table 4, Table 6, and Table 14 and Table 17** provides some representative species of a probe of the invention useful for QRT-PCR.

The oligonucleotides and analogs thereof may be RNA or DNA, or analogs of RNA or DNA, commonly referred to as antisense oligomers or antisense oligonucleotides. Such RNA
30 or DNA analogs comprise but are not limited to 2'-O-alkyl sugar modifications, methylphosphonate, phosphorothiate, phosphorodithioate, formacetal, 3'-thioformacetal,

sulfone, sulfamate, and nitroxide backbone modifications, and analogs wherein the base moieties have been modified. In addition, analogs of oligomers may be polymers in which the sugar moiety has been modified or replaced by another suitable moiety, resulting in polymers which include, but are not limited to, morpholino analogs and peptide nucleic acid (PNA) analogs (Egholm, et al. Peptide Nucleic Acids (PNA)--Oligonucleotide Analogues with an Achiral Peptide Backbone, (1992)).

Probes may also be mixtures of any of the oligonucleotide analog types together or in combination with native DNA or RNA and may also include linker species. At the same time, the oligonucleotides and analogs thereof may be used alone or in combination with one or more additional oligonucleotides or analogs thereof.

As used herein, the term "products of the biomarker" or "biomarker products" refers to a species of RNA or a species of protein (wherein a species of RNA or protein can include multiple copies) isolated and/or derived from a sample including a tissue sample, a lymph sample, a lymph tissue sample, or a blood sample, or a fraction of a blood sample which corresponds to the biomarker (i.e., is transcribed from the gene or genetic element or is translated from RNA which is transcribed from the gene or genetic element). See **Table 3** and **Table 13**. The RNA may be pre-mRNA, mRNA, spliced variants of mRNA and the like. The protein may be in its native state or post-translationally processed in any one of various ways.

As used herein, "a plurality of" or "a set of" refers to two or more, for example, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more 10 or more etc.

As used herein, "pre-selected region", "predefined region", or "unique position" refers to a localized area on a substrate which is, was, or is intended to be used for the deposit of a nucleic acid and is otherwise referred to herein in the alternative as a "selected region" or simply a "region." The pre-selected region may have any convenient shape, e.g., circular, rectangular, elliptical, wedge-shaped, etc. In some embodiments, a pre-selected region is smaller than about 1 cm², more preferably less than 1 mm², still more preferably less than 0.5 mm², and in some embodiments less than 0.1 mm². A nucleic acid member at a "pre-selected region", "predefined region", or "unique position" is one whose identity (e.g., sequence) can be determined by virtue of its position at the region or unique position.

As used herein, the terms “prophylactic agent” and “prophylactic agents” refer to any compound(s) which can be used to prevent polyp formation, development, recurrence or onset. In certain embodiments, the term “prophylactic agent” refers to a compound identified in the screening assays described herein. In certain other embodiments, the term “prophylactic agent” refers to an agent other than a compound identified in the screening assays described herein which is known to be useful for, or has been or is currently being used to prevent or impede the onset, development and/or progression or transformation of one or more colorectal pathologies including one or more polyps or subtypes of polyps.

As used herein, the phrase “prophylactically effective amount” refers to the amount of a therapy (e.g., a prophylactic agent) which is sufficient to result in the prevention of the development, recurrence or onset or progression or transformation of one or more colorectal pathologies including one or more polyps or subtypes of polyps; the reduction or amelioration of the progression and/or severity of one or more colorectal pathologies including one or more polyps or subtypes of polyps; or the prevention of colorectal pathology including polyps or subtypes of polyps advancing to colorectal cancer.

As used herein, the terms “protein” and “polypeptide” are used interchangeably to refer to a chain of amino acids linked together by peptide bonds. In a specific embodiment, a protein is composed of less than 200, less than 175, less than 150, less than 125, less than 100, less than 50, less than 45, less than 40, less than 35, less than 30, less than 25, less than 20, less than 15, less than 10, or less than 5 amino acids linked together by peptide bonds. In another embodiment, a protein is composed of at least 200, at least 250, at least 300, at least 350, at least 400, at least 450, at least 500 or more amino acids linked together by peptide bonds.

A “protein coding region” refers to the portion of the mRNA encoding a polypeptide.

As used herein the “reference population” or “test population” refers to one or more populations of “control samples” used to develop one or more classifier. In one embodiment a single reference population can be divided into subpopulations. In another embodiment, two or more reference populations can be used. In some instances a classifier can be developed to differentiate between individuals with one or more colorectal pathologies or one or more polyps or one or more subtypes of polyps and individuals without the same colorectal pathology or one or more polyps or one or more subtypes of polyps. In some instances a first reference population would be comprised of individuals with the one or more colorectal pathologies and a

second reference population would be comprised of individuals without the one or more colorectal pathologies. The “reference population” or “test population” can be comprised of control samples from a number of individuals diagnosed with one or more colorectal pathologies including one or more polyps or one or more subtypes of polyps and individuals not having the colorectal pathologies or not having the one or more polyps or not having the one or more subtypes of polyps as determined using conventional diagnostic techniques. Note that in some embodiments the population of individuals having one or more colorectal pathologies can be selected to include individuals having a single subtype of polyp or one or more subtypes of polyps. In other embodiments the individuals who do not have one or more colon pathologies can include individuals who have been diagnosed with other disease or diseases. In another embodiment the individuals who do not have one or more colon pathologies can include individuals who have been diagnosed with other cancers. In one embodiment the “reference population” or “test population” is comprised of a roughly equivalent number of “control samples” from each trait subgroup (e.g., in this instance wherein said trait is a determination of status with regards to the presence of colorectal pathology). In another embodiment, each trait subgroup (e.g. having or not having colorectal pathology) of the “reference population” has a similar distribution with regards to other traits e.g., age, sex, drug status, etc.

As used herein, the term “selectively binds” in the context of proteins encompassed by the invention refers to the specific interaction of any two of a peptide, a protein, a polypeptide an antibody, wherein the interaction preferentially occurs as between any two of a peptide, protein, polypeptide and antibody preferentially as compared with any other peptide, protein, polypeptide and antibody. For example, when the two molecules are protein molecules, a structure on the first molecule recognizes and binds to a structure on the second molecule, rather than to other proteins. “Selective binding”, “Selective binding”, as the term is used herein, means that a molecule binds its specific binding partner with at least 2-fold greater affinity, and preferably at least 10-fold, 20-fold, 50-fold, 100-fold or higher affinity than it binds a non-specific molecule.

As used herein “selective hybridization” can refer to a hybridization which occurs as between a polynucleotide and an RNA or protein product of the biomarker of the invention wherein the hybridization is such that the polynucleotide binds to the RNA products of the biomarker of the invention preferentially to the RNA products of other genes in the genome in question. In a preferred embodiment a polynucleotide which “selectively hybridizes” is one

which hybridizes with a selectivity of greater than 70%, greater than 80%, greater than 90% and most preferably on 100% (ie cross hybridization with other RNA species preferably occurs at less than 30%, less than 20%, less than 10%). As would be understood to a person skilled in the art, a polynucleotide which "selectively hybridizes" to the RNA products of a biomarker of the invention can be determined by taking into account the length and composition.

As used herein, "specifically hybridizes", "specific hybridization" can refer to hybridization which occurs when two nucleic acid sequences are substantially complementary (at least about 65% complementary over a stretch of at least 14 to 25 nucleotides, preferably at least about 75% complementary, more preferably at least about 90% complementary). See Kanehisa, M., 1984, Nucleic acids Res., 12:203, incorporated herein by reference. As a result, it is expected that a certain degree of mismatch is tolerated. Such mismatch may be small, such as a mono-, di- or tri-nucleotide. Alternatively, a region of mismatch can encompass loops, which are defined as regions in which there exists a mismatch in an uninterrupted series of four or more nucleotides. Numerous factors influence the efficiency and selectivity of hybridization of two nucleic acids, for example, the hybridization of a nucleic acid member on an array to a target nucleic acid sequence. These factors include nucleic acid member length, nucleotide sequence and/or composition, hybridization temperature, buffer composition and potential for steric hindrance in the region to which the nucleic acid member is required to hybridize. A positive correlation exists between the nucleic acid length and both the efficiency and accuracy with which a nucleic acid will anneal to a target sequence. In particular, longer sequences have a higher melting temperature (T_M) than do shorter ones, and are less likely to be repeated within a given target sequence, thereby minimizing non-specific hybridization. Hybridization temperature varies inversely with nucleic acid member annealing efficiency. Similarly the concentration of organic solvents, e.g., formamide, in a hybridization mixture varies inversely with annealing efficiency, while increases in salt concentration in the hybridization mixture facilitate annealing. Under stringent annealing conditions, longer nucleic acids, hybridize more efficiently than do shorter ones, which are sufficient under more permissive conditions.

As used herein, "spotting" or "attaching" refers to a process of depositing a nucleic acid member onto a solid substrate to form a nucleic acid array such that the nucleic acid is stably bound to the solid substrate via covalent bonds, hydrogen bonds or ionic interactions.

As used herein, “stably associated” refers to a nucleic acid that is stably bound to a solid substrate to form an array via covalent bonds, hydrogen bonds or ionic interactions such that the nucleic acid retains its unique pre-selected position relative to all other nucleic acids that are stably associated with an array, or to all other pre-selected regions on the solid substrate under
5 conditions in which an array is typically analyzed (i.e., during one or more steps of hybridization, washes, and/or scanning, etc.).

As used herein, “substrate” or “support” when referring to an array refers to a material capable of supporting or localizing an oligonucleotide or cDNA member. The support may be biological, non-biological, organic, inorganic, or a combination of any of these, existing as
10 particles, strands, precipitates, gels, sheets, tubing, spheres, beads, containers, capillaries, pads, slices, films, plates, slides, chips, etc. Often, the substrate is a silicon or glass surface, (poly)tetrafluoroethylene, (poly)vinylidene difluoride, polystyrene, polycarbonate, a charged membrane, such as nylon 66 or nitrocellulose, or combinations thereof. In one embodiment, the support is glass. In some embodiments, at least one surface of the substrate will be substantially
15 flat. In some embodiments, the support will contain reactive groups, including, but not limited to, carboxyl, amino, hydroxyl, thiol, and the like. In one embodiment, the support is optically transparent.

As herein used, the term “standard stringent conditions” means hybridization will occur only if there is at least 95% and preferably, at least 97% identity between the sequences, wherein
20 the region of identity comprises at least 10 nucleotides. In one embodiment, the sequences hybridize under stringent conditions following incubation of the sequences overnight at 42°C, followed by stringent washes (0.2X SSC at 65°C). The degree of stringency of washing can be varied by changing the temperature, pH, ionic strength, divalent cation concentration, volume and duration of the washing. For example, the stringency of hybridization may be varied by
25 conducting the hybridization at varying temperatures below the melting temperatures of the probes. The melting temperature of the probe may be calculated using the following formulas:

For oligonucleotide probes, between 14 and 70 nucleotides in length, the melting temperature (T_m) in degrees Celcius may be calculated using the formula: $T_m = 81.5 + 16.6(\log [Na^+]) + 0.41(\text{fraction G+C}) - (600/N)$ where N is the length of the oligonucleotide.

30 For example, the hybridization temperature may be decreased in increments of 5 °C from 68°C to 42 °C in a hybridization buffer having a Na^+ concentration of approximately 1M.

Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be “moderate stringency” conditions above 50 °C and “low stringency” conditions below 50 °C. A specific example of “moderate stringency” hybridization conditions is when the above hybridization is conducted at 55°C. A specific example of “low stringency” hybridization conditions is when the above hybridization is conducted at 45 °C.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation $T_m = 81.5 + 16.6(\log [Na^+]) + 0.41(\text{fraction G} + \text{C}) - (0.63\% \text{ formamide}) - (600/N)$, where N is the length of the probe.

For example, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42°C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50 °C. These conditions are considered to be “moderate stringency” conditions above 25% formamide and “low stringency” conditions below 25% formamide. A specific example of “moderate stringency” hybridization conditions is when the above hybridization is conducted at 30% formamide. A specific example of “low stringency” hybridization conditions is when the above hybridization is conducted at 10% formamide.

As used herein, the term “significant match”, when referring to nucleic acid sequences, means that two nucleic acid sequences exhibit at least 65% identity, at least 70%, at least 75%, at least 80%, at least 85%, and preferably, at least 90% identity, using comparison methods well known in the art (i.e., Altschul, S.F. et al., 1997, *Nucl. Acids Res.*, 25:3389-3402; Schäffer, A.A. et al., 1999, *Bioinformatics* 15:1000-1011). As used herein, “significant match” encompasses non-contiguous or scattered identical nucleotides so long as the sequences exhibit at least 65%, and preferably, at least 70%, at least 75%, at least 80%, at least 85%, and preferably, at least 90% identity, when maximally aligned using alignment methods routine in the art.

As used herein, the term “synergistic” refers to a combination of a compound identified using one of the methods described herein, and another therapy (e.g., agent), which is more effective than the additive effects of the therapies. In some embodiments, such other therapy has been or is currently being to prevent, treat, or ameliorate one or more colorectal pathologies including one or more polyps or one or more subtypes of polyps. A synergistic effect of a

combination of therapies (e.g., prophylactic or therapeutic agents) permits the use of lower dosages of one or more of the therapies and/or less frequent administration of said therapies to an individual with colorectal pathologies including polyps or subtypes of polyps. The ability to utilize lower dosages of a therapy (e.g., a prophylactic or therapeutic agent) and/or to administer said therapy less frequently reduces the toxicity associated with the administration of said agent to an individual without reducing the efficacy of said therapies in the prevention or treatment of colorectal pathology including polyps or subtypes of polyps. In addition, a synergistic effect can result in improved efficacy of therapies (e.g., agents) in the prevention or treatment of colorectal pathology including polyps or subtypes of polyps. Finally, a synergistic effect of a combination of therapies (e.g., prophylactic or therapeutic agents) may avoid or reduce adverse or unwanted side effects associated with the use of either therapy alone.

As used herein, a “therapeutic agent” or “agent” refers to a compound that increases or decreases the expression of a polynucleotide or polypeptide sequences that are differentially expressed in a sample from an individual having one or more colorectal pathologies including polyps or a subtype of polyps. The invention provides for a “therapeutic agent” that 1) prevents the formation of colorectal pathology 2) reduces, delays, or eliminates advancement or transformation of colorectal pathology and/or 3) restores one or more expression profiles of one or more colorectal pathology indicative nucleic acids or polypeptides of a patient to a profile more similar to that of a normal individual when administered to a patient. In addition, the terms “therapeutic agent” and “therapeutic agents” refer to any compound(s) which can be used in the treatment or prevention of colorectal pathology or polyps or a subtype of polyps. In certain embodiments, the term “therapeutic agent” refers to a compound identified in the screening assays described herein. In other embodiments, the term “therapeutic agent” refers to an agent other than a compound identified in the screening assays described herein which is known to be useful for, or has been or is currently being used to treat or prevent colorectal pathology or polyps or subtypes of polyps.

As used herein, the term “therapeutically effective amount” refers to that amount of a therapy (e.g., a therapeutic agent) sufficient to treat one or more colorectal pathologies including polyps or one or more subtypes of polyps; prevent one or more colorectal pathologies including polyps or one or more subtypes of polyps; prevent colorectal pathologies including polyps or one or more subtypes of polyps from transforming and/or advancing to colorectal cancer, cause

regression of colorectal pathology, polyps or one or more subtypes of polyps, or to enhance or improve the therapeutic effect(s) of another therapy (e.g., therapeutic agent). In a specific embodiment, a therapeutically effective amount refers to the amount of a therapy (e.g., a therapeutic agent) that modulates gene expression of the products of the biomarkers of the inventions. In some embodiments, a therapeutically effective amount of a therapy (e.g., a therapeutic agent) modulates gene expression of the products of the biomarkers of the invention at least 5%, preferably at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 100% relative to a control therapeutic agent such as phosphate buffered saline ("PBS").

As used herein, the terms "treat", "treatment" and "treating" refer to the prevention of one or more colorectal pathologies including polyp formation or the formation of one or more subtypes of polyps, development, recurrence onset or transformation of one or more colorectal pathologies and, the reduction or amelioration of the progression and/or severity of one or more colorectal pathologies including polyps or subtypes thereof resulting from the administration of one or more compounds identified in accordance the methods of the invention, or a combination of one or more compounds identified in accordance with the invention and another therapy.

As used herein, a "tissue nucleic acid sample", refers to nucleic acids isolated and/or derived from tissue, for example polyp tissue, colon tissue, rectum tissue, lymphoid tissue, and the like. In some embodiments, a tissue nucleic acid sample is total RNA, mRNA or is a nucleic acid corresponding to RNA, for example, cDNA. A tissue nucleic acid sample can also include a PCR product derived from total RNA, mRNA or cDNA.

(C) Samples for Use in the Invention

Samples for use in the invention include refers to any one of various type of molecules, cells and/or tissues which can be isolated and/or derived from a test subject and/or control subject, and which contains one or more biomarker products. The sample can be isolated and/or derived from any fluid, cell or tissue. The sample can also be one which is isolated and/or derived from any fluid and/or tissue which predominantly comprises blood cells.

The sample which is isolated and/or derived from an individual can be assayed for gene expression products, particularly genes expression products differentially expressed in individuals with or without one or more colorectal pathologies. In one embodiment, the sample is

a fluid sample, a lymph sample, a lymph tissue sample or a blood sample. In one embodiment the sample is isolated and/or derived from peripheral blood. Alternately, the sample may be isolated and/or derived from alternate sources, including from any one of various types of lymphoid tissue.

5 Examples of samples isolated and/or derived from blood include samples of whole blood, serum-reduced whole blood, serum-depleted blood, and serum-depleted and erythrocyte depleted blood.

 Unless otherwise indicated herein, samples obtained from any individual may be used in accordance with the methods of the invention. Examples of individuals from which such a
10 sample may be obtained and utilized in accordance with the methods of the invention include, but are not limited to, individuals suspected of having one or more colorectal pathologies, individuals diagnosed as having one or more colorectal pathologies; individuals that have not been diagnosed with having one or more colorectal pathologies; individuals who have been confirmed as not having one or more colorectal pathologies.

15 In a further embodiment, the individual from whom a sample may be obtained is a test subject wherein it is unknown whether the person has one or more colorectal pathologies or not. In another embodiment, the individual from whom a sample may be obtained is a test subject wherein it is unknown whether the person has one or more colorectal pathologies or not.

Blood

20 In one aspect of the invention, a sample of blood is obtained from an individual according to methods well known in the art. A sample of blood may be obtained from an individual, for example a subject having one or more colorectal pathologies, suspected of having one or more colorectal pathologies or not having one or more colorectal pathologies. In some embodiments, a drop of blood is collected from a simple pin prick made in the skin of an individual. Blood
25 may be drawn from an individual from any part of the body (*e.g.*, a finger, a hand, a wrist, an arm, a leg, a foot, an ankle, a stomach, and a neck) using techniques known to one of skill in the art, in particular methods of phlebotomy known in the art.

 The amount of blood collected will vary depending upon the site of collection, the amount required for a method of the invention, and the comfort of the individual. However, an
30 advantage of one embodiment of the present invention is that the amount of blood required to implement the methods of the present invention can be so small that more invasive procedures

are not required to obtain the sample. For example, in some embodiments, all that is required is a drop of blood. This drop of blood can be obtained, for example, from a simple pinprick. In some embodiments, any amount of blood is collected that is sufficient to detect the expression of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 or all of the genes in **Table 1** and **Table 2** and **Table 11** and **Table 12**. As such, in some embodiments, the amount of blood that is collected is 1 ml or less, 0.5 ml or less, 0.1 ml or less, or 0.01 ml or less. However, the present invention is not limited to such embodiments. In some embodiments more blood is available and in some embodiments, more blood can be used to effect the methods of the present invention. As such, in various specific embodiments, 0.001 ml, 0.005 ml, 0.01 ml, 0.05 ml, 0.1 ml, 0.15 ml, 0.2 ml, 0.25 ml, 0.5 ml, 0.75 ml, 1 ml, 1.5 ml, 2 ml, 3 ml, 4 ml, 5 ml, 10 ml, 15 ml or more of blood is collected from a subject. In another embodiment, 0.001 ml to 15ml, 0.01 ml to 10 ml, 0.1 ml to 10 ml, 0.1 ml to 5 ml, 1 to 5 ml of blood is collected from an individual. In a further embodiment, 0.001-100ml, preferably 0.01-50ml, more preferably 0.01-25ml and most preferably 0.01-1ml of blood is collected from an individual.

In some embodiments of the present invention, blood is stored within a K3/EDTA tube (e.g., from Becton Dickinson). In another embodiment, one can utilize tubes for storing blood which contain stabilizing agents such as disclosed in U.S. Patent No. 6,617,170 (which is incorporated herein by reference). In another embodiment the PAXgeneTM blood RNA system: provided by PreAnalytiX, a Qiagen/BD company, may be used to collect blood. In yet another embodiment, the TempusTM blood RNA collection tubes, offered by Applied Biosystems may be used. TempusTM collection tubes provide a closed evacuated plastic tube containing RNA stabilizing reagent for whole blood collection.

The blood collected is in some embodiments utilized immediately or within 1 hour, 2 hours, 3 hours, 4 hours, 5 hours or 6 hours or is optionally stored at temperatures such as 4 °C, or at -20°C prior to use in accordance with the methods of the invention. In some embodiments, a portion of the blood sample is used in accordance with the invention at a first instance of time whereas one or more remaining portions of the blood sample (or fractions thereof) are stored for a period of time for later use. For longer-term storage, storage methods well known in the art, such as storage at cryo temperatures (e.g. below -60 °C) can be used. In some embodiments, in addition to storage of the blood or instead of storage of the blood, plasma, serum, isolated

nucleic acid or proteins are stored for a period of time for later use in accordance with methods known in the art.

In one aspect, whole blood is obtained from an individual according to the methods of phlebotomy well known in the art. Whole blood includes blood which can be used as is, and
5 includes blood wherein the serum or plasma has been removed or reduced, and the RNA or mRNA from the remaining blood sample has been isolated in accordance with methods well known in the art (*e.g.*, using, in some embodiments, gentle centrifugation at 300 to 800 xg for 5 to 10 minutes). In a specific embodiment, whole blood (*i.e.*, unfractionated blood) obtained from a subject is mixed with lysing buffer (*e.g.*, Lysis Buffer (1L): 0.6g EDTA; 1.0g KHCO₂, 8.2g
10 NH₄Cl adjusted to pH 7.4 (using NaOH)), the sample is centrifuged and the cell pellet retained, and RNA or mRNA extracted in accordance with methods known in the art ("lysed blood") (see for example Sambrook et al.). In one embodiment, it is helpful to use unfractionated whole blood is preferred since it avoids the costly and time-consuming process to separate out the cell types within the blood (Kimoto, 1998, Mol. Gen. Genet 258:233-239; Chelly J et al., 1989, Proc.
15 Nat. Acad. Sci. USA 86:2617-2621; Chelly J et al., 1988, Nature 333:858-860).

In some embodiments of the present invention, whole blood collected from an individual is fractionated (*i.e.*, separated into components) before isolated products of the biomarkers from the sample. In one embodiment, blood is serum depleted (or serum reduced). In another embodiment the blood is plasma depleted (or plasma reduced). In yet other embodiments blood
20 is erythrocyte depleted or reduced. In some embodiments erythrocyte reduction is performed by preferentially lysing the red blood cells. In other embodiments, erythrocyte depletion or reduction is performed by lysing the red blood cells and further fractionating the remaining cells. In yet other embodiments erythrocyte depletion or reduction is performed but the remaining cells are not further fractionated. In other embodiments blood cells are separated from whole blood
25 collected from an individual using other techniques known in the art. For example, blood collected from an individual can be subjected to Ficoll-Hypaque (Pharmacia) gradient centrifugation. Such centrifugation may separate various types of cells from a blood sample. In particular, Ficoll-Hypaque gradient centrifugation is useful to isolate peripheral blood leukocytes (PBLs) which can be used in accordance with the methods of the invention.

30 By way of example but not limitation, macrophages can be obtained as follows. Mononuclear cells are isolated from peripheral blood of a subject, by syringe removal of blood

followed by Ficoll-Hypaque gradient centrifugation. Tissue culture dishes are pre-coated with the subject's own serum or with AB+ human serum and incubated at 37°C for one hour. Non-adherent cells are removed by pipetting. Cold (4°C) 1mM EDTA in phosphate-buffered saline is added to the adherent cells left in the dish and the dishes are left at room temperature for
5 fifteen minutes. The cells are harvested, washed with RPMI buffer and suspended in RPMI buffer. Increased numbers of macrophages can be obtained by incubating at 37°C with macrophage-colony stimulating factor (M-CSF). Antibodies against macrophage specific surface markers, such as Mac-1, can be labeled by conjugation of an affinity compound to such molecules to facilitate detection and separation of macrophages. Affinity compounds that can be
10 used include but are not limited to biotin, photobiotin, fluorescein isothiocyanate (FITC), or phycoerythrin (PE), or other compounds known in the art. Cells retaining labeled antibodies are then separated from cells that do not bind such antibodies by techniques known in the art such as, but not limited to, various cell sorting methods, affinity chromatography, and panning.

Blood cells can be sorted using a using a fluorescence activated cell sorter (FACS).
15 Fluorescence activated cell sorting (FACS) is a known method for separating particles, including cells, based on the fluorescent properties of the particles. See, for example, Kamarch, 1987, Methods Enzymol 151:150-165. Laser excitation of fluorescent moieties in the individual particles results in a small electrical charge allowing electromagnetic separation of positive and negative particles from a mixture. An antibody or ligand used to detect a blood cell antigenic
20 determinant present on the cell surface of particular blood cells is labeled with a fluorochrome, such as FITC or phycoerythrin. The cells are incubated with the fluorescently labeled antibody or ligand for a time period sufficient to allow the labeled antibody or ligand to bind to cells. The cells are processed through the cell sorter, allowing separation of the cells of interest from other cells. FACS sorted particles can be directly deposited into individual wells of microtiter plates to
25 facilitate separation.

Magnetic beads can be also used to separate blood cells in some embodiments of the present invention. For example, blood cells can be sorted using a using a magnetic activated cell sorting (MACS) technique, a method for separating particles based on their ability to bind magnetic beads (0.5-100 m diameter). A variety of useful modifications can be performed on the
30 magnetic microspheres, including covalent addition of an antibody which specifically recognizes a cell-solid phase surface molecule or hapten. A magnetic field is then applied, to physically

manipulate the selected beads. In a specific embodiment, antibodies to a blood cell surface marker are coupled to magnetic beads. The beads are then mixed with the blood cell culture to allow binding. Cells are then passed through a magnetic field to separate out cells having the blood cell surface markers of interest. These cells can then be isolated.

5 In some embodiments, the surface of a culture dish may be coated with antibodies, and used to separate blood cells by a method called panning. Separate dishes can be coated with antibody specific to particular blood cells. Cells can be added first to a dish coated with blood cell specific antibodies of interest. After thorough rinsing, the cells left bound to the dish will be cells that express the blood cell markers of interest. Examples of cell surface antigenic
10 determinants or markers include, but are not limited to, CD2 for T lymphocytes and natural killer cells, CD3 for T lymphocytes, CD11a for leukocytes, CD28 for T lymphocytes, CD19 for B lymphocytes, CD20 for B lymphocytes, CD21 for B lymphocytes, CD22 for B lymphocytes, CD23 for B lymphocytes, CD29 for leukocytes, CD14 for monocytes, CD41 for platelets, CD61 for platelets, CD66 for granulocytes, CD67 for granulocytes and CD68 for monocytes and
15 macrophages.

Whole blood can be separated into cell types such as leukocytes, platelets, erythrocytes, *etc.* and such cell types can be used in accordance with the methods of the invention. Leukocytes can be further separated into granulocytes and agranulocytes using standard techniques and such cells can be used in accordance with the methods of the invention.
20 Granulocytes can be separated into cell types such as neutrophils, eosinophils, and basophils using standard techniques and such cells can be used in accordance with the methods of the invention. Agranulocytes can be separated into lymphocytes (*e.g.*, T lymphocytes and B lymphocytes) and monocytes using standard techniques and such cells can be used in accordance with the methods of the invention. T lymphocytes can be separated from B lymphocytes and
25 helper T cells separated from cytotoxic T cells using standard techniques and such cells can be used in accordance with the methods of the invention. Separated blood cells (*e.g.*, leukocytes) can be frozen by standard techniques prior to use in the present methods.

(D) RNA Preparation

In one aspect of the invention, RNA is isolated from an individual in order to measure the
30 RNA products of the biomarkers of the invention. RNA is isolated from a sample from individuals diagnosed as having one or more colorectal pathologies including one or more polyps

or one or more subtype of polyps, individuals not having one or more colorectal pathologies, not having one or more polyps or not having a subtype of polyps, or test subjects.

In some embodiments, RNA is isolated from blood which is erythrocyte depleted by the following protocol. Lysis Buffer is added to blood sample in a ratio of 3 parts Lysis Buffer to 1 part blood (Lysis Buffer (1L) 0.6g EDTA; 1.0g KHCO₂, 8.2g NH₄Cl adjusted to pH 7.4 (using NaOH)). Sample is mixed and placed on ice for 5-10 minutes until transparent. Lysed sample is centrifuged at 1000 rpm for 10 minutes at 4°C, and supernatant is aspirated. Pellet is resuspended in 5ml Lysis Buffer, and centrifuged again at 1000 rpm for 10 minutes at 4°C. Pelleted cells are homogenized using TRIzol® (GIBCO/BRL) in a ratio of approximately 6ml of TRIzol® for every 10ml of the original blood sample and vortexed well. Samples are left for 5 minutes at room temperature. RNA is extracted using 1.2 ml of chloroform per 1 ml of TRIzol®. Sample is centrifuged at 12,000 x g for 5 minutes at 4°C and upper layer is collected. To upper layer, isopropanol is added in ratio of 0.5 ml per 1 ml of TRIzol®. Sample is left overnight at -20°C or for one hour at -20°C. RNA is pelleted in accordance with known methods, RNA pellet air dried, and pellet resuspended in DEPC treated ddH₂O. RNA samples can also be stored in 75% ethanol where the samples are stable at room temperature for transportation.

In other aspects, RNA is prepared by first collecting blood into a PAXgene™ collection tube and then subsequently isolating the RNA using the PAXgene™ blood RNA isolated system provided by PreAnalytiX, a Qiagen/BD company. In another embodiment, RNA is prepared by first collecting blood into any known stabilizing solution (e.g. a PAXgene™ collection or a TEMPUS® collection tube and then isolating the RNA using any method known to a person skilled in the art.

In other aspects globin reduced or depleted RNA is prepared. In one embodiment RNA is isolated first and then is subsequently treated to remove globin mRNA using one of any technique known in the art. For example, one can hybridize DNA primers and/or probes specific for globin RNA and utilize RNase H to selectively degrade globin mRNA. In other embodiments RNA is isolated in a manner which removes the globin RNA during the RNA isolation steps (for example reducing globin RNA by selectively removing globin RNA using globin primers and/or probes attached to paramagnetic particles).

In other aspects of the invention RNA is prepared using one or more known commercial kits for isolating RNA (including isolating total RNA or mRNA and the like) such as oligo dT based purification methods, Qiagen® RNA isolation methods, LeukoLOCKT Total RNA Isolation System, MagMAX-96 Blood Technology from Ambion, Promega® polyA mRNA isolation system and the like.

Purity and integrity of RNA can be assessed by absorbance at 260/280nm and agarose gel electrophoresis followed by inspection under ultraviolet light. In some embodiments RNA integrity is assessed using more sensitive techniques such as the Agilent 2100 Bioanalyzer 6000 RNA Nano Chip.

10 (E) Biomarkers of the Invention

In one aspect, the invention provides biomarkers and biomarker combinations wherein the measure of the level of expression of the product or products of said biomarkers is indicative of the presence of one or more colorectal pathologies.

Table 1 is a list of biomarkers of one aspect of the invention. Each biomarker is differentially expressed in samples from individuals having or not having polyps using microarray assays. The table provides the Hugo Gene name, symbol and locus link ID; the RNA and protein accession number; and also includes both the p value (which represents the statistical significance of the observed differential expression) and a measure of the fold change as between the average measured level of individuals having polyps and the average measured level of individuals not having polyps.

Table 2 is a selection of those genes listed in **Table 1** and lists the gene symbol and the associated locus link ID for the biomarkers of the invention. The table also provides the fold change and direction of differential gene expression in individuals having polyps as compared to individuals not having polyps. As described, differential expression of the genes between individuals having or not having polyps can be identified using a non-parametric Wilcoxon-Mann-Whitney test or a parametric t test. The results of the tests are also shown in **Table 2**.

Table 11 shows genes identified as differentially expressed in samples from individuals having “high risk polyps” as compared with individuals not having high risk polyps (ie having low risk polyps or having no pathology at all) using microarray as described in **Example 2**. The table provides the gene name, gene ID; a representative human RNA accession number, and also provides the p value, the fold change (as between the average of individuals classified as

having high risk polyps as compared with the average of individuals having low risk polyps), along with the coefficient of variation for both the high risk polyp individuals and the low risk polyp individuals (the standard deviation of the normalized intensity divided by the mean normalized intensity). Column 1 is AffySpotID, column 2 is Fold Change, column 3 is p value, Column 4 is CV (Coefficient of Variation) (High RiskPolyp), column 5 is CV(Low Risk Polyp), column 6 is Gene ID, column 7 is the HUGO Gene Symbol, column 8 is the Human RNA Accession Number and column 9 is the Gene Description.

Table 12 shows 48 biomarkers tested for differentially expression by QRT-PCR in samples from individuals having colorectal cancer and individuals not having colorectal cancer. The 48 biomarkers were tested using QRT-PCR. The table provides the gene symbol, locus link ID, and gene description for each biomarker. The table also includes the p value (which represents the statistical significance of the observed differential expression), the measure of the fold change as between the average measured level of individuals having colorectal cancer and the average measured level of individuals not having colorectal cancer and the direction of the differential expression between individuals having colorectal cancer and not having colorectal cancer.

Other biomarkers of the invention are described within the specification. The invention thus encompasses the use of those methods known to a person skilled in the art to measure the expression of these biomarkers and combinations of biomarkers for each of the purposes outlined above.

As would be understood by a person skilled in the art, the locus link ID can be used to determine the sequence of all the RNA products and all the protein products of the biomarkers of the invention.

(F) Combinations of Biomarkers

In one embodiment, combinations of biomarkers of the present invention includes any combination of the biomarkers listed in **Table 1**, **Table 2**, **Table 11**, or **Table 12**. For instance, the number of possible combinations of a subset n of m genes in any of the tables above is described in Feller, *Intro to Probability Theory*, Third Edition, volume 1, 1968, ed. J. Wiley, using the general formula:

$$m!/(n)! (m-n)!$$

For example, where n is 2 and m is 8, the number of combinations of biomarkers is:

$$\begin{aligned} \frac{8!}{2! (8-2)!} &= \frac{8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1) (6 \times 5 \times 4 \times 3 \times 2 \times 1)} \\ &= 40320/1440 = 28 \end{aligned}$$

unique two-gene combinations. The measurement of the gene expression of each of these two-gene combinations can independently be used to determine whether a patient has one or more colorectal pathologies. In another specific embodiment in which m is 8 and n is three, there are $8!/3!(8-3)!$ unique three-gene combinations. Each of these unique three-gene combinations can independently serve as a model for determining whether a patient has one or more colorectal pathologies.

10 (G) Testing Combinations of Biomarkers by Generating Formulas Resulting from One or More Classifiers

The invention further provides a means of testing combinations of biomarkers from **Table 1, Table 2, Table 11, or Table 12** or subsets thereof for their ability to test for one or more colorectal pathologies or one or more subtypes of colorectal pathology. Also provided are methods of evaluating the combinations tested tested for their ability to test an individual for the presence of one or more colorectal pathologies or one or more subtypes of colorectal pathology. In order to test combinations of biomarkers and generate classifiers, a mathematical model of the invention can be used. A mathematical model of the invention can be used to test each selected combination of biomarkers from all combinations of biomarkers or a selected subset thereof.

20 In some embodiments, it is useful to further select biomarkers to be tested as combinations. in one embodiment, one can select individual biomarkers on the basis of the p value as a measure of the likelihood that the the individual biomarker can distinguish as between the two phenotypic trait subgroups. Thus in one embodiment, biomarkers are chosen to test in combination by input into a model wherein the p value of each biomarker is less than 0.2, 0.1, 25 0.5; less than 0.1, less than 0.05, less than 0.01, less than 0.005, less than 0.001, less than 0.0005, less than 0.0001, less than 0.00005, less than 0.00001, less than 0.000005, less than 0.000001 etc. We have also surprisingly found that even biomarkers which demonstrate a p value of greater than 0.2 (and thus would normally not be considered to be a useful individual biomarker) do significantly increase the ability of a combination of biomarkers in which they are included to 30 distinguish as between two phenotypic trait subgroups. In other embodiments, biomarkers for input into the model to test in combination are chosen on the basis of the fold change of

differential expression of the product of the biomarker as between the the two phenotypic trait subgroups. Note that in measuring differential fold change in blood, the fold change differences can be quite small, thus in some embodiments, selection of biomarker subsets for input into classifier is based on a differential fold change where the fold change is greater than 1.01, 1.02, 1.03, 1.04, 1.05, 1.06, 1.07, 1.08, 1.09, 1.1, 1.125, 1.15, 1.175, 1, 1.2, 1.225, 1.25, 1.275, 1.30, greater than 1.3, greater than 1.4, greater than 1.5, greater than 1.6, greater than 1.7, greater than 1.8, greater than 1.9, greater than 2.0, greater than 2.1, greater than 2.2, greater than 2.3, greater than 2.4, greater than 2.5, greater than 2.6, greater than 2.7, greater than 2.8, greater than 2.9, greater than 3.0, greater than 3.1, greater than 3.2, greater than 3.3, greater than 3.4 greater than 3.5, greater than 4.0 and the like. In yet other embodiments in order to select subsets of biomarkers to test in combination, one can also take into account the coefficient of variation as a variability of the data representing the level of expression of the product of the biomarker amongst individuals within a phenotypic trait subgroup. In some embodiments, it is helpful to select biomarkers on the basis of a combination of factors including p value, fold change, and coefficient of variation as would be understood by a person skilled in the art. In some embodiments, biomarkers are first selected as outlined above on the basis of the p value resulting from the biomarker data and then a subselection of said biomarkers is chosen on the basis of the differential fold change determined from the biomarker data. In other embodiments, biomarkers are first selected on the basis of differential fold change, and then subselection is made on the basis of p value. In some embodiments, the use of one or more of the selection criteria and subsequent ranking permits the selection of the top 2.5%, 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20%, 30%, 40%, 50% or more of the ranked biomarkers for input into the model. In some embodiments, the desired number of selected biomarkers can be 4,000; 3,000; 2,000; 1,000; 900; 800; 700; 600; 500; 400; 300; 200; 190; 180; 170; 160; 150; 140; 130; 120; 110; 100; 90; 80; 70; 60; 50; 40; 30; 20; or 10. In other embodiments, the selection criteria noted above can be set on the basis of the desired number of selected biomarkers for use in the model. As would be understood, one can select therefore all of the individually identified biomarkers or subsets of the individually identified biomarkers and test all possible combinations of the selected biomarkers to identify useful combinations of biomarkers. In another embodiment, one can select a subset of biomarkers and then test all possible combinations of 2 biomarkers from that subset, 3 biomarkers from that subset, 4 biomarkers from that subset, 5 biomarkers from that subset, 6

biomarkers from that subset 7 biomarkers from that subset, 8 biomarkers from that subset 9
biomarkers from that subset or 10 biomarkers from that subset in order to identify useful
combinations of biomarkers. A selection criteria to determine the number of selected individual
biomarkers to test in combination, and to select the number of possible combinations of
5 biomarkers will depend upon the resources available for obtaining the biomarker data and/or the
computer resources available for calculating and evaluating classifiers resulting from the model.

The classifier generated by the mathematical model can be subsequently evaluated by
determining the ability of the classifier to correctly call each individual for one of the two
phenotypic traits of the population used to generate the classifier (ie having or not having one or
10 more colorectal pathologies). In a preferred embodiment, the individuals of the training
population used to derive the model are different from the individuals of the training population
used to test the model. As would be understood by a person skilled in the art, this allows one to
predict the ability of the combinations as to their ability to properly characterize an individual
whose phenotypic trait characterization is unknown.

15 The data which is input into the mathematical model can be any data which is
representative of the expression level of the product of the biomarkers being evaluated.
Mathematical models useful in accordance with the invention include those using both
supervised or unsupervised learning techniques. In a preferred embodiment of the invention, the
mathematical model chosen uses supervised learning in conjunction with a "training population"
20 to evaluate each of the possible combination of biomarkers of the invention. In one embodiment
of the invention, the mathematical model used is selected from the following: a regression
model, a logistic regression model, a neural network, a clustering model, principal component
analysis, nearest neighbour classifier analysis, linear discriminant analysis, quadratic
discriminant analysis, a support vector machine, a decision tree, a genetic algorithm, classifier
25 optimization using bagging, classifier optimization using boosting, classifier optimization using
the Random Subspace Method, a projection pursuit, genetic programming and weighted voting.
In a preferred embodiment, a logistic regression model is used. In another preferred
embodiment, a neural network model is used.

The results of applying a mathematical model of the invention to the data will generate
30 one or more classifiers using one or more biomarkers. In some embodiments, multiple classifiers
are created which are satisfactory for the given purpose (e.g. all have sufficient AUC and/or

sensitivity and/or specificity). In this instance, in some embodiments, a formula is generated which utilizes more than one classifier. For example, a formula can be generated which utilizes classifiers in series (e.g. first obtains results of classifier A, then classifier B e.g. Classifier A differentiates pathology from non pathology; classifier B then determines whether the pathology is colorectal cancer or not colorectal cancer). In another embodiment, a formula can be generated which results from weighting the results of more than one classifier. For example, the results of each classifier can be given a score of 1 and an indication of probability of a test subject having one or more colorectal pathologies is the result of the aggregate score of each of the selected classifiers of a given formula. Other possible combinations and weightings of classifiers would be understood and are encompassed herein.

Classifiers generated can be used to test an unknown or test subject. In one embodiment, the results from equations generated by logistic regression to answer the question does an individual have one or more colorectal pathologies or is an individual "normal". In yet another embodiment of the invention, the answer to the question above may be an answer of non-determinable.

In one embodiment of the invention, each classifier is evaluated for its ability to properly characterize each individual of the training population using those methods known to a person skilled in the art. For example one can evaluate the classifier using cross validation, Leave One out Cross Validation (LOOCV), n-fold cross validation, jackknife analysis using standard statistical methods as disclosed. In another embodiment of the invention, each classifier is evaluated for its ability to properly characterize those individuals of the training population which were not used to generate the classifier.

In one embodiment, the method used to evaluate the classifier for its ability to properly characterize each individual of the training population is a method which evaluates the classifier's sensitivity (TPF, true positive fraction) and 1-specificity (TNF, true negative fraction). In one embodiment, the method used to test the classifier is Receiver Operating Characteristic ("ROC") which provides several parameters to evaluate both the sensitivity and specificity of the result of the equation generated. In one embodiment using the Receiver Operating Characteristic ("ROC") the ROC area (area under the curve) is used to evaluate the equations. A ROC area greater than 0.5, 0.6, 0.7, 0.8, 0.9 is preferred. A perfect ROC area score of 1.0 indicates with both 100% sensitivity and 100% specificity. In some embodiments

classifiers are selected on the basis of the score. For example, where the scoring system used is receiver operating characteristic (ROC) curve score determined by an area under the ROC curve, in some embodiments, those classifiers with scores of greater than 0.95, 0.9, 0.85, 0.8, 0.7, 0.65, 0.6, 0.55 0.5 or 0.45 are chosen. In other embodiments, where specificity is important to the use
5 of the classifier, a sensitivity threshold can be set and classifiers ranked on the basis of the specificity chosen. For example classifiers with a cutoff for specificity of greater than 0.95, 0.9, 0.85, 0.8, 0.7, 0.65, 0.6, 0.55 0.5 or 0.45 can be chosen. Similarly, the specificity threshold can be set and classifiers ranked on the basis of sensitivity greater than 0.95, 0.9, 0.85, 0.8, 0.7, 0.65, 0.6, 0.55 0.5 or 0.45 can be chosen. Thus in some embodiments, only the top 10 ranking
10 classifiers, the top 20 ranking classifiers, or the top 100 ranking classifiers are selected.

As would be understood by a person skilled in the art, the utility of the combinations and classifiers determined by a mathematical model will depend upon the phenotypes of the populations used to generate the data for input into the model. Examples of specific embodiments are described more thoroughly herein.

15 (H) Populations for Input Into the Mathematical Models

Populations used for input should be chosen so as to result in statistically significant resulting classifier. In some embodiments, the reference or training population includes between 10 and 30 subjects. In another embodiment the reference population contains between 30-50 subjects. In still other embodiments, the reference population includes two or more populations
20 each containing between 50 and 100, 100 and 500, between 500 and 1000, or more than 1000 subjects. The reference population includes two or more subpopulations. In a preferred embodiment, the phenotypic trait characteristics of the subpopulations are similar but for the diagnosis with respect to the presence of one or more colorectal pathologies, for example the distribution within the subpopulations are similar with regards to the age and sex of the
25 subpopulations. It is also preferred that the subpopulations are of roughly equivalent numbers. It is to be understood that the methods herein do not require using data from every member of a population, but instead may rely on data from a subset of a population in question.

For example, for a reference or test population for input into a mathematical model to identify those biomarkers which are useful in identifying an individual as having any polyps or
30 not having any polyps, the reference population is comprised of individuals having polyps (the first subpopulation), and individuals not having polyps (the second subpopulation). For purposes

of characterizing the subpopulations as having or not having polyps, any verified method can be used including digital rectal examination, fecal occult blood testing, rigid sigmoidoscopy, flexible sigmoidoscopy, double-contrast barium enema, colonoscopy, and histological examination. Preferably only those individuals whose diagnoses are certain are utilized as part
5 of the reference population.

In another embodiment, to identify those biomarkers which are useful in identifying an individual as having high risk polyps or not, the reference population is comprised of individuals having high risk polyps (the first subpopulation), and individuals not having high risk polyps (the second subpopulation) where high risk polyps are the following: Tubulovillous Adenoma,
10 Villous Adenoma, Cancer High Grade Dysplasia and Tubular Adenoma where the Tubular Adenoma is greater than 10mm. For purposes of characterizing the subpopulations as having or not having high risk polyps, any verified method can be used including digital rectal examination, fecal occult blood testing, rigid sigmoidoscopy, flexible sigmoidoscopy, double-contrast barium enema, colonoscopy, and histological examination.

15 In yet another embodiment, to test biomarkers which are useful in identifying an individual as having early stage of colorectal cancer or not, the reference population can, for example be comprised of individuals having localized colorectal cancer as compared with individuals with other types of colorectal cancer (e.g. late stage).

In another embodiment, to identify those biomarkers which are useful in identifying an
20 individual as having high risk polyps or not, the reference population is comprised of individuals having high risk polyps (the first subpopulation), and individuals not having high risk polyps (the second subpopulation) where high risk polyps are the following: Tubulovillous Adenoma; Villous Adenoma; Cancer; High Grade Dysplasia; and Tubular Adenoma.

25 **(I) Data for Input into the Mathematical Models to Identify Classifiers for Testing for Colorectal Pathology**

Data for input into the mathematical models is data representative of the level of the products of the biomarkers of the invention. As such the data is the measure of the level of expression of the products of the biomarkers of the invention including either mRNA and/or protein.

30 In one embodiment of the invention, the RNA products of the biomarkers of the invention which are measured are the population of RNA products including the mRNA, and all

of the spliced variants of the mRNA. In another embodiment of the invention the products measured are all of the mRNA products expressed in blood. In yet another embodiment of the invention, the products measured include one or more specific spliced variants of the mRNA which are expressed in blood. In yet another embodiment of the invention, the products
 5 measured are the RNA products listed in **Table 3** or **Table 13**.

Protein products of the biomarkers of the invention are also included within the scope of the invention. To practice the invention, measurement of the protein products of the biomarkers of the invention can be used for purposes of testing for one or more colorectal pathologies. More particularly, measurement of those populations of protein products of the biomarkers which are
 10 differentially expressed in individuals having or not having any polyps are useful for purposes of testing and are encompassed herein.

In one embodiment of the invention the protein products are those translated from the biomarkers listed in **Table 1**, **Table 2**, **Table 11**, or **Table 12**. In another embodiment, the protein products are those which are expressed in blood. In yet another embodiment of the
 15 invention, the protein products are those corresponding to the proteins listed in **Table 3** or **Table 13**.

In yet another embodiment, data reflective of the level of expression of a combination of protein products and RNA products of the biomarkers are used. As would be understood by a person skilled in the art, other combinations of input data can be utilized to generate classifiers
 20 useful in accordance with the invention.

In other embodiments, as would be understood by one of ordinary skill in the art, data reflective of each biomarker in each member of the population is not necessary so long as there are data for sufficient members of each reference population to permit creation of a classifier. For example, data representative of biomarkers in 99%, 95%, 90%, 85%, 80%, or 75% of
 25 members of a population may suffice in given circumstances.

(J) Mathematical Models

Formulae for use with the methods described herein may generally have the form:

$$V = C + \Sigma \beta_{ij} f(X_i) + \Sigma \beta_{ij} f(X_i, X_j) + \Sigma \beta_{ijk} f(X_i, X_j, X_k) + \dots$$

Wherein V is a value indicating the probability that a test subject has one or more colorectal
 30 pathologies, X_i is a level of one or more products of an i th biomarker in a sample from the test subject, β_i is a coefficient for a term involving only the i th biomarker, β_{ij} is a coefficient for a

term that is a function of the i th and j th biomarkers, β_{ijk} is a coefficient for a term that is a function of the i th, j th and k th biomarkers, and C is a constant. Still other terms may find themselves in this formula, such as terms depending on four or more biomarkers.

By 'indicates' is meant that V might be an actual probability (a number varying between 0 and 1), or V might be a quantity from which a probability can be readily derived.

There are various forms of functions $f(X_i, X_j, \dots)$ that depend on expression levels of the various biomarkers. For example, the functions may be polynomials in those expression levels, *i.e.*, involve products of the various biomarkers raised to numeric powers. Examples include: $X_i X_j^2$, $X_i X_j X_k$, $(X_i X_j)^{1/2}$, $X_i X_j + X_i X_k$. The functions may additionally or alternatively involve logarithms, exponentials, or still other functions of the expression levels.

In certain embodiments, the $f(X_i, X_j, \dots)$ depend on ratios of the biomarker expression levels, *i.e.*, $f(X_i, X_j) = X_i/X_j$.

Regression Models

In some embodiments the expression data for some or all of the biomarkers identified in the present invention are used in a regression model, such as but not limited to a logistic regression model or a linear regression model, so as to identify classifiers useful in diagnosing one or more colorectal pathologies. The regression model is used to test various combinations of two or more of the biomarkers identified in **Table 1**, **Table 2**, **Table 11**, or **Table 12** to generate classifiers. In the case of regression models, the classifiers which result are in the form of equations which provide a dependent variable Y , which represents the presence or absence of a given phenotype where the data representing the expression of each of the biomarkers in the equation is multiplied by a weighted coefficient as generated by the regression model. The classifiers generated can be used to analyze expression data from a test subject and provide a result indicative of the probability of a test subject having one or more colorectal pathologies. In general, a multiple regression equation of interest can be written

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

where Y , the dependent variable, indicates presence (when Y is positive) or absence (when Y is negative) of the biological feature (e.g., absence or presence of one or more colorectal pathologies) associated with the first subgroup. This model says that the dependent variable Y depends on k explanatory variables (the measured characteristic values for the k select genes (e.g., the biomarkers) from subjects in the first and second subgroups in the reference

population), plus an error term that encompasses various unspecified omitted factors. In the above-identified model, the parameter β_1 gauges the effect of the first explanatory variable X_1 on the dependent variable Y (e.g., a weighting factor), holding the other explanatory variables constant. Similarly, β_2 gives the effect of the explanatory variable X_2 on Y , holding the remaining explanatory variables constant.

A logistic regression model is a non-linear transformation of the linear regression. The logistic regression model is often referred to as the “logit” model and can be expressed as

$$\ln[p/(1-p)] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad \text{or}$$

$$[p/(1-p)] = \exp^\alpha \exp^{\beta_1 X_1} \exp^{\beta_2 X_2} \times \dots \times \exp^{\beta_k X_k} \exp^\varepsilon$$

where,

α and ε are constants

$\ln$ is the natural logarithm, $\log_e$, where $e=2.71828\dots$,

p is the probability that the event Y occurs, $p(Y=1)$,

$p/(1-p)$ is the “odds ratio”,

$\ln[p/(1-p)]$ is the log odds ratio, or “logit”, and

all other components of the model are the same as the general linear regression equation described above. It will be appreciated by those of skill in the art that the term for α and ε can be folded into a single constant. Indeed, in preferred embodiments, a single term is used to represent α and ε . The “logistic” distribution is an S-shaped distribution function. The logit distribution constrains the estimated probabilities (p) to lie between 0 and 1.

In some embodiments of the present invention, the logistic regression model is fit by maximum likelihood estimation (MLE). In other words, the coefficients (e.g., α , β_1 , β_2 , ...) are determined by maximum likelihood. A likelihood is a conditional probability (e.g., $P(Y|X)$, the probability of Y given X). The likelihood function (L) measures the probability of observing the particular set of dependent variable values ($Y_1, Y_2, \dots, Y_n$) that occur in the sample data set. It is written as the probability of the product of the dependent variables:

$$L = \text{Prob} (Y_1 * Y_2 *** Y_n)$$

The higher the likelihood function, the higher the probability of observing the Y s in the sample. MLE involves finding the coefficients (α , β_1 , β_2 , ...) that makes the log of the likelihood function ($LL < 0$) as large as possible or -2 times the log of the likelihood function ($-2LL$) as small as possible. In MLE, some initial estimates of the parameters α , β_1 , β_2 , ... are made. Then

the likelihood of the data given these parameter estimates is computed. The parameter estimates are improved the likelihood of the data is recalculated. This process is repeated until the parameter estimates do not change much (for example, a change of less than .01 or .001 in the probability). Examples of logistic regression and fitting logistic regression models are
5 found in Hastie, *The Elements of Statistical Learning*, Springer, New York, 2001, pp. 95-100 which is incorporated herein in its entirety.

Neural Networks

In another embodiment, the expression measured for each of the biomarkers of the present invention can be used to train a neural network. A neural network is a two-stage regression or
10 classification model. A neural network can be binary or non binary. A neural network has a layered structure that includes a layer of input units (and the bias) connected by a layer of weights to a layer of output units. For regression, the layer of output units typically includes just one output unit. However, neural networks can handle multiple quantitative responses in a seamless fashion. As such a neural network can be applied to allow identification of biomarkers
15 which differentiate as between more than two populations (ie more than two phenotypic traits). In one specific example, a neural network can be trained using expression data from the products of the biomarkers in Table 1, Table 2, Table 11, or Table 12 to identify those combinations of biomarkers which are specific for one or more colorectal pathologies. As a result, the trained neural network can be used to directly identify combinations of biomarkers useful to test for one
20 or more colorectal pathologies. In some embodiments, the back-propagation neural network (see, for example Abdi, 1994, "A neural network primer", J. Biol System. 2, 247-283) containing a single hidden layer of ten neurons (ten hidden units) found in EasyNN-Plus version 4.0g software package (Neural Planner Software Inc.) is used.

Neural networks are described in Duda *et al.*, 2001, *Pattern Classification*, Second
25 Edition, John Wiley & Sons, Inc., New York; and Hastie *et al.*, 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York which is incorporated herein in its entirety.

Singular Value Decomposition (SVD) and Principal Component Analysis (PCA)

Singular value decomposition (SVD) and Principal Component Analysis (PCA) are common techniques for analysis of multivariate data, and we have found that gene expression
30 data is well suited to analysis using SVD/PCA. SVD or equivalently, in this case, PCA, is defined as follows:

Singular value decomposition (SVD) and Principal Component Analysis (PCA) are common techniques for analysis of multivariate data, and we have found gene expression data are well suited to analysis using SVD/PCA. SVD or equivalently, in this case, PCA, is defined as follows:

5 Let G be an $m \times n$ gene expression matrix with rank r , and $m \geq n$, and therefore $r \leq n$ where m is a row and n is a column of data of the matrix. In the case of microarray data, g_{ij} is the level of one or more products of the i th biomarker in the j th assay. The elements of the i th row of G form the n -dimensional vector b_i (where b is a biomarker), which we refer to as the transcriptional response of the i th biomarker. Alternatively, the elements of the j th column of G
 10 form the m -dimensional vector a_j , which we refer to as the expression profile (or gene expression profile) of the j th assay.

The equation for singular value decomposition of G is the following:

$$G = USV^T$$

where U is an $m \times n$ matrix, S is an $n \times n$ diagonal matrix, and V^T is also an $n \times n$ matrix. The
 15 columns of U are called the left singular vectors, $\{u_k\}$, and form an orthonormal basis for the assay expression profiles, so that $u_i \cdot u_j = 1$ for $i = j$, and $u_i \cdot u_j = 0$ otherwise. The rows of V^T contain the elements of the right singular vectors, $\{v_k\}$, and form an orthonormal basis for the gene transcriptional responses. The elements of S are only nonzero on the diagonal, and are called the singular values. Thus, $S = \text{diag}(s_1, \dots, s_n)$. Furthermore, $s_k > 0$ for $1 \leq k \leq r$, and $s_i = 0$
 20 for $(r+1) \leq k \leq n$. By convention, the ordering of the singular vectors is determined by high-to-low sorting of singular values, with the highest singular value in the upper left index of the S matrix. Note that for a square, symmetric matrix X , singular value decomposition is equivalent to diagonalization, or solution of the eigenvalue problem.

Other Mathematical Models

25 The pattern classification and statistical techniques described above are merely examples of the types of models that can be used to construct classifiers useful for diagnosing or detecting one or more colorectal pathologies, for example clustering as described on pages 211-256 of Duda and Hart, *Pattern Classification and Scene Analysis*, 1973, John Wiley & Sons, Inc., New York, incorporated herein by reference in its entirety; Principal component analysis, (see for
 30 Jolliffe, 1986, *Principal Component Analysis*, Springer, New York, incorporated herein by

reference); nearest neighbour classifier analysis, (see for example Duda, *Pattern Classification*, Second Edition, 2001, John Wiley & Sons, Inc; and Hastie, 2001, *The Elements of Statistical Learning*, Springer, New York); linear discriminant analysis, (see for example Duda, *Pattern Classification*, Second Edition, 2001, John Wiley & Sons, Inc; and Hastie, 2001, *The Elements of Statistical Learning*, Springer, New York; Venables & Ripley, 1997, *Modern Applied Statistics with s-plus*, Springer, New York); Support Vector Machines (see, for example, Cristianini and Shawe-Taylor, 2000, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, Boser *et al.*, 1992, "A training algorithm for optimal margin classifiers, in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, ACM Press, Pittsburgh, PA, pp. 142-152; Vapnik, 1998, *Statistical Learning Theory*, Wiley, New York, incorporated herein by reference.)

Computer implementation

The methods described herein are preferably performed by a suitably programmed computer. The computer system for use with the methods described herein, as further described herein, is configured to accept and to process data and may be a single-processor or multi-processor computer system. Examples of suitable computer systems include, but are not limited to, any one of various combinations of mainframe computers, minicomputers, personal computers, laptop computers, notebook computers, hand-held computers, personal digital assistants, mobile phones, set-top boxes, microprocessor-based consumer electronics, programmable consumer electronics, and the like. Additionally, the methods of the invention may be practiced on networked computers, CPU-clusters, workstations, and so-called mainframe computers. The computer system may be a locally accessed computer, a remotely accessed computer system (e.g. server), or a combination of both. Depending on the application and purpose, the computer system may have access or be accessible to "the internet" [World Wide Web (WWW)]. It will be appreciated that the computer system may be a stand-alone system or a distributed system comprising multiple devices communicating with each other through a network. Depending on the application and purpose, the computer system may be a static or mobile computer system. One of ordinary skill in the art will possess the necessary knowledge and skills for selecting, obtaining and utilizing a suitable computer system for practicing any aspect of the invention.

It is therefore consistent with the description herein that various methods and formulae are implemented, in the form of computer program instructions, and executed on a computer as also described herein. Suitable programming languages for expressing the program instructions include, but are not limited to, one or more languages selected from the group consisting of: C, C++, an embodiment of FORTRAN such as FORTRAN77 or FORTRAN90, Java, Visual Basic, Perl, Tcl/Tk, JavaScript, and ADA. It is to be understood that various aspects of the methods may be written in different computing languages from one another, where such languages are preferred for particular applications, and the various aspects are caused to communicate with one another by appropriate system-level-tools available on a given computer.

The computer program instructions are stored in a computer memory during execution, and may additionally be stored on any of various forms of computer-readable media known in the art, such as, but not limited to, CD-Rom, CD-R, CD-RW, flash memory, memory cards, memory sticks, DVD-Rom, USB-sticks, optical discs, or high capacity network storage drives. It is thus consistent with ordinary practice of the present invention that the computer program instructions can be delivered to a user on a transferable medium such as a CD-Rom, and also delivered over a computer network, such as by downloading over the Internet through a web-interface.

FIG. 1 shows a schematic of a general-purpose computer system 100 suitable for practicing the methods described herein. The computer system 100, shown as a self-contained unit but not necessarily so limited, comprises at least one data processing unit (CPU) 102, a memory 104, which will typically include both high speed random access memory as well as non-volatile memory (such as one or more magnetic disk drives) but may be simply flash memory, a user interface 108, optionally a disk 110 controlled by a disk controller 112, and at least one optional network or other communication interface card 114 for communicating with other computers as well as other devices. At least the CPU 102, memory 104, user interface 108, disk controller where present, and network interface card, communicate with one another via at least one communication bus 106.

Memory 104 stores procedures and data, typically including: an operating system 140 for providing basic system services; application programs 152 such as user level programs for viewing and manipulating data, evaluating formulae for the purpose of diagnosing a test subject; authoring tools for assisting with the writing of computer programs; a file system 142, a user

interface controller 144 for handling communications with a user via user interface 108, and optionally one or more databases 146 for storing microarray data and other information, optionally a graphics controller 148 for controlling display of data, and optionally a floating point coprocessor 150 dedicated to carrying out mathematical operations. The methods of the present invention may also draw upon functions contained in one or more dynamically linked libraries, not shown in FIG. 1, but stored either in Memory 104, or on disk 110, or accessible via network interface connection 114.

User interface 108 may comprise a display 128, a mouse 126, and a keyboard 130. Although shown as separate components in FIG. 1, one or more of these user interface components can be integrated with one another in embodiments such as handheld computers. Display 128 may be a cathode ray tube (CRT), or flat-screen display such as an LCD based on active matrix or TFT embodiments, or may be an electroluminescent display, based on light emitting organic molecules such as conjugated small molecules or polymers. Other embodiments of a user interface not shown in FIG. 1 include, *e.g.*, several buttons on a keypad, a card-reader, a touch-screen with or without a dedicated touching device, a trackpad, a trackball, or a microphone used in conjunction with voice-recognition software, or any combination thereof, or a security-device such as a fingerprint sensor or a retinal scanner that prohibits an unauthorized user from accessing data and programs stored in system 100.

System 100 may also be connected to an output device such as a printer (not shown), either directly through a dedicated printer cable connected to a serial or USB port, or wirelessly, or via a network connection.

The database 146 may instead, optionally, be stored on disk 110 in circumstances where the amount of data in the database is too great to be efficiently stored in memory 104. The database may also instead, or in part, be stored on one or more remote computers that communicate with computer system 100 through network interface connection 114.

The network interface 134 may be a connection to the internet or to a local area network via a cable and modem, or ethernet, firewire, or USB connectivity, or a digital subscriber line. Preferably the computer network connection is wireless, *e.g.*, utilizing CDMA, GSM, or GPRS, or bluetooth, or standards such as 802.11a, 802.11b, or 802.11g.

It would be understood that various embodiments and configurations and distributions of the components of system 10 across different devices and locations are consistent with practice

of the methods described herein. For example, a user may use a handheld embodiment that accepts data from a test subject, and transmits that data across a network connection to another device or location wherein the data is analyzed according to a formulae described herein. A result of such an analysis can be stored at the other location and/or additionally transmitted back
5 to the handheld embodiment. In such a configuration, the act of accepting data from a test subject can include the act of a user inputting the information. The network connection can include a web-based interface to a remote site at, for example, a healthcare provider. Alternatively, system 10 can be a device such as a handheld device that accepts data from the test subject, analyzes the data, such as by inputting the data into a formula as further described
10 herein, and generating a result that is displayed to the user. The result can then be, optionally, transmitted back to a remote location via a network interface such as a wireless interface. System 100 may further be configured to permit a user to transmit by e-mail results of an analysis directly to some other party, such as a healthcare provider, or a diagnostic facility, or a patient.

15 **(K) Use of the Biomarkers of the Invention for Testing, Screening or Diagnosing a Test Subject**

As would be understood by a person skilled in the art, the identification of one or more biomarkers can be used to allow for the testing, screening or diagnosis of one or more colorectal pathologies including polyps or one or more subtypes of polyps within a test subject by
20 measuring the expression of the products of the biomarkers (gene) in the test subject (the "test subject").

In one embodiment, the results from the test subject are compared with the a control wherein the control can be results from one or more individuals having colorectal pathology, having polyps, having one or more subtypes of polyps and/or one or more individuals not
25 having any colorectal pathology, not having any polyps or not having one or more specific subtypes of colorectal polyps.

In another embodiment, one can input data reflective of the expression of the products of the biomarkers of the test subject into a formula of the invention resulting in a determination of whether said test subject has one or more colorectal pathologies. It is not necessary to use the
30 same formula used to test the biomarker combination for its ability to test for colorectal pathologies as to diagnose an individual using the biomarker combination identified. Data

representative of the products of the biomarkers of the invention (including RNA and/or Protein) is input into a formula of the invention so as to determine a probability of a test subject having one or more colorectal pathologies. The data can be generated using any technique known to measure the level of expression of either the RNA and protein products of the biomarkers of the invention.

In one embodiment, use of the formula results in a determination of whether the test subject has polyps or does not have polyps. For example, using logistic regression as the model, Y is used as a predictor of polyps, where when $Y > 0$ a person is diagnosed as having polyps and where $Y < 0$, a person is diagnosed as not having polyps. In yet another embodiment, one can also include a third category of prediction wherein diagnosis is indeterminable. For example, one can determine the standard deviation inherent within the methodology used to measure gene expression of the biomarkers (δ). If $Y < \delta$ but > 0 or $Y > -\delta$ but < 0 , then the test results are considered indeterminable.

(L) Polynucleotides Used to Measure the Products of the Biomarkers of the Invention

Polynucleotides capable of specifically or selectively binding to the RNA products of the biomarkers of the invention are used to measure the level of expression of the biomarkers. For example: oligonucleotides, cDNA, DNA, RNA, PCR products, synthetic DNA, synthetic RNA, or other combinations of naturally occurring or modified nucleotides which specifically and/or selectively hybridize to one or more of the RNA products of the biomarker of the invention are useful in accordance with the invention.

In a preferred embodiment, the oligonucleotides, cDNA, DNA, RNA, PCR products, synthetic DNA, synthetic RNA, or other combinations of naturally occurring or modified nucleotides oligonucleotides which both specifically and selectively hybridize to one or more of the RNA products of the biomarker of the invention are used.

(M) Techniques to Measure the RNA Products of the Biomarkers of the Invention Array Hybridization

In one embodiment of the invention, the polynucleotide used to measure the RNA products of the biomarkers of the invention can be used as nucleic acid members localized on a support to comprise an array according to one aspect of the invention. The length of a nucleic acid member can range from 8 to 1000 nucleotides in length and are chosen so as to be specific for the RNA products of the biomarkers of the invention. In one embodiment, these members are selective for the RNA products of the biomarkers of the invention. The nucleic acid members may be single

or double stranded, and/or may be oligonucleotides or PCR fragments amplified from cDNA. In some embodiments oligonucleotides are approximately 20-30 nucleotides in length. ESTs are in some embodiments 100 to 600 nucleotides in length. It will be understood to a person skilled in the art that one can utilize portions of the expressed regions of the biomarkers of the invention as
5 a probe on the array. More particularly oligonucleotides complementary to the genes of the invention and or cDNA or ESTs derived from the genes of the invention are useful. For oligonucleotide based arrays, the selection of oligonucleotides corresponding to the gene of interest which are useful as probes is well understood in the art. More particularly it is important to choose regions which will permit hybridization to the target nucleic acids. Factors such as the
10 T_m of the oligonucleotide, the percent GC content, the degree of secondary structure and the length of nucleic acid are important factors. See for example US Patent No. 6,551,784.

As described, microarrays can be used to identify and select genes differentially expressed in individuals having or not having one or more colorectal pathologies, one or more polyps or one or more subtypes of polyps, and can be used to diagnose or detect polyps or one or more
15 subtypes of polyps using the biomarkers of the invention. Genes identified as differentially expressed using microarrays can be seen in **Table 1**, and **Table 11**.

Construction of a nucleic acid array

In the subject methods, an array of nucleic acid members stably associated with the surface of a substantially support is contacted with a sample comprising target nucleic acids
20 under hybridization conditions sufficient to produce a hybridization pattern of complementary nucleic acid members/target complexes in which one or more complementary nucleic acid members at unique positions on the array specifically hybridize to target nucleic acids. The identity of target nucleic acids which hybridize can be determined with reference to location of nucleic acid members on the array.

25 The nucleic acid members may be produced using established techniques such as polymerase chain reaction (PCR) and reverse transcription (RT). These methods are similar to those currently known in the art (see e.g., *PCR Strategies*, Michael A. Innis (Editor), et al. (1995) and *PCR: Introduction to Biotechniques Series*, C. R. Newton, A. Graham (1997)). Amplified nucleic acids are purified by methods well known in the art (e.g., column purification or alcohol
30 precipitation). A nucleic acid is considered pure when it has been isolated so as to be substantially free of primers and incomplete products produced during the synthesis of the

desired nucleic acid. In some embodiments, a purified nucleic acid will also be substantially free of contaminants which may hinder or otherwise mask the specific binding activity of the molecule.

5 An array, according to one aspect of the invention, comprises a plurality of nucleic acids attached to one surface of a support at a density exceeding 20 different nucleic acids/cm², wherein each of the nucleic acids is attached to the surface of the support in a non-identical pre-selected region (e.g. a microarray). Each associated sample on the array comprises a nucleic acid composition, of known identity, usually of known sequence, as described in greater detail below. Any conceivable substrate may be employed in the invention.

10 In one embodiment, the nucleic acid attached to the surface of the support is DNA. In a preferred embodiment, the nucleic acid attached to the surface of the support is cDNA or RNA. In another preferred embodiment, the nucleic acid attached to the surface of the support is cDNA synthesized by polymerase chain reaction (PCR). In some embodiments, a nucleic acid member in the array, according to the invention, is at least 10, 25 or 50 nucleotides in length. In one
15 embodiment, a nucleic acid member is at least 150 nucleotides in length. In some embodiments, a nucleic acid member is less than 1000 nucleotides in length. More preferably, a nucleic acid member is less than 500 nucleotides in length.

In the arrays of the invention, the nucleic acid compositions are stably associated with the surface of a support. In one embodiment, the support may be a flexible or rigid support. By
20 "stably associated" is meant that each nucleic acid member maintains a unique position relative to the support under hybridization and washing conditions. As such, the samples are non-covalently or covalently stably associated with the support surface. Examples of non-covalent association include non-specific adsorption, binding based on electrostatic interactions (e.g., ion pair interactions), hydrophobic interactions, hydrogen bonding interactions, specific binding
25 through a specific binding pair member covalently attached to the support surface, and the like. Examples of covalent binding include covalent bonds formed between the nucleic acids and a functional group present on the surface of the rigid support (e.g., --OH), where the functional group may be naturally occurring or present as a member of an introduced linking group, as described in greater detail below

30 The amount of nucleic acid present in each composition will be sufficient to provide for adequate hybridization and detection of target nucleic acid sequences during the assay in which

the array is employed. Generally, the amount of each nucleic acid member stably associated with the support of the array is at least about 0.001 ng, preferably at least about 0.02 ng and more preferably at least about 0.05 ng, where the amount may be as high as 1000 ng or higher, but will usually not exceed about 20 ng. Where the nucleic acid member is "spotted" onto the support in
5 a spot comprising an overall circular dimension, the diameter of the "spot" will generally range from about 10 to 5,000 μm , usually from about 20 to 2,000 μm and more usually from about 100 to 200 μm .

Control nucleic acid members may be present on the array including nucleic acid members comprising oligonucleotides or nucleic acids corresponding to genomic DNA,
10 housekeeping genes, vector sequences, plant nucleic acid sequence, negative and positive control genes, and the like. Control nucleic acid members are calibrating or control genes whose function is not to tell whether a particular "key" gene of interest is expressed, but rather to provide other useful information, such as background or basal level of expression.

Other control nucleic acids are spotted on the array and used as target expression control
15 nucleic acids and mismatch control nucleotides to monitor non-specific binding or cross-hybridization to a nucleic acid in the sample other than the target to which the probe is directed. Mismatch probes thus indicate whether a hybridization is specific or not. For example, if the target is present, the perfectly matched probes should be consistently brighter than the mismatched probes. In addition, if all control mismatches are present, the mismatch probes are
20 used to detect a mutation.

Use of a Microarray

Nucleic acid arrays according to the invention can be used to assay nucleic acids in a sample comprising one or more target nucleic acid sequences (ie such as RNA products of the biomarkers of the invention). The arrays of the subject invention can be used for testing,
25 screening, and/or diagnosis of one or more colorectal pathologies including polyps or one or more subtypes of polyps, or screening for therapeutic targets and the like.

The arrays are also useful in broad scale expression screening for drug discovery and research, such as the effect of a particular active agent on the expression pattern of biomarkers of the invention, where such information is used to reveal drug efficacy and toxicity, environmental
30 monitoring, disease research and the like.

Arrays can be made using at least one, more preferably a combination of these sequences, as a means of diagnosing colon pathology or one or more subtypes of colon pathology.

The choice of a standard sample would be well understood by a person skilled in the art, and would include a sample complementary to RNA isolated from one or more normal
5 individuals, wherein a normal individual is an individual not having polyps.

Preparation of Nucleic Acid Sample for Hybridization to an Array

The samples for hybridization with the arrays according to the invention are in some embodiments derived from total RNA from blood. In another embodiment, targets for the arrays are derived from mRNA from blood.

10 The nucleic acid sample is capable of binding to a nucleic acid member of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation.

As used herein, a “nucleic acid derived from an mRNA transcript: or a “nucleic acid corresponding to an mRNA” refers to a nucleic acid for which synthesis of the mRNA transcript
15 or a sub-sequence thereof has ultimately served as a template. Thus, a cDNA reverse transcribed from an mRNA, an RNA transcribed from that cDNA, a DNA amplified from the cDNA, an RNA transcribed from the amplified DNA, etc., are all derived from or correspond to the mRNA transcript and detection of such derived or corresponding products is indicative of or proportional to the presence and/or abundance of the original transcript in a sample. Thus,
20 suitable nucleic acid samples include, but are not limited to, mRNA transcripts of a gene or genes, cDNA reverse transcribed from the mRNA, cRNA transcribed from the cDNA, DNA amplified from a gene or genes, RNA transcribed from amplified DNA, and the like. The nucleic acid samples used herein are in some embodiments derived from blood. Nucleic acids can be single- or double-stranded DNA, RNA, or DNA-RNA hybrids synthesized from human blood
25 using methods known in the art, for example, reverse transcription or PCR.

In the simplest embodiment, such a nucleic acid sample comprises total mRNA or a nucleic acid sample corresponding to mRNA (e.g., cDNA) isolated from blood samples. In another embodiment, total mRNA is isolated from a given sample using, for example, an acid guanidinium-phenol-chloroform extraction method and polyA+ mRNA is isolated by oligo dT
30 column chromatography or by using (dT)_n magnetic beads (see, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual* (2nd ed.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989), or

Current Protocols in Molecular Biology, F. Ausubel et al., ed. Greene Publishing and Wiley-Interscience, New York (1987). In a preferred embodiment, total RNA is extracted using TRIzol® reagent (GIBCO/BRL, Invitrogen Life Technologies, Cat. No. 15596). Purity and integrity of RNA is assessed by absorbance at 260/280nm and agarose gel electrophoresis
5 followed by inspection under ultraviolet light.

In some embodiments, it is desirable to amplify the nucleic acid sample prior to hybridization, for example, when only limited amounts of sample can be used (e.g. drop of blood). One of skill in the art will appreciate that whatever amplification method is used, if a quantitative result is desired, care must be taken to use a method that maintains or controls for
10 the relative frequencies of the amplified nucleic acids. Methods of "quantitative" amplification are well known to those of skill in the art. For example, quantitative PCR involves simultaneously co-amplifying a known quantity of a control sequence using the same primers. This provides an internal standard that may be used to calibrate the PCR reaction. The high density array may then include probes specific to the internal standard for quantification of the
15 amplified nucleic acid. Detailed protocols for quantitative PCR are provided in *PCR Protocols, A Guide to Methods and Applications*, Innis et al., Academic Press, Inc. N.Y., (1990).

Other suitable amplification methods include, but are not limited to polymerase chain reaction (PCR) (Innis, et al., *PCR Protocols. A Guide to Methods and Application*. Academic Press, Inc. San Diego, (1990)), ligase chain reaction (LCR) (see Wu and Wallace, 1989,
20 *Genomics*, 4:560; Landegren, et al., 1988, *Science*, 241:1077 and Barringer, et al., 1990, *Gene*, 89:117, transcription amplification (Kwoh, et al., 1989, *Proc. Natl. Acad. Sci. USA*, 86: 1173), and self-sustained sequence replication (Guatelli, et al., 1990, *Proc. Nat. Acad. Sci. USA*, 87: 1874).

In a particularly preferred embodiment, the nucleic acid sample mRNA is reverse
25 transcribed with a reverse transcriptase and a primer consisting of oligo dT and a sequence encoding the phage T7 promoter to provide single-stranded DNA template. The second DNA strand is polymerized using a DNA polymerase. After synthesis of double-stranded cDNA, T7 RNA polymerase is added and RNA is transcribed from the cDNA template. Successive rounds of transcription from each single cDNA template results in amplified RNA. Methods of *in vitro*
30 transcription are well known to those of skill in the art (see, e.g., Sambrook, *supra.*) and this particular method is described in detail by Van Gelder, et al., 1990, *Proc. Natl. Acad. Sci. USA*,

87: 1663-1667 who demonstrate that *in vitro* amplification according to this method preserves the relative frequencies of the various RNA transcripts. Moreover, Eberwine et al. *Proc. Natl. Acad. Sci. USA*, 89: 3010-3014 provide a protocol that uses two rounds of amplification via *in vitro* transcription to achieve greater than 10^6 fold amplification of the original starting material
5 thereby permitting expression monitoring even where biological samples are limited.

Labeling of Nucleic Acid Sample or Nucleic Acid Probe

Nucleic acid samples are labelled so as to allow detection of hybridization to an array of the invention. Any analytically detectable marker that is attached to or incorporated into a molecule may be used in the invention. An analytically detectable marker refers to any
10 molecule, moiety or atom which is analytically detected and quantified.

Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., DynabeadsTM), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Patents teaching the use of such labels include
15 U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and
20 4,366,241, the entireties of which are incorporated by reference herein.

Means of detecting such labels are well known to those of skill in the art. Thus, for example, radiolabels may be detected using photographic film or scintillation counters, fluorescent markers may be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction
25 product produced by the action of the enzyme on the substrate, and colorimetric labels are detected by simply visualizing the colored label.

The labels may be incorporated by any of a number of means well known to those of skill in the art. However, in one embodiment, the label is simultaneously incorporated during the amplification step in the preparation of the sample nucleic acids. Thus, for example, polymerase
30 chain reaction (PCR) with labeled primers or labeled nucleotides will provide a labeled amplification product. In a preferred embodiment, transcription amplification, as described

above, using a labeled nucleotide (e.g. fluorescein-labeled UTP and/or CTP) incorporates a label into the transcribed nucleic acids.

Alternatively, a label may be added directly to the original nucleic acid sample (e.g., mRNA, polyA mRNA, cDNA, etc.) or to the amplification product after the amplification is
5 completed. Means of attaching labels to nucleic acids are well known to those of skill in the art and include, for example, nick translation or end-labeling (e.g. with a labeled RNA) by kinasing of the nucleic acid and subsequent attachment (ligation) of a nucleic acid linker joining the sample nucleic acid to a label (e.g., a fluorophore).

In another embodiment, the fluorescent modifications are by cyanine dyes e.g. Cy-3/Cy-5
10 dUTP, Cy-3/Cy-5 dCTP (Amersham Pharmacia) or alexa dyes (Khan, et al., 1998, *Cancer Res.* 58:5009-5013).

In one embodiment, the two Nucleic Acid Sample samples used for comparison are labeled with different fluorescent dyes which produce distinguishable detection signals, for example, nucleic acid samples made from normal intestinal cells are labeled with Cy5 and
15 nucleic acid samples made from intestinal tissue cells are labeled with Cy3. The differently labeled target samples are hybridized to the same microarray simultaneously. In a preferred embodiment, the labeled nucleic acid samples are purified using methods known in the art, e.g., by ethanol purification or column purification.

In another embodiment, the nucleic acid samples will include one or more control
20 molecules which hybridize to control probes on the microarray to normalize signals generated from the microarray. In one embodiment, labeled normalization nucleic acid samples are nucleic acid sequences that are perfectly complementary to control oligonucleotides that are spotted onto the microarray as described above. In another embodiment, labeled normalization nucleic acid samples are nucleic acid sequences that are 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80% or 75%
25 complementary to control oligonucleotides that are spotted onto the microarray as described above. The signals obtained from the normalization controls after hybridization provide a control for variations in hybridization conditions, label intensity, "reading" efficiency and other factors that may cause the signal of a perfect hybridization to vary between arrays. In one embodiment, signals (e.g., fluorescence intensity) read from all other probes in the array are
30 divided by the signal (e.g., fluorescence intensity) from the control probes, thereby normalizing the measurements.

Preferred normalization nucleic acid samples are selected to reflect the average length of the other nucleic acid samples present in the sample, however, they are selected to cover a range of lengths. The normalization control(s) also can be selected to reflect the (average) base composition of the other probes in the array, however, in one embodiment, only one or a few
5 normalization probes are used and they are selected such that they hybridize well (i.e., have no secondary structure and do not self hybridize) and do not match any nucleic acids on the array.

Normalization probes are localized at any position in the array or at multiple positions throughout the array to control for spatial variation in hybridization efficiency. In one embodiment, normalization controls are located at the corners or edges of the array as well as in
10 the middle.

Hybridization Conditions

Nucleic acid hybridization involves providing a nucleic acid sample under conditions where the sample and the complementary nucleic acid member can form stable hybrid duplexes through complementary base pairing. The nucleic acids that do not form hybrid duplexes are
15 then washed away leaving the hybridized nucleic acids to be detected, typically through detection of an attached detectable label. It is generally recognized that nucleic acids are denatured by increasing the temperature or decreasing the salt concentration of the buffer containing the nucleic acids. Under low stringency conditions (e.g., low temperature and/or high salt) hybrid duplexes (e.g., DNA:DNA, RNA:RNA, or RNA:DNA) will form even where the
20 annealed sequences are not perfectly complementary. Thus specificity of hybridization is reduced at lower stringency. Conversely, at higher stringency (e.g., higher temperature or lower salt) successful hybridization requires fewer mismatches.

The invention provides for hybridization conditions comprising the Dig hybridization mix (Boehringer); or formamide-based hybridization solutions, for example as described in Ausubel
25 et al., *supra* and Sambrook et al. *supra*.

Methods of optimizing hybridization conditions are well known to those of skill in the art (see, e.g., *Laboratory Techniques in Biochemistry and Molecular Biology*, Vol. 24: *Hybridization With Nucleic acid Probes*, P. Tijssen, ed. Elsevier, N.Y., (1993)).

Following hybridization, non-hybridized labeled or unlabeled nucleic acid is removed
30 from the support surface, conveniently by washing, thereby generating a pattern of hybridized target nucleic acid on the substrate surface. A variety of wash solutions are known to those of

skill in the art and may be used. The resultant hybridization patterns of labeled, hybridized oligonucleotides and/or nucleic acids may be visualized or detected in a variety of ways, with the particular manner of detection being chosen based on the particular label of the test nucleic acid, where representative detection means include scintillation counting, autoradiography, fluorescence measurement, calorimetric measurement, light emission measurement and the like.

Image Acquisition and Data Analysis

Following hybridization and any washing step(s) and/or subsequent treatments, as described above, the resultant hybridization pattern is detected. In detecting or visualizing the hybridization pattern, the intensity or signal value of the label will be not only be detected but quantified, by which is meant that the signal from each spot of the hybridization will be measured and compared to a unit value corresponding to the signal emitted by a known number of end labeled target nucleic acids to obtain a count or absolute value of the copy number of each end-labeled target that is hybridized to a particular spot on the array in the hybridization pattern.

Methods for analyzing the data collected from hybridization to arrays are well known in the art. For example, where detection of hybridization involves a fluorescent label, data analysis can include the steps of determining fluorescent intensity as a function of substrate position from the data collected, removing outliers, i.e., data deviating from a predetermined statistical distribution, and calculating the relative binding affinity of the test nucleic acids from the remaining data. The resulting data is displayed as an image with the intensity in each region varying according to the binding affinity between associated oligonucleotides and/or nucleic acids and the test nucleic acids.

The following detection protocol is used for the simultaneous analysis of two samples to be compared, where each sample is labeled with a different fluorescent dye.

Each element of the microarray is scanned for the first fluorescent color. The intensity of the fluorescence at each array element is proportional to the expression level of that gene in the sample. The scanning operation is repeated for the second fluorescent label. The ratio of the two fluorescent intensities provides a highly accurate and quantitative measurement of the relative gene expression level in the two samples.

In a preferred embodiment, fluorescence intensities of immobilized nucleic acid sequences were determined from images taken with a custom confocal microscope equipped with laser excitation sources and interference filters appropriate for the Cy3 and Cy5 fluors. Separate scans

were taken for each fluor at a resolution of $225\ \mu\text{m}^2$ per pixel and 65,536 gray levels. Image segmentation to identify areas of hybridization, normalization of the intensities between the two fluor images, and calculation of the normalized mean fluorescent values at each target are as described (Khan, et al., 1998, *Cancer Res.* 58:5009-5013. Chen, et al., 1997, *Biomed. Optics* 2:364-374). Normalization between the images is used to adjust for the different efficiencies in labeling and detection with the two different fluors. This is achieved by equilibrating to a value of one the signal intensity ratio of a set of internal control genes spotted on the array.

In another preferred embodiment, the array is scanned in the Cy3 and Cy5 channels and stored as separate 16-bit TIFF images. The images are incorporated and analysed using software which includes a gridding process to capture the hybridization intensity data from each spot on the array. The fluorescence intensity and background-subtracted hybridization intensity of each spot is collected and a ratio of measured mean intensities of Cy5 to Cy3 is calculated. A linear regression approach is used for normalization and assumes that a scatter plot of the measured Cy5 versus Cy3 intensities should have a slope of one. The average of the ratios is calculated and used to rescale the data and adjust the slope to one. A ratio of expression not equal to 1 is used as an indication of differential gene expression.

In a particular embodiment, where it is desired to quantify the transcription level (and thereby expression) of one or more nucleic acid sequences in a sample, the nucleic acid sample is one in which the concentration of the mRNA transcript(s) of the gene or genes, or the concentration of the nucleic acids derived from the mRNA transcript(s), is proportional to the transcription level (and therefore expression level) of that gene. Similarly, it is preferred that the hybridization signal intensity be proportional to the amount of hybridized nucleic acid. While it is preferred that the proportionality be relatively strict (e.g., a doubling in transcription rate results in a doubling in mRNA transcript in the sample nucleic acid pool and a doubling in hybridization signal), one of skill will appreciate that the proportionality can be more relaxed and even non-linear and still provide meaningful results. Thus, for example, an assay where a 5 fold difference in concentration of the sample mRNA results in a 3- to 6-fold difference in hybridization intensity is sufficient for most purposes. Where more precise quantification is required, appropriate controls are run to correct for variations introduced in sample preparation and hybridization as described herein. In addition, serial dilutions of "standard" mRNA samples are used to prepare calibration curves according to methods well known to those of skill in the

art. Of course, where simple detection of the presence or absence of a transcript is desired, no elaborate control or calibration is required.

For example, if a nucleic acid member on an array is not labeled after hybridization, this indicates that the gene comprising that nucleic acid member is not expressed in either sample. If
5 a nucleic acid member is labeled with a single color, it indicates that a labeled gene was expressed only in one sample. The labeling of a nucleic acid member comprising an array with both colors indicates that the gene was expressed in both samples. Even genes expressed once per cell are detected (1 part in 100,000 sensitivity). A difference in expression intensity in the two samples being compared is indicative of differential expression, the ratio of the intensity in
10 the two samples being not equal to 1.0, greater than 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 2.0, 3.0, 4.0 and the like or less than 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 and the like.

PCR

In one aspect of the invention, the level of the expression of the RNA products of the biomarkers of the invention can be measured by amplifying the RNA products of the biomarkers
15 from a sample by first using reverse transcription (RT). Either in combination, or as a second reaction step, the reverse transcribed product can then be amplified with the polymerase chain reaction (PCR). In accordance with one embodiment of the invention, the PCR can be QRT-PCR as would be understood to a person skilled in the art.

Total RNA, or mRNA from a sample is used as a template and a primer specific to the
20 transcribed portion of a biomarker of the invention is used to initiate reverse transcription. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook et al., 1989, *supra*. Primer design can be accomplished utilizing commercially available software (e.g., Primer Designer 1.0, Scientific Software *etc.*). The product of the reverse transcription is subsequently used as a template for PCR.

25 PCR provides a method for rapidly amplifying a particular nucleic acid sequence by using multiple cycles of DNA replication catalyzed by a thermostable, DNA-dependent DNA polymerase to amplify the target sequence of interest. PCR requires the presence of a nucleic acid to be amplified, two single-stranded oligonucleotide primers flanking the sequence to be amplified, a DNA polymerase, deoxyribonucleoside triphosphates, a buffer and salts.

30 The method of PCR is well known in the art. PCR, is performed as described in Mullis and Faloona, 1987, *Methods Enzymol.*, 155: 335, which is incorporated herein by reference.

PCR is performed using template DNA (at least 1fg; more usefully, 1-1000 ng) and at least 25 pmol of oligonucleotide primers. A typical reaction mixture includes: 2 ml of DNA, 25 pmol of oligonucleotide primer, 2.5 ml of 10H PCR buffer 1 (Perkin-Elmer, Foster City, CA), 0.4 ml of 1.25 mM dNTP, 0.15 ml (or 2.5 units) of Taq DNA polymerase (Perkin Elmer, Foster City, CA) and deionized water to a total volume of 25 ml. Mineral oil is overlaid and the PCR is performed using a programmable thermal cycler.

The length and temperature of each step of a PCR cycle, as well as the number of cycles, are adjusted according to the stringency requirements in effect. Annealing temperature and timing are determined both by the efficiency with which a primer is expected to anneal to a template and the degree of mismatch that is to be tolerated. The ability to optimize the stringency of primer annealing conditions is well within the knowledge of one of moderate skill in the art. An annealing temperature of between 30 °C and 72 °C is used. Initial denaturation of the template molecules normally occurs at between 92 °C and 99 °C for 4 minutes, followed by 20-40 cycles consisting of denaturation (94-99°C for 15 seconds to 1 minute), annealing (temperature determined as discussed above; 1-2 minutes), and extension (72 °C for 1 minute). The final extension step is generally carried out for 4 minutes at 72 °C, and may be followed by an indefinite (0-24 hour) step at 4 °C.

QRT-PCR (Quantitative real time RT-PCR), can also be performed to provide a quantitative measure of gene expression levels. Similar to reverse transcription PCR, QRT-PCR reverse transcription and PCR can be performed in two steps, or reverse transcription combined with PCR can be performed concurrently. One of these techniques, for which there are commercially available kits such as Taqman (Perkin Elmer, Foster City, CA), is performed with a transcript-specific antisense probe. This probe is specific for the PCR product (e.g. a nucleic acid fragment derived from a gene) and is prepared with a quencher and fluorescent reporter probe complexed to the 5' end of the oligonucleotide. Different fluorescent markers are attached to different reporters, allowing for measurement of two products in one reaction. When Taq DNA polymerase is activated, it cleaves off the fluorescent reporters of the probe bound to the template by virtue of its 5'-to-3' exonuclease activity. In the absence of the quenchers, the reporters now fluoresce. The color change in the reporters is proportional to the amount of each specific product and is measured by a fluorometer; therefore, the amount of each color is measured and the PCR product is quantified. The PCR reactions can be performed in 96 well

plates, 384 well plates and the like so that samples derived from many individuals are processed and measured simultaneously. The Taqman system has the additional advantage of not requiring gel electrophoresis and allows for quantification when used with a standard curve.

5 A second technique useful for detecting PCR products quantitatively without is to use an intercalating dye such as the commercially available QuantiTect SYBR Green PCR (Qiagen, Valencia California). QRT-PCR is performed using SYBR green as a fluorescent label which is incorporated into the PCR product during the PCR stage and produces a fluorescence proportional to the amount of PCR product.

10 Both Taqman and QuantiTect SYBR systems can be used subsequent to reverse transcription of RNA. Reverse transcription can either be performed in the same reaction mixture as the PCR step (one-step protocol) or reverse transcription can be performed first prior to amplification utilizing PCR (two-step protocol).

15 Additionally, other systems to quantitatively measure mRNA expression products are known including Molecular Beacons® which uses a probe having a fluorescent molecule and a quencher molecule, the probe capable of forming a hairpin structure such that when in the hairpin form, the fluorescence molecule is quenched, and when hybridized the fluorescence increases giving a quantitative measurement of gene expression.

20 Additional techniques to quantitatively measure RNA expression include, but are not limited to, polymerase chain reaction, ligase chain reaction, Qbeta replicase (see, *e.g.*, International Application No. PCT/US87/00880), isothermal amplification method (see, *e.g.*, Walker et al. (1992) PNAS 89:382-396), strand displacement amplification (SDA), repair chain reaction, Asymmetric Quantitative PCR (see, *e.g.*, U.S. Publication No. US20030134307A1) and the multiplex microsphere bead assay described in Fuja et al., 2004, Journal of Biotechnology 108:193-205.

25 The level of gene expression can be measured by amplifying RNA from a sample using transcription based amplification systems (TAS), including nucleic acid sequence amplification (NASBA) and 3SR. See, *e.g.*, Kwoh et al (1989) PNAS USA 86:1173; International Publication No. WO 88/10315; and U.S. Patent No. 6,329,179. In NASBA, the nucleic acids may be prepared for amplification using conventional phenol/chloroform extraction, heat denaturation, 30 treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer that has

target specific sequences. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target specific primer, followed by polymerization. The double-stranded DNA molecules are then multiply transcribed
5 by a polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNA's are reverse transcribed into double stranded DNA, and transcribed once with a polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target specific sequences.

Several techniques may be used to separate amplification products. For example, amplification products may be separated by agarose, agarose-acrylamide or polyacrylamide gel
10 electrophoresis using conventional methods. See Sambrook et al., 1989. Several techniques for detecting PCR products quantitatively without electrophoresis may also be used according to the invention (see for example *PCR Protocols, A Guide to Methods and Applications*, Innis et al., Academic Press, Inc. N.Y., (1990)). For example, chromatographic techniques may be employed to effect separation. There are many kinds of chromatography which may be used in
15 the present invention: adsorption, partition, ion-exchange and molecular sieve, HPLC, and many specialized techniques for using them including column, paper, thin-layer and gas chromatography (Freifelder, *Physical Biochemistry Applications to Biochemistry and Molecular Biology*, 2nd ed., Wm. Freeman and Co., New York, N.Y., 1982).

Another example of a separation methodology is done by covalently labeling the
20 oligonucleotide primers used in a PCR reaction with various types of small molecule ligands. In one such separation, a different ligand is present on each oligonucleotide. A molecule, perhaps an antibody or avidin if the ligand is biotin, that specifically binds to one of the ligands is used to coat the surface of a plate such as a 96 well ELISA plate. Upon application of the PCR reactions to the surface of such a prepared plate, the PCR products are bound with specificity to the
25 surface. After washing the plate to remove unbound reagents, a solution containing a second molecule that binds to the first ligand is added. This second molecule is linked to some kind of reporter system. The second molecule only binds to the plate if a PCR product has been produced whereby both oligonucleotide primers are incorporated into the final PCR products. The amount of the PCR product is then detected and quantified in a commercial plate reader
30 much as ELISA reactions are detected and quantified. An ELISA-like system such as the one

described here has been developed by the Raggio Italgene company under the C-Track trade name.

Amplification products must be visualized in order to confirm amplification of the nucleic acid sequences of interest. One typical visualization method involves staining of a gel with
5 ethidium bromide and visualization under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the amplification products may then be exposed to x-ray film or visualized under the appropriate stimulating spectra, following separation.

In one embodiment, visualization is achieved indirectly. Following separation of
10 amplification products, a labeled, nucleic acid probe is brought into contact with the amplified nucleic acid sequence of interest. The probe in one embodiment is conjugated to a chromophore but may be radiolabeled. In another embodiment, the probe is conjugated to a binding partner, such as an antibody or biotin, where the other member of the binding pair carries a detectable moiety.

15 In another embodiment, detection is by Southern blotting and hybridization with a labeled probe. The techniques involved in Southern blotting are well known to those of skill in the art and may be found in many standard books on molecular protocols. See Sambrook et al., 1989, *supra*. Briefly, amplification products are separated by gel electrophoresis. The gel is then contacted with a membrane, such as nitrocellulose, permitting transfer of the nucleic acid and
20 non-covalent binding. Subsequently, the membrane is incubated with a chromophore-conjugated probe that is capable of hybridizing with a target amplification product. Detection is by exposure of the membrane to x-ray film or ion-emitting detection devices.

One example of the foregoing is described in U.S. Patent No. 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and
25 transfer of nucleic acids. The apparatus permits electrophoresis and blotting without external manipulation of the gel and is ideally suited to carrying out methods according to the present invention.

One embodiment of the invention includes the primers and probes in **Table 4, 6, 16 or 17** can be used for use in measuring the expression of the biomarkers of the invention.

Nuclease Protection Assays

In another embodiment of the invention, Nuclease protection assays (including both ribonuclease protection assays and S1 nuclease assays) can be used to detect and quantitate the RNA products of the biomarkers of the invention. In nuclease protection assays, an antisense probe (labeled with, *e.g.*, radiolabeled or nonisotopic) hybridizes in solution to an RNA sample. Following hybridization, single-stranded, unhybridized probe and RNA are degraded by nucleases. An acrylamide gel is used to separate the remaining protected fragments. Typically, solution hybridization is more efficient than membrane-based hybridization, and it can accommodate up to 100 μ g of sample RNA, compared with the 20-30 μ g maximum of blot hybridizations.

The ribonuclease protection assay, which is the most common type of nuclease protection assay, requires the use of RNA probes. Oligonucleotides and other single-stranded DNA probes can only be used in assays containing S1 nuclease. The single-stranded, antisense probe must typically be completely homologous to target RNA to prevent cleavage of the probe:target hybrid by nuclease.

Northern Blots

A standard Northern blot assay can also be used to ascertain an RNA transcript size, identify alternatively spliced RNA transcripts, and the relative amounts of RNA products of the biomarker of the invention, in accordance with conventional Northern hybridization techniques known to those persons of ordinary skill in the art. In Northern blots, RNA samples are first separated by size via electrophoresis in an agarose gel under denaturing conditions. The RNA is then transferred to a membrane, crosslinked and hybridized with a labeled probe. Nonisotopic or high specific activity radiolabeled probes can be used including random-primed, nick-translated, or PCR-generated DNA probes, in vitro transcribed RNA probes, and oligonucleotides. Additionally, sequences with only partial homology (*e.g.*, cDNA from a different species or genomic DNA fragments that might contain an exon) may be used as probes. The labeled probe, *e.g.*, a radiolabelled cDNA, either containing the full-length, single stranded DNA or a fragment of that DNA sequence may be at least 20, at least 30, at least 50, or at least 100 consecutive nucleotides in length. The probe can be labeled by any of the many different methods known to those skilled in this art. The labels most commonly employed for these studies are radioactive elements, enzymes, chemicals that fluoresce when exposed to ultraviolet light, and others. A

number of fluorescent materials are known and can be utilized as labels. These include, but are not limited to, fluorescein, rhodamine, auramine, Texas Red, AMCA blue and Lucifer Yellow. A particular detecting material is anti-rabbit antibody prepared in goats and conjugated with fluorescein through an isothiocyanate. Proteins can also be labeled with a radioactive element or
5 with an enzyme. The radioactive label can be detected by any of the currently available counting procedures. Non-limiting examples of isotopes include ^3H , ^{14}C , ^{32}P , ^{35}S , ^{36}Cl , ^{51}Cr , ^{57}Co , ^{58}Co , ^{59}Fe , ^{90}Y , ^{125}I , ^{131}I , and ^{186}Re . Enzyme labels are likewise useful, and can be detected by any of the presently utilized colorimetric, spectrophotometric, fluorospectrophotometric, amperometric or gasometric techniques. The enzyme is conjugated to the selected particle by reaction with
10 bridging molecules such as carbodiimides, diisocyanates, glutaraldehyde and the like. Any enzymes known to one of skill in the art can be utilized. Examples of such enzymes include, but are not limited to, peroxidase, beta-D-galactosidase, urease, glucose oxidase plus peroxidase and alkaline phosphatase. U.S. Patent Nos. 3,654,090, 3,850,752, and 4,016,043 are referred to by way of example for their disclosure of alternate labeling material and methods.

15 **(N) Techniques to Measure the Protein Products of the Biomarkers of the Invention**
Antibody Based Methodologies

Standard techniques can also be utilized for determining the amount of the protein or proteins of interest present in a sample. For example, standard techniques can be employed using, *e.g.*, immunoassays such as, for example, Western blot, immunoprecipitation followed by
20 sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), immunocytochemistry, and the like to determine the amount of the protein or proteins of interest present in a sample. A preferred agent for detecting a protein of interest is an antibody capable of binding to a protein of interest, in one embodiment an antibody with a detectable label.

For such detection methods, protein from the sample to be analyzed can easily be isolated
25 using techniques which are well known to those of skill in the art. Protein isolation methods can, for example, be such as those described in Harlow and Lane (Harlow, E. and Lane, D., *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1988)).

In some embodiments, methods for the detection of the protein or proteins of interest
30 involve their detection via interaction with a protein-specific antibody. For example, antibodies directed a protein of interest can be utilized as described herein. Antibodies can be generated

utilizing standard techniques well known to those of skill in the art. See, *e.g.*, Section 15.13.2 of this application and Section 5.2 of U.S. Publication No. 20040018200 for a more detailed discussion of such antibody generation techniques, which is incorporated herein by reference. Briefly, such antibodies can be polyclonal, or monoclonal. An intact antibody, or an antibody
5 fragment (*e.g.*, Fab or F(ab')₂) can, for example, be used. In some embodiments, the antibody is a human or humanized antibody.

Table 5 and **Table 15** are tables showing, in one embodiment of the invention, antibodies which are used to detect the proteins of the biomarkers of the invention.

For example, antibodies, or fragments of antibodies, specific for a protein of interest can
10 be used to quantitatively or qualitatively detect the presence of the protein. This can be accomplished, for example, by immunofluorescence techniques. Antibodies (or fragments thereof) can, additionally, be employed histologically, as in immunofluorescence or immunoelectron microscopy, for *in situ* detection of a protein of interest. *In situ* detection can be accomplished by removing a histological specimen (*e.g.*, a biopsy specimen) from a patient, and
15 applying thereto a labeled antibody thereto that is directed to a protein. The antibody (or fragment) can be applied by overlaying the labeled antibody (or fragment) onto a biological sample. Through the use of such a procedure, it is possible to determine not only the presence of the protein of interest, but also its distribution, its presence in cells (*e.g.*, intestinal cells and lymphocytes) within the sample. A wide variety of well-known histological methods (such as
20 staining procedures) can be utilized in order to achieve such *in situ* detection.

Immunoassays for a protein of interest typically comprise incubating a biological sample of a detectably labeled antibody capable of identifying a protein of interest, and detecting the bound antibody by any of a number of techniques well-known in the art. As discussed in more detail, below, the term "labeled" can refer to direct labeling of the antibody via, *e.g.*, coupling
25 (*i.e.*, physically linking) a detectable substance to the antibody, and can also refer to indirect labeling of the antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody.

For example, the biological sample can be brought in contact with and immobilized onto
30 a solid phase support or carrier such as nitrocellulose, or other support which is capable of immobilizing cells, cell particles or soluble proteins. The support can then be washed with

suitable buffers followed by treatment with the detectably labeled fingerprint gene-specific antibody. The solid phase support can then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on support can then be detected by conventional means.

5 By "solid phase support or carrier " in the context of proteinaceous agents is intended any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified
10 celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material can have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration can be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface can be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding
15 antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

One of the ways in which a specific antibody can be detectably labeled is by linking the same to an enzyme and use in an enzyme immunoassay (EIA) (Voller, A., "The Enzyme Linked Immunosorbent Assay (ELISA)", 1978, Diagnostic Horizons 2:1-7, Microbiological Associates Quarterly Publication, Walkersville, MD); Voller, A. et al., 1978, J. Clin. Pathol. 31:507-520;
20 Butler, J.E., 1981, Meth. Enzymol. 73:482-523; Maggio, E. (ed.), 1980, *Enzyme Immunoassay*, CRC Press, Boca Raton, FL; Ishikawa, E. et al., (eds.), 1981, *Enzyme Immunoassay*, Kigaku Shoin, Tokyo). The enzyme which is bound to the antibody will react with an appropriate substrate, in one embodiment a chromogenic substrate, in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, fluorimetric or by
25 visual means. Enzymes which can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and
30 acetylcholinesterase. The detection can be accomplished by colorimetric methods which employ a chromogenic substrate for the enzyme. Detection can also be accomplished by visual

comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

Detection can also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect a protein of interest through the use of a radioimmunoassay (RIA) (see, for example, Weintraub, B., Principles of Radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March, 1986, which is incorporated by reference herein). The radioactive isotope (*e.g.*, ^{125}I , ^{131}I , ^{35}S or ^3H) can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography.

It is also possible to label the antibody with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wavelength, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthalaldehyde and fluorescamine.

The antibody can also be detectably labeled using fluorescence emitting metals such as ^{152}Eu , or others of the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylenetriaminepentacetic acid (DTPA) or ethylenediamine-tetraacetic acid (EDTA).

The antibody also can be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful chemiluminescent labeling compounds are luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound can be used to label the antibody of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems in, which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

Protein Arrays

Polypeptides which specifically and/or selectively bind to the protein products of the biomarkers of the invention can be immobilized on a protein array. The protein array can be used as a tool, *e.g.*, to test individual samples (such as isolated cells, tissue, lymph, lymph tissue, blood, synovial fluid, sera, biopsies, and the like) for the presence of the polypeptides protein products of the biomarkers of the invention. The protein array can also include antibodies as well as other ligands, *e.g.*, that bind to the polypeptides encoded by the biomarkers of the invention.

Methods of producing polypeptide arrays are described, *e.g.*, in De Wildt et al., 2000, Nature Biotech. 18:989-994; Lueking et al., 1999, Anal. Biochem. 270:103-111; Ge, 2000, Nuc. Acids Res. 28:e3; MacBeath and Schreiber, 2000, Science 289:1760-1763; International Publication Nos. WO 01/40803 and WO 99/51773A1; and U.S. Patent No. 6,406,921. Polypeptides for the array can be spotted at high speed, *e.g.*, using commercially available robotic apparatus, *e.g.*, from Genetic MicroSystems and Affymetrix (Santa Clara, California, USA) or BioRobotics (Cambridge, UK). The array substrate can be, for example, nitrocellulose, plastic, glass, *e.g.*, surface-modified glass. The array can also include a porous matrix, *e.g.*, acrylamide, agarose, or another polymer.

For example, the array can be an array of antibodies, *e.g.*, as described in De Wildt, *supra*. Cells that produce the polypeptide ligands can be grown on a filter in an arrayed format. Polypeptide production is induced, and the expressed antibodies are immobilized to the filter at the location of the cell. Information about the extent of binding at each address of the array can be stored as a profile, *e.g.*, in a computer database.

In one embodiment the array is an array of protein products of the 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 or all or any combination of the biomarkers of the invention. In one aspect, the invention provides for antibodies that are bound to an array which selectively bind to the protein products of the biomarkers of the invention.

(O) Protein Production

Standard recombinant nucleic acid methods can be used to express a polypeptide or antibody of the invention (*e.g.*, a protein products of a biomarker of the invention). Generally, a nucleic acid sequence encoding the polypeptide is cloned into a nucleic acid expression vector. Of course, if the protein includes multiple polypeptide chains, each chain must be cloned into an expression vector, *e.g.*, the same or different vectors, that are expressed in the same or different

cells. If the protein is sufficiently small, *i.e.*, the protein is a peptide of less than 50 amino acids, the protein can be synthesized using automated organic synthetic methods. Polypeptides comprising the 5' region, 3' region or internal coding region of a biomarker of the invention, are expressed from nucleic acid expression vectors containing only those nucleotide sequences
5 corresponding to the 5' region, 3' region or internal coding region of a biomarker of the invention. Methods for producing antibodies directed to protein products of a biomarker of the invention, or polypeptides encoded by the 5' region, 3' region or internal coding regions of a biomarker of the invention.

The expression vector for expressing the polypeptide can include, in addition to the
10 segment encoding the polypeptide or fragment thereof, regulatory sequences, including for example, a promoter, operably linked to the nucleic acid(s) of interest. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK,
15 pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene, La Jolla, California, USA); pTrc99A, pKK223-3, pKK233-3, pDR540, and pRIT5 (Pharmacia, Uppsala, Sweden). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia). One preferred class of preferred libraries is the display library, which is described below.

20 Methods well known to those skilled in the art can be used to construct vectors containing a polynucleotide of the invention and appropriate transcriptional/translational control signals. These methods include *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination. See, for example, the techniques described in Sambrook & Russell, *Molecular Cloning: A Laboratory Manual*, 3rd Edition, Cold Spring
25 Harbor Laboratory, N.Y. (2001) and Ausubel *et al.*, *Current Protocols in Molecular Biology* (Greene Publishing Associates and Wiley Interscience, N.Y. (1989). Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P, and trc. Eukaryotic promoters
30 include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, mouse metallothionein-I, and various art-known tissue specific promoters. In specific

embodiments, the promoter is an inducible promoter. In other embodiments, the promoter is a constitutive promoter. In yet other embodiments, the promoter is a tissue-specific promoter.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* auxotrophic markers (such as *URA3*, *LEU2*, *HIS3*, and *TRP1* genes), and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), *a-factor*, acid phosphatase, or heat shock proteins, among others. The polynucleotide of the invention is assembled in appropriate phase with translation initiation and termination sequences, and in some embodiments, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, a nucleic acid of the invention can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression-vectors for bacteria are constructed by inserting a polynucleotide of the invention together with suitable translation initiation and termination signals, optionally in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacteria can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and pGEM1 (Promega, Madison, Wisconsin, USA).

The present invention provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention also provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with

a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

The present invention further provides host cells containing the vectors of the present invention, wherein the nucleic acid has been introduced into the host cell using known
5 transformation, transfection or infection methods. The host cell can be a eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected, for example, by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. *et al.*, *Basic Methods in Molecular Biology*
10 (1986)). Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

Any host/vector system can be used to express one or more of the proteins products of the biomarkers of the invention including those listed in **Table 3** and/or **Table 13**. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by
15 Sambrook *et al.*, in *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is incorporated herein by reference in its entirety. The most preferred host cells are those which do not normally express the particular polypeptide or which expresses the polypeptide at low natural level.

In a specific embodiment, the host cells are engineered to express an endogenous gene
20 comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences
25 may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting, including polyadenylation signals. mRNA stability elements, splice sites, leader sequences for enhancing
30 or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The host of the present invention may also be a yeast or other fungi. In yeast, a number of vectors containing constitutive or inducible promoters may be used. For a review see, Ausubel et al. (eds), *Current Protocols in Molecular Biology*, Vol. 2, Greene Publish. Assoc. & Wiley Interscience, Ch. 13 (1988); Grant et al., 1987, "Expression and Secretion Vectors for
5 Yeast", *Methods Enzymol.* 153:516-544; Glover, *DNA Cloning*, Vol. II, IRL Press, Wash., D.C., Ch. 3 (1986); Bitter, 1987, "Heterologous Gene Expression in Yeast", *Methods Enzymol.* 152:673-684; and Strathern et al. (eds), *The Molecular Biology of the Yeast Saccharomyces*, Cold Spring Harbor Press, Vols. I and II (1982).

Potentially suitable yeast strains include *Saccharomyces cerevisiae*,
10 *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli*, enterobacteriaceae such as *Serratia marescans*, bacilli such as *Bacillus subtilis*, *Salmonella typhimurium*, pseudomonads or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced
15 therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the monkey COS cells such as
20 COS-7 lines of monkey kidney fibroblasts, described by Gluzman, 1981, *Cell* 23:175 (1981), Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK, C127, 3T3, or Jurkat cells, and other cell lines capable of expressing a compatible vector.
25 Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome-binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing
30 agents. Recombinant polypeptides produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size

exclusion chromatography steps. In some embodiments, the template nucleic acid also encodes a polypeptide tag, *e.g.*, penta- or hexa-histidine.

Recombinant proteins can be isolated using a technique well-known in the art. Scopes (*Protein Purification: Principles and Practice*, Springer-Verlag, New York (1994)), for example, provides a number of general methods for purifying recombinant (and non-recombinant) proteins. The methods include, *e.g.*, ion-exchange chromatography, size-exclusion chromatography, affinity chromatography, selective precipitation, dialysis, and hydrophobic interaction chromatography.

(P) Methods for Identifying Compounds for Use in the Prevention, Treatment, or Amelioration of One or More Colorectal Pathologies

Methods to Identify Compounds that Modulate the Expression or Activity of a Biomarker

The present invention provides methods of identifying compounds that bind to the products of the biomarkers of the invention. The present invention also provides methods for identifying compounds that modulate the expression and/or activity of the products of the biomarkers of the invention. The compounds identified via such methods are useful for the development of one or more animal models to study colorectal pathology including polyps or one or more subtypes of polyps. Further, the compounds identified via such methods are useful as lead compounds in the development of prophylactic and therapeutic compositions for prevention, treatment, and/or amelioration of one or more colorectal pathologies including one or more polyps or one or more subtypes of polyps. Such methods are particularly useful in that the effort and great expense involved in testing potential prophylactics and therapeutics *in vivo* is efficiently focused on those compounds identified via the *in vitro* and *ex vivo* methods described herein.

The present invention provides a method for identifying a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies said method comprising: (a) contacting a cell expressing a protein products of one or more biomarkers of the invention or a fragment thereof, or RNA products of one or more biomarkers of the invention or a fragment thereof with a test compound; and (b) determining the ability of the test compound to bind to the protein products, protein fragment, RNA products, or RNA portion so that if a compound binds to the protein products, protein fragment, RNA products, RNA portions, a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal

pathologies is identified. The cell, for example, can be a prokaryotic cell, yeast cell, viral cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the protein products, protein fragment, RNA products, or RNA portion can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the protein products, protein fragment, RNA products, or RNA portion can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In a specific embodiment, the assay comprises contacting a cell which expresses a protein products of one or more biomarkers of the invention or a fragment thereof, or a RNA products of one or more biomarkers of the invention or a fragment thereof, with a known compound which binds the protein products, protein fragment, RNA products, or RNA portion to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the protein products, protein fragment, RNA products, or RNA portion, wherein determining the ability of the test compound to interact with the protein products, protein fragment, RNA products, or RNA portion comprises determining the ability of the test compound to preferentially bind to the protein products, protein fragment, RNA products, or RNA portion as compared to the known compound.

Binding of the test compound to the protein products or protein fragment can be determined either directly or indirectly. In a specific embodiment, the assay includes contacting a protein products of one or more biomarkers of the invention or a fragment thereof, or one or more RNA products of one or more biomarkers of the invention or a portion thereof with a known compound which binds the protein products, protein fragment, RNA products, or RNA portion to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the protein products, protein fragment, RNA products, or RNA portion, wherein determining the ability of the test compound to interact with the protein products, protein fragment, RNA products, or RNA portion comprises determining the ability of the test compound to preferentially bind to the protein products, protein fragment, RNA products, or RNA portion as compared to the known compound.

Techniques well known in the art can be used to determine the binding between a test compound and a protein products of a biomarker of the invention or a fragment thereof, or a RNA products of a biomarker of the invention or a portion thereof.

In some embodiments of the above assay methods of the present invention, it may be
5 desirable to immobilize a RNA products of a biomarker of the invention or a portion thereof, or its target molecule to facilitate separation of complexed from uncomplexed forms of the RNA products or RNA portion, the target molecule or both, as well as to accommodate automation of the assay. In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either a protein products of a biomarker of the invention or a
10 fragment thereof, or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a protein products of a biomarker of the invention or a fragment thereof can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one
15 embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase (GST) fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical; St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or a protein products of a biomarker of
20 the invention or a fragment thereof, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of binding of a
25 protein products of a biomarker of the invention or a fragment thereof can be determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a protein products of a biomarker of the invention or a fragment thereof, or a target molecule can be immobilized utilizing conjugation of biotin and
30 streptavidin. A biotinylated protein products of a biomarker of the invention or a target molecule can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the

art (*e.g.*, biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with a protein products of a biomarker of the invention or a fragment thereof can be derivatized to the wells of the plate, and protein trapped in the wells by antibody conjugation. Methods for
5 detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with a protein products of a biomarker of the invention, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with a protein products of a biomarker of the invention or a fragment thereof, or target molecule.

10 The interaction or binding of a protein products of a biomarker of the invention or a fragment thereof to a test compound can also be determined using such proteins or protein fragments as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J. Biol. Chem. 268:12046-12054; Bartel et al. (1993) Bio/Techniques 14:920-924; Iwabuchi et al. (1993)
15 Oncogene 8:1693-1696; and International Publication No. WO 94/10300).

The present invention provides a method for identifying a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies, said method comprising: (a) contacting a cell expressing a protein or RNA products of one or more biomarkers of the invention with a test compound; (b) determining the amount of the protein or
20 RNA products present in (a); and (c) comparing the amount in (a) to that present in a corresponding control cell that has not been contacted with the test compound, so that if the amount of the protein or RNA products is altered relative to the amount in the control, a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies is identified. In a specific embodiment, the expression level(s) is altered by 5%,
25 10%, 15%, 25%, 30%, 40%, 50%, 5 to 25%, 10 to 30%, at least 1 fold, at least 1.5 fold, at least 2 fold, 4 fold, 5 fold, 10 fold, 25 fold, 1 to 10 fold, or 5 to 25 fold relative to the expression level in the control as determined by utilizing an assay described herein (*e.g.*, a microarray or QRT-PCR) or an assay well known to one of skill in the art. In alternate embodiments, such a method comprises determining the amount of the protein or RNA products of at least 2, at least 3, at least
30 4, at least 5, at least 6, at least 7, at least 8, 1 to 3, 1 to 5, 1-8, all or any combination of the

biomarkers of the invention present in the cell and comparing the amounts to those present in the control.

The cells utilized in the cell-based assays described herein can be engineered to express a biomarker of the invention utilizing techniques known in the art. *See, e.g.*, Section III entitled
5 “Recombinant Expression Vectors and Host Cells” of U.S. Patent No. 6,245,527, which is incorporated herein by reference. Alternatively, cells that endogenously express a biomarker of the invention can be used. For example, intestinal cells may be used.

In a specific embodiment, intestinal cells are isolated from a “normal” individual, or an individual with one or more colorectal pathologies and are incubated in the presence and
10 absence of a test compound for varying amounts of time (*i.e.*, 30 min, 1 hr, 5 hr, 24 hr, 48 hr and 96 hrs). When screening for prophylactic or therapeutic agents, a clone of the full sequence of a biomarker of the invention or functional portion thereof is used to transfect the cells. The transfected cells are cultured for varying amounts of time (*i.e.*, 1, 2, 3, 5, 7, 10, or 14 days) in the presence or absence of test compound. Following incubation, target nucleic acid samples are
15 prepared from the cells and hybridized to a nucleic acid probe corresponding to a nucleic acid sequence which are differentially expressed in individuals with one or more colorectal pathologies. The nucleic acid probe is labeled, for example, with a radioactive label, according to methods well-known in the art and described herein. Hybridization is carried out by northern blot, for example as described in Ausubel et al., *supra* or Sambrook et al., *supra*). The
20 differential hybridization, as defined herein, of the target to the samples on the array from normal relative to RNA from samples having one or more colorectal pathologies is indicative of the level of expression of RNA corresponding to a differentially expressed specific nucleic acid sequence. A change in the level of expression of the target sequence as a result of the incubation step in the presence of the test compound, is indicative of a compound that increases or decreases
25 the expression of the corresponding polyp biomarker specific nucleic acid sequence.

The present invention also provides a method for identifying a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies, said method comprises: (a) contacting a cell-free extract (*e.g.*, an intestinal cell extract) with a nucleic acid sequence encoding a protein or RNA products of one or more biomarkers of the invention and a
30 test compound; (b) determining the amount of the protein or RNA product present in (a); and (c) comparing the amount(s) in (a) to that present to a corresponding control that has not been

contacted with the test compound, so that if the amount of the protein or RNA product is altered relative to the amount in the control, a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies is identified. In a specific embodiment, the expression level(s) is altered by 5%, 10%, 15%, 25%, 30%, 40%, 50%, 5 to 25%, 10 to 30%, at least 1 fold, at least 1.5 fold, at least 2 fold, 4 fold, 5 fold, 10 fold, 25 fold, 1 to 10 fold, or 5 to 25 fold relative to the expression level in the control sample determined by utilizing an assay described herein (*e.g.*, a microarray or QRT-PCR) or an assay well known to one of skill in the art. In alternate embodiments, such a method comprises determining the amount of a protein or RNA product of at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, 1 to 3, 1 to 5, 1-8, all or any combination of the biomarkers of the invention present in the extract and comparing the amounts to those present in the control.

In certain embodiments, the amount of RNA product of a biomarker of the invention is determined, in other embodiments, the amount of protein products of a biomarker of the invention is determined, while in still other embodiments, the amount of RNA and protein products of a biomarker of the invention is determined. Standard methods and compositions for determining the amount of RNA or protein products of a biomarker of the invention can be utilized. Such methods and compositions are described in detail above.

Kits to Identify Compounds that Modulate the Expression or Activity of a Biomarker

In specific embodiments, in a screening assay described herein, the amount of protein or RNA product of a biomarker of the invention is determined utilizing kits. Such kits comprise materials and reagents required for measuring the expression of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8 protein or RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, all or any combination of the biomarkers of the invention. In specific embodiments, the kits may further comprise one or more additional reagents employed in the various methods, such as: (1) reagents for purifying RNA from blood; (2) primers for generating test nucleic acids; (3) dNTPs and/or rNTPs (either premixed or separate), optionally with one or more uniquely labeled dNTPs and/or rNTPs (*e.g.*, biotinylated or Cy3 or Cy5 tagged dNTPs); (4) post synthesis labeling reagents, such as chemically active derivatives of fluorescent dyes; (5) enzymes, such as reverse transcriptases, DNA polymerases, and the like; (6) various buffer mediums, *e.g.*, hybridization and washing buffers; (7) labeled probe purification reagents and components, like spin columns, etc.; and (8)

protein purification reagents; (9) signal generation and detection reagents, *e.g.*, streptavidin-alkaline phosphatase conjugate, chemifluorescent or chemiluminescent substrate, and the like. In particular embodiments, the kits comprise prelabeled quality controlled protein and or RNA transcript (in some embodiments, mRNA) for use as a control.

5 In some embodiments, the kits are QRT-PCR kits. In other embodiments, the kits are nucleic acid arrays and protein arrays. Such kits according to the subject invention will at least comprise an array having associated protein or nucleic acid members of the invention and packaging means therefore. Alternatively the protein or nucleic acid members of the invention may be prepackaged onto an array.

10 In a specific embodiment, kits for measuring a RNA product of a biomarker of the invention comprise materials and reagents that are necessary for measuring the expression of the RNA product. For example, a microarray or QRT-PCR kit may be used and contain only those reagents and materials necessary for measuring the levels of RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, all or any combination of the
15 biomarkers of the invention. Alternatively, in some embodiments, the kits can comprise materials and reagents that are not limited to those required to measure the levels of RNA products of 1, 2, 3, 4, 5, 6, 7, 8 all or any combination of the biomarkers of the invention. For example, a microarray kit may contain reagents and materials necessary for measuring the levels of RNA products 1, 2, 3, 4, 5, 6, 7, 8, all or any combination of the biomarkers of the invention,
20 in addition to reagents and materials necessary for measuring the levels of the RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, or more genes other than the biomarkers of the invention. In a specific embodiment, a microarray or QRT-PCR kit contains reagents and materials necessary for measuring the levels of RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, all or any
25 combination of the biomarkers of the invention, and 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, or more genes that are not biomarkers of the invention, or 1-10, 1-100, 1-150, 1-200, 1-300, 1-400, 1-500, 1-1000, 25-100, 25-200, 25-300, 25-400, 25-500, 25-1000, 100-150, 100-200, 100-300, 100-400, 100-500, 100-1000 or 500-1000 or more genes that are not biomarkers of the invention.

30 For nucleic acid microarray kits, the kits generally comprise probes attached or localized to a support surface. The probes may be labeled with a detectable label. In a specific

embodiment, the probes are specific for the 5' region, the 3' region, the internal coding region, an exon(s), an intron(s), an exon junction(s), or an exon-intron junction(s), of 1, 2, 3, 4, 5, 6, 7, 8, all or any combination of the biomarkers of the invention. The microarray kits may comprise instructions for performing the assay and methods for interpreting and analyzing the data
5 resulting from the performance of the assay. The kits may also comprise hybridization reagents and/or reagents necessary for detecting a signal produced when a probe hybridizes to a target nucleic acid sequence. Generally, the materials and reagents for the microarray kits are in one or more containers. Each component of the kit is generally in its own a suitable container.

For QRT-PCR kits, the kits generally comprise pre-selected primers specific for
10 particular RNA products (*e.g.*, an exon(s), an intron(s), an exon junction(s), and an exon-intron junction(s)) of 1, 2, 3, 4, 5, 6, 7, 8, all or any combination of the biomarkers of the invention. The QRT-PCR kits may also comprise enzymes suitable for reverse transcribing and/or amplifying nucleic acids (*e.g.*, polymerases such as Taq, enzymes such as reverse transcriptase etc.), and deoxynucleotides and buffers needed for the reaction mixture for reverse transcription
15 and amplification. The QRT-PCR kits may also comprise biomarker specific sets of primers specific for 1, 2, 3, 4, 5, 6, 7, 8, all or any combination of the biomarkers of the invention. The QRT-PCR kits may also comprise biomarker specific probes which are specific for the sequences amplified from 1, 2, 3, 4, 5, 6, 7, 8, all or any combination of the biomarkers of the invention using the biomarker specific sets of primers. The probes may or may not be labeled
20 with a detectable label (*e.g.*, a fluorescent label). In some embodiments, when contemplating multiplexing it is helpful if the probes are labeled with a different detectable label (*e.g.* FAM or HEX) Each component of the QRT-PCR kit is generally in its own suitable container. Thus, these kits generally comprise distinct containers suitable for each individual reagent, enzyme, primer and probe. Further, the QRT-PCR kits may comprise instructions for performing the
25 assay and methods for interpreting and analyzing the data resulting from the performance of the assay.

For antibody based kits, the kit can comprise, for example: (1) a first antibody (which may or may not be attached to a support) which binds to protein of interest (*e.g.*, a protein products of 1, 2, 3, 4, 5, 6, 7, 8, all or any combination of the biomarkers of the invention); and,
30 optionally, (2) a second, different antibody which binds to either the protein, or the first antibody and is conjugated to a detectable label (*e.g.*, a fluorescent label, radioactive isotope or enzyme).

The antibody-based kits may also comprise beads for conducting an immunoprecipitation. Each component of the antibody-based kits is generally in its own suitable container. Thus, these kits generally comprise distinct containers suitable for each antibody. Further, the antibody-based kits may comprise instructions for performing the assay and methods for interpreting and analyzing the data resulting from the performance of the assay.

Reporter gene-based assays may also be conducted to identify a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies. In a specific embodiment, the present invention provides a method for identifying a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies, said method comprising: (a) contacting a compound with a cell expressing a reporter gene construct comprising a reporter gene operably linked to a regulatory element of a biomarker of the invention (*e.g.*, a promoter/enhancer element); (b) measuring the expression of said reporter gene; and (c) comparing the amount in (a) to that present in a corresponding control cell that has not been contacted with the test compound, so that if the amount of expressed reporter gene is altered relative to the amount in the control cell, a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies is identified. In accordance with this embodiment, the cell may naturally express the biomarker or be engineered to express the biomarker. In another embodiment, the present invention provides a method for identifying a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies, said method comprising: (a) contacting a compound with a cell-free extract and a reporter gene construct comprising a reporter gene operably linked to a regulatory element of a biomarker of the invention (*e.g.*, a promoter/enhancer element); (b) measuring the expression of said reporter gene; and (c) comparing the amount in (a) to that present in a corresponding control that has not been contacted with the test compound, so that if the amount of expressed reporter gene is altered relative to the amount in the control, a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies is identified.

Any reporter gene well-known to one of skill in the art may be used in reporter gene constructs used in accordance with the methods of the invention. Reporter genes refer to a nucleotide sequence encoding a RNA transcript or protein that is readily detectable either by its presence (by, *e.g.*, RT-PCR, Northern blot, Western Blot, ELISA, *etc.*) or activity. Non-limiting examples of reporter genes are listed in **Table 10**. Reporter genes may be obtained and the

nucleotide sequence of the elements determined by any method well-known to one of skill in the art. The nucleotide sequence of a reporter gene can be obtained, *e.g.*, from the literature or a database such as GenBank. Alternatively, a polynucleotide encoding a reporter gene may be generated from nucleic acid from a suitable source. If a clone containing a nucleic acid encoding
5 a particular reporter gene is not available, but the sequence of the reporter gene is known, a nucleic acid encoding the reporter gene may be chemically synthesized or obtained from a suitable source (*e.g.*, a cDNA library, or a cDNA library generated from, or nucleic acid, preferably poly A+ RNA, isolated from, any tissue or cells expressing the reporter gene) by PCR amplification. Once the nucleotide sequence of a reporter gene is determined, the nucleotide
10 sequence of the reporter gene may be manipulated using methods well-known in the art for the manipulation of nucleotide sequences, *e.g.*, recombinant DNA techniques, site directed mutagenesis, PCR, etc. (see, for example, the techniques described in Sambrook et al., 1990, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY and Ausubel et al., eds., 1998, Current Protocols in Molecular Biology, John Wiley
15 & Sons, NY, which are both incorporated by reference herein in their entireties), to generate reporter genes having a different amino acid sequence, for example to create amino acid substitutions, deletions, and/or insertions.

In accordance with the invention, cells that naturally or normally express one or more, all or any combination of the biomarkers of the invention can be used in the methods described
20 herein. Alternatively, cells can be engineered to express one or more, all or any combination of the biomarkers of the invention, or a reporter gene using techniques well-known in the art and used in the methods described herein. Examples of such techniques include, but are not to, calcium phosphate precipitation (see, *e.g.*, Graham & Van der Eb, 1978, Virology 52:546), dextran-mediated transfection, calcium phosphate mediated transfection, polybrene mediated
25 transfection, protoplast fusion, electroporation, encapsulation of the nucleic acid in liposomes, and direct microinjection of the nucleic acid into nuclei.

In a specific embodiment, the cells used in the methods described herein are intestinal cells or cell lines, lymphocytes (T or B lymphocytes), monocytes, neutrophils, macrophages, eosinophils, basophils, erythrocytes or platelets. In a preferred embodiment, the cells used in the
30 methods described herein are intestinal cells. In another preferred embodiment, the cells used in

the methods described herein are lymphocytes. In another embodiment, the cells used in the methods described herein are immortalized cell lines derived from a source, *e.g.*, a tissue.

Any cell-free extract that permits the translation, and optionally but preferably, the transcription, of a nucleic acid can be used in accordance with the methods described herein.

5 The cell-free extract may be isolated from cells of any species origin. For example, the cell-free translation extract may be isolated from human cells, cultured mouse cells, cultured rat cells, Chinese hamster ovary (CHO) cells, *Xenopus* oocytes, rabbit reticulocytes, wheat germ, or rye embryo (see, *e.g.*, Krieg & Melton, 1984, *Nature* 308:203 and Dignam et al., 1990 *Methods Enzymol.* 182:194-203). Alternatively, the cell-free translation extract, *e.g.*, rabbit reticulocyte
10 lysates and wheat germ extract, can be purchased from, *e.g.*, Promega, (Madison, WI). In a preferred embodiment, the cell-free extract is an extract isolated from human cells. In a specific embodiment, the human cells are HeLa cells, lymphocytes, or intestinal cells or cell lines.

In addition to the ability to modulate the expression levels of RNA and/or protein products a biomarker of the invention, it may be desirable, at least in certain instances, that
15 compounds modulate the activity of a protein products of a biomarker of the invention. Thus, the present invention provides methods of identifying compounds to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies, comprising methods for identifying compounds that modulate the activity of a protein products of one or more biomarkers of the invention. Such methods can comprise: (a) contacting a cell expressing a
20 protein products of one or more biomarkers of the invention with a test compound; (b) determining the activity level of the protein products; and (c) comparing the activity level to that in a corresponding control cell that has not been contacted with the test compound, so that if the level of activity in (a) is altered relative to the level of activity in the control cell, a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies is
25 identified. In a specific embodiment, the activity level(s) is altered by 5%, 10%, 15%, 25%, 30%, 40%, 50%, 5 to 25%, 10 to 30%, at least 1 fold, at least 1.5 fold, at least 2 fold, 4 fold, 5 fold, 10 fold, 25 fold, 1 to 10 fold, or 5 to 25 fold relative to the activity level in the control as determined by utilizing an assay described herein (*e.g.*, a microarray or QRT-PCR) or an assay well known to one of skill in the art. In alternate embodiments, such a method comprises
30 determining the activity level of a protein products of at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, 1 to 5, 1-10, 5-10, 5-15,

or 10-15, or more, or all or any combination of the biomarkers of the invention present in the cell and comparing the activity levels to those present in the control.

The present invention provides methods of identifying compounds to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies, comprising: (a) contacting a cell-free extract with a nucleic acid encoding a protein products of one or more biomarkers of the invention and a test compound; (b) determining the activity level of the protein products; and (c) comparing the activity level to that in a corresponding control that has not been contacted with the test compound, so that if the level of activity in (a) is altered relative to the level of activity in the control, a compound to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies is identified. In a specific embodiment, the activity level(s) is altered by 5%, 10%, 15%, 25%, 30%, 40%, 50%, 5 to 25%, 10 to 30%, at least 1 fold, at least 1.5 fold, at least 2 fold, 4 fold, 5 fold, 10 fold, 25 fold, 1 to 10 fold, or 5 to 25 fold relative to the activity level in the control as determined by utilizing an assay described herein (e.g., a microarray or QRT-PCR) or an assay well known to one of skill in the art. In alternate embodiments, such a method comprises determining the activity level of a protein products of at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, 1 to 3, 1 to 5, 1-8 all or any combination of the biomarkers of the invention present in the sample and comparing the activity levels to those present in the control.

Standard techniques can be utilized to determine the level of activity of a protein products of a biomarker of the invention. Activities of protein productss of biomarkers of the invention that can be determined using techniques well known in the art.

Method to Utilize the Biological Activity of the Compounds

Upon identification of compounds to be tested for an ability to prevent, treat, or ameliorate one or more colorectal pathologies (for convenience referred to herein as a "lead" compound), the compounds can be further investigated. For example, the compounds identified via the present methods can be further tested *in vivo* in accepted animal models of polyp formation. Further, the compounds identified via the methods can be analyzed with respect to their specificity. Techniques for such additional compound investigation are well known to one of skill in the art.

In one embodiment, the effect of a lead compound can be assayed by measuring the cell growth or viability of the target cell. Such assays can be carried out with representative cells of

cell types involved in polyp formation (*e.g.*, intestinal cells; cells isolated from different portions of the gastrointestinal system and the like). Alternatively, instead of culturing cells from a patient, a lead compound may be screened using cells of a cell line.

Many assays well-known in the art can be used to assess the survival and/or growth of a patient cell or cell line following exposure to a lead compound; for example, cell proliferation can be assayed by measuring Bromodeoxyuridine (BrdU) incorporation (see, *e.g.*, Hoshino et al., 1986, *Int. J. Cancer* 38, 369; Campana et al., 1988, *J. Immunol. Meth.* 107:79) or (³H)-thymidine incorporation (see, *e.g.*, Chen, J., 1996, *Oncogene* 13:1395-403; Jeoung, J., 1995, *J. Biol. Chem.* 270:18367-73), by direct cell count, by detecting changes in transcription, translation or activity of known genes such as proto-oncogenes (*e.g.*, *fos*, *myc*) or cell cycle markers (Rb, *cdc2*, cyclin A, D1, D2, D3, E, etc). The levels of such protein and RNA (*e.g.*, mRNA) and activity can be determined by any method well known in the art. For example, protein can be quantitated by known immunological based methods such as Western blotting or immunoprecipitation using commercially available antibodies. mRNA can be quantitated using methods that are well known and routine in the art, for example, using northern analysis, RNase protection, the polymerase chain reaction in connection with the reverse transcription. Cell viability can be assessed by using trypan-blue staining or other cell death or viability markers known in the art. In a specific embodiment, the level of cellular ATP is measured to determine cell viability. Differentiation can be assessed, for example, visually based on changes in morphology.

20 Animal Models

Compounds can be tested in suitable animal model systems prior to use in humans. Such animal model systems include but are not limited to rats, mice, chicken, cows, monkeys, pigs, dogs, rabbits, etc. Any animal system well-known in the art may be used. In certain embodiments, compounds are tested in a mouse model. Compounds can be administered repeatedly.

Accepted animal models can be utilized to determine the efficacy of the compounds identified via the methods described above for the prevention, treatment, and/or amelioration of one or more colorectal pathologies.

Toxicity

30 The toxicity and/or efficacy of a compound identified in accordance with the invention can be determined by standard pharmaceutical procedures in cell cultures or experimental

animals, *e.g.*, for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). Cells and cell lines that can be used to assess the cytotoxicity of a compound identified in accordance with the invention include, but are not limited to, peripheral blood mononuclear cells (PBMCs), Caco-2 cells, and
5 Huh7 cells. The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. A compound identified in accordance with the invention that exhibits large therapeutic indices is preferred. While a compound identified in accordance with the invention that exhibits toxic side effects may be used, care should be taken to design a delivery system that targets such agents to the site of affected tissue in order to
10 minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage of a compound identified in accordance with the invention for use in humans. The dosage of such agents lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this
15 range depending upon the dosage form employed and the route of administration utilized. For any agent used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (*i.e.*, the concentration of the compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture.
20 Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

Design of Congeners or Analogs

The compounds which display the desired biological activity can be used as lead compounds for the development or design of congeners or analogs having useful
25 pharmacological activity. For example, once a lead compound is identified, molecular modeling techniques can be used to design variants of the compound that can be more effective. Examples of molecular modeling systems are the CHARM and QUANTA programs (Polygen Corporation, Waltham, MA). CHARM performs the energy minimization and molecular dynamics functions. QUANTA performs the construction, graphic modelling and analysis of molecular structure.

QUANTA allows interactive construction, modification, visualization, and analysis of the behavior of molecules with each other.

A number of articles review computer modeling of drugs interactive with specific proteins, such as Rotivinen et al., 1988, Acta Pharmaceutical Fennica 97:159-166; Ripka, 1998, 5 New Scientist 54-57; McKinaly & Rossmann, 1989, Annu. Rev. Pharmacol. Toxicol. 29:111-122; Perry & Davies, OSAR: Quantitative Structure-Activity Relationships in Drug Design pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis & Dean, 1989, Proc. R. Soc. Lond. 236:125-140 and 141-162; Askew et al., 1989, J. Am. Chem. Soc. 111:1082-1090. Other computer programs that screen and graphically depict chemicals are available from companies such as BioDesign, 10 Inc. (Pasadena, California), Allelix, Inc. (Mississauga, Ontario, Canada), and Hypercube, Inc. (Cambridge, Ontario). Although these are primarily designed for application to drugs specific to particular proteins, they can be adapted to design of drugs specific to any identified region. The analogs and congeners can be tested for binding to the proteins of interest (*i.e.*, the protein products of a biomarker of the invention) using the above-described screens for biologic activity. 15 Alternatively, lead compounds with little or no biologic activity, as ascertained in the screen, can also be used to design analogs and congeners of the compound that have biologic activity.

Compounds

Compounds that can be tested and identified methods described herein can include, but are not limited to, compounds obtained from any commercial source, including Aldrich (1001 20 West St. Paul Ave., Milwaukee, WI 53233), Sigma Chemical (P.O. Box 14508, St. Louis, MO 63178), Fluka Chemie AG (Industriestrasse 25, CH-9471 Buchs, Switzerland (Fluka Chemical Corp. 980 South 2nd Street, Ronkonkoma, NY 11779)), Eastman Chemical Company, Fine Chemicals (P.O. Box 431, Kingsport, TN 37662), Boehringer Mannheim GmbH (Sandhofer Strasse 116, D-68298 Mannheim), Takasago (4 Volvo Drive, Rockleigh, NJ 07647), SST 25 Corporation (635 Brighton Road, Clifton, NJ 07012), Ferro (111 West Irene Road, Zachary, LA 70791), Riedel-deHaen Aktiengesellschaft (P.O. Box D-30918, Seelze, Germany), PPG Industries Inc., Fine Chemicals (One PPG Place, 34th Floor, Pittsburgh, PA 15272). Further any kind of natural products may be screened using the methods of the invention, including microbial, fungal, plant or animal extracts.

30 Compounds from large libraries of synthetic or natural compounds can be screened. Numerous means are currently used for random and directed synthesis of saccharide, peptide,

and nucleic acid-based compounds. Synthetic compound libraries are commercially available from a number of companies including Maybridge Chemical Co. (Trevillet, Cornwall, UK), Comgenex (Princeton, NJ), Brandon Associates (Merrimack, NH), and Microsource (New Milford, CT). A rare chemical library is available from Aldrich (Milwaukee, WI).

5 Combinatorial libraries are available and are prepared. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available from e.g., Pan Laboratories (Bothell, WA) or MycoSearch (NC), or are readily produceable by methods well known in the art. Additionally, natural and synthetically produced libraries and compounds are readily modified through conventional chemical, physical, and biochemical means.

10 Furthermore, diversity libraries of test compounds, including small molecule test compounds, may be utilized. Libraries screened using the methods of the present invention can comprise a variety of types of compounds. Examples of libraries that can be screened in accordance with the methods of the invention include, but are not limited to, peptoids; random biooligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous
15 polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; and small molecule libraries (in some embodiments, small organic molecule libraries). In some embodiments, the compounds in the libraries screened are nucleic acid or peptide molecules. In a non-limiting example, peptide molecules can exist in a phage display library. In other embodiments, the types of compounds
20 include, but are not limited to, peptide analogs including peptides comprising non-naturally occurring amino acids, e.g., D-amino acids, phosphorous analogs of amino acids, such as α -amino phosphoric acids and α -amino phosphonic acids, or amino acids having non-peptide linkages, nucleic acid analogs such as phosphorothioates and PNAs, hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin,
25 acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose. Libraries of polypeptides or proteins can also be used in the assays of the invention.

In a specific embodiment, the combinatorial libraries are small organic molecule libraries including, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones,
30 pyrrolidines, morpholino compounds, and benzodiazepines. In another embodiment, the combinatorial libraries comprise peptoids; random bio-oligomers; benzodiazepines; diversomers

such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; or carbohydrate libraries. Combinatorial libraries are themselves commercially available. For example, libraries may be commercially obtained from, *e.g.*, Specs and BioSpecs B.V. (Rijswijk, The Netherlands), Chembridge Corporation (San Diego, CA), Contract Service Company (Dolgoprudny, Moscow Region, Russia), Comgenex USA Inc. (Princeton, NJ), Maybridge Chemicals Ltd. (Cornwall PL34 OHW, United Kingdom), Asinex (Moscow, Russia), ComGenex (Princeton, New Jersey), Ru, Tripos, Inc. (St. Louis, Missouri), ChemStar, Ltd (Moscow, Russia), 3D Pharmaceuticals (Exton, Pennsylvania), and Martek Biosciences (Columbia, Maryland).

In a preferred embodiment, the library is preselected so that the compounds of the library are more amenable for cellular uptake. For example, compounds are selected based on specific parameters such as, but not limited to, size, lipophilicity, hydrophilicity, and hydrogen bonding, which enhance the likelihood of compounds getting into the cells. In another embodiment, the compounds are analyzed by three-dimensional or four-dimensional computer computation programs.

The combinatorial compound library for use in accordance with the methods of the present invention may be synthesized. There is a great interest in synthetic methods directed toward the creation of large collections of small organic compounds, or libraries, which could be screened for pharmacological, biological or other activity. The synthetic methods applied to create vast combinatorial libraries are performed in solution or in the solid phase, *i.e.*, on a support. Solid-phase synthesis makes it easier to conduct multi-step reactions and to drive reactions to completion with high yields because excess reagents can be easily added and washed away after each reaction step. Solid-phase combinatorial synthesis also tends to improve isolation, purification and screening. However, the more traditional solution phase chemistry supports a wider variety of organic reactions than solid-phase chemistry.

Combinatorial compound libraries of the present invention may be synthesized using the apparatus described in U.S. Patent No. 6,190,619 to Kilcoin et al., which is hereby incorporated by reference in its entirety. U.S. Patent No. 6,190,619 discloses a synthesis apparatus capable of holding a plurality of reaction vessels for parallel synthesis of multiple discrete compounds or for combinatorial libraries of compounds.

In one embodiment, the combinatorial compound library can be synthesized in solution. The method disclosed in U.S. Patent No. 6,194,612 to Boger et al., which is hereby incorporated by reference in its entirety, features compounds useful as templates for solution phase synthesis of combinatorial libraries. The template is designed to permit reaction products to be easily
5 purified from unreacted reactants using liquid/liquid or solid/liquid extractions. The compounds produced by combinatorial synthesis using the template will in some embodiments be small organic molecules. Some compounds in the library may mimic the effects of non-peptides or peptides. In contrast to solid phase synthesis of combinatorial compound libraries, liquid phase synthesis does not require the use of specialized protocols for monitoring the individual steps of
10 a multistep solid phase synthesis (Egner et al., 1995, J.Org. Chem. 60:2652; Anderson et al., 1995, J. Org. Chem. 60:2650; Fitch et al., 1994, J. Org. Chem. 59:7955; Look et al., 1994, J. Org. Chem. 49:7588; Metzger et al., 1993, Angew. Chem., Int. Ed. Engl. 32:894; Youngquist et al., 1994, Rapid Commun. Mass Spect. 8:77; Chu et al., 1995, J. Am. Chem. Soc. 117:5419; Brummel et al., 1994, Science 264:399; and Stevanovic et al., 1993, Bioorg. Med. Chem. Lett.
15 3:431).

Combinatorial compound libraries useful for the methods of the present invention can be synthesized on supports. In one embodiment, a split synthesis method, a protocol of separating and mixing supports during the synthesis, is used to synthesize a library of compounds on supports (see *e.g.* Lam et al., 1997, Chem. Rev. 97:41-448; Ohlmeyer et al., 1993, Proc. Natl.
20 Acad. Sci. USA 90:10922-10926 and references cited therein). Each support in the final library has substantially one type of compound attached to its surface. Other methods for synthesizing combinatorial libraries on supports, wherein one product is attached to each support, will be known to those of skill in the art (*see, e.g.*, Nefzi et al., 1997, Chem. Rev. 97:449-472).

In some embodiments of the present invention, compounds can be attached to supports
25 via linkers. Linkers can be integral and part of the support, or they may be nonintegral that are either synthesized on the support or attached thereto after synthesis. Linkers are useful not only for providing points of compound attachment to the support, but also for allowing different groups of molecules to be cleaved from the support under different conditions, depending on the nature of the linker. For example, linkers can be, *inter alia*, electrophilically cleaved,
30 nucleophilically cleaved, photocleavable, enzymatically cleaved, cleaved by metals, cleaved under reductive conditions or cleaved under oxidative conditions. In a preferred embodiment,

the compounds are cleaved from the support prior to high throughput screening of the compounds.

If the library comprises arrays or microarrays of compounds, wherein each compound has an address or identifier, the compound can be deconvoluted, *e.g.*, by cross-referencing the
5 positive sample to original compound list that was applied to the individual test assays.

If the library is a peptide or nucleic acid library, the sequence of the compound can be determined by direct sequencing of the peptide or nucleic acid. Such methods are well known to one of skill in the art.

A number of physico-chemical techniques can be used for the *de novo* characterization of
10 compounds. Examples of such techniques include, but are not limited to, mass spectrometry, NMR spectroscopy, X-ray crystallography and vibrational spectroscopy.

(Q) Use of Identified Compounds to Prevent, Treat, or Ameliorate One or more Colorectal Pathologies

The present invention provides methods of preventing, treating, or ameliorating one or
15 more colorectal pathologies, said methods comprising administering to a subject in need thereof one or more compounds identified in accordance with the methods of the invention. In a preferred embodiment, the individual is human.

In one embodiment, the invention provides a method of preventing, treating, or ameliorating one or more colorectal pathologies, said method comprising administering to an
20 individual in need thereof a dose of a prophylactically or therapeutically effective amount of one or more compounds identified in accordance with the methods of the invention. In a specific embodiment, a compound identified in accordance with the methods of the invention is not administered to prevent, treat, or ameliorate one or more colorectal pathologies, if such compound has been used previously to prevent, treat, or ameliorate one or more colorectal
25 pathologies. In another embodiment, a compound identified in accordance with the methods of the invention is not administered to prevent, treat, or ameliorate one or more colorectal pathologies, if such compound has suggested to be used to prevent, treat, or ameliorate one or more colorectal pathologies. In another embodiment, a compound identified in accordance with the methods of the invention specifically binds to and/or alters the expression and/or activity
30 level of a protein or RNA products of only one biomarker of the invention. In another embodiment, a compound identified in accordance with the methods of the invention is not

administered to prevent, treat, or ameliorate one or more colorectal pathologies, if such compound binds to and/or alters the expression and/or activity of a protein or RNA products of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or more or all or any combination of the biomarkers of **Tables 2** or **6**. In yet another embodiment, a compound identified in accordance
5 with the methods of the invention binds to and/or alters the expression and/or activity level of a protein or RNA products of at least 2, at least 3, at least 4, at least 5, at least 10, at least 15, or more biomarkers of the invention.

The invention also provides methods of preventing, treating, or ameliorating one or more colorectal pathologies, said methods comprising administering to a subject in need thereof one or
10 more of the compounds identified utilizing the screening methods described herein, and one or more other therapies (*e.g.*, prophylactic or therapeutic agents and surgery). In a specific embodiment, such therapies are currently being used, have been used or are known to be useful in the prevention, treatment, or amelioration of one or more colorectal pathologies (including, but not limited to the prophylactic or therapeutic agents listed herein). The therapies (*e.g.*,
15 prophylactic or therapeutic agents) of the combination therapies of the invention can be administered sequentially or concurrently. In a specific embodiment, the combination therapies of the invention comprise a compound identified in accordance with the invention and at least one other therapy that has the same mechanism of action as said compound. In another specific embodiment, the combination therapies of the invention comprise a compound identified in
20 accordance with the methods of the invention and at least one other therapy (*e.g.*, prophylactic or therapeutic agent) which has a different mechanism of action than said compound. The combination therapies of the present invention improve the prophylactic or therapeutic effect of a compound of the invention by functioning together with the compound to have an additive or synergistic effect. The combination therapies of the present invention reduce the side effects
25 associated with the therapies (*e.g.*, prophylactic or therapeutic agents).

The prophylactic or therapeutic agents of the combination therapies can be administered to a subject in the same pharmaceutical composition. Alternatively, the prophylactic or therapeutic agents of the combination therapies can be administered concurrently to a subject in separate pharmaceutical compositions. The prophylactic or therapeutic agents may be
30 administered to a subject by the same or different routes of administration.

In specific embodiment, a pharmaceutical composition comprising one or more compounds identified in an assay described herein is administered to an individual, in some embodiments a human, to prevent, treat, or ameliorate one or more colorectal pathologies. In accordance with the invention, the pharmaceutical composition may also comprise one or more prophylactic or therapeutic agents. In some embodiments, such agents are currently being used, have been used or are known to be useful in the prevention, treatment, or amelioration of one or more colorectal pathologies.

(R) Compounds of the Invention

Representative, non-limiting examples of compounds that can be used in accordance with the methods of the invention to prevent, treat, and/or ameliorate one or more colorectal pathologies are described in detail below.

First, such compounds can include, for example, antisense, ribozyme, or triple helix compounds that can downregulate the expression or activity of a protein or RNA products of a biomarker of the invention. Such compounds are described in detail in the subsection below.

Second, such compounds can include, for example, antibody compositions that can modulate the expression of a protein or RNA products of a biomarker of the invention, or the activity of a protein products of a biomarker of the invention. In a specific embodiment, the antibody compositions downregulate the expression a protein or RNA products of a biomarker of the invention, or the activity of a protein products of a biomarker of the invention. Such compounds are described in detail in the subsection below.

Third, such compounds can include, for example, protein products of a biomarker of the invention. The invention encompasses the use of peptides or peptide mimetics selected to mimic a protein products of a biomarker of the invention to prevent, treat, or ameliorate one or more colorectal pathologies. Further, such compounds can include, for example, dominant-negative polypeptides that can modulate the expression a protein or RNA products of a biomarker of the invention, or the activity protein products of a biomarker of the invention.

The methods also encompass the use derivatives, analogs and fragments of protein products of a biomarker of the invention to prevent, treat, or ameliorate one or more colorectal pathologies. In particular, the invention encompasses the use of fragments of protein products of a biomarker of the invention comprising one or more domains of such a protein(s) to prevent, treat, or ameliorate one or more colorectal pathologies. In another specific embodiment, the

invention encompasses the use of a protein products of a biomarker of the invention, or an analog, derivative or fragment of such a protein which is expressed as a fusion, or chimeric protein products (comprising the protein, fragment, analog, or derivative joined via a peptide bond to a heterologous protein sequence).

5 In specific embodiments, an antisense oligonucleotide of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, or more of biomarkers of the invention are administered to prevent, treat, or ameliorate one or more colorectal pathologies. In other embodiments, one or more of protein products of a biomarker of the invention or a fragment, analog, or derivative thereof are administered to prevent, treat, or
10 ameliorate one or more colorectal pathologies. In other embodiment, one or more antibodies that specifically bind to protein products of the invention are administered to prevent, treat, or ameliorate one or more colorectal pathologies. In other embodiments, one or more dominant-negative polypeptides are administered to prevent, treat, or ameliorate one or more colorectal pathologies.

15 **Antisense, Ribozyme, Triple-Helix Compositions**

Standard techniques can be utilized to produce antisense, triple helix, or ribozyme molecules reactive to one or more of the genes listed in **Tables 2** or **6**, and transcripts of the genes listed in **Tables 2** or **6**, for use as part of the methods described herein. First, standard techniques can be utilized for the production of antisense nucleic acid molecules, *i.e.*, molecules
20 which are complementary to a sense nucleic acid encoding a polypeptide of interest, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, *e.g.*, all or part of the protein coding region (or open reading frame). An
25 antisense nucleic acid molecule can be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a polypeptide of interest. The non-coding regions ("5' and 3' untranslated regions") are the 5' and 3' sequences that flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or
30 50 nucleotides or more in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art.

For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxy-hydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest).

Antisense nucleic acid molecules administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA encoding the polypeptide of interest to thereby inhibit expression, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue, *e.g.*, a joint (*e.g.*, a knee, hip, elbow, and knuckle), site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell, *e.g.*, a T cell or intestinal cell, surface, *e.g.*, by

linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using vectors, *e.g.*, gene therapy vectors, described below. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of interest can be an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α -units, the strands run parallel to each other (Gaultier et al., 1987, Nucleic Acids Res. 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-*o*-methylribonucleotide (Inoue et al., 1987, Nucleic Acids Res. 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330).

Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region, and can also be generated using standard techniques. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach, 1988, Nature 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity for a nucleic acid molecule encoding a polypeptide of interest can be designed based upon the nucleotide sequence of a cDNA disclosed herein. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, an mRNA encoding a polypeptide of interest can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel and Szostak, 1993, Science 261:1411-1418.

Triple helical structures can also be generated using well known techniques. For example, expression of a polypeptide of interest can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the gene encoding the polypeptide (*e.g.*, the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells. *See generally* Helene, 1991, Anticancer Drug Des. 6(6):569-84; Helene, 1992, Ann. N.Y. Acad. Sci. 660:27-36; and Maher, 1992, Bioassays 14(12):807-15.

In various embodiments, nucleic acid compositions can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (*see* Hyrup et al., 1996, *Bioorganic & Medicinal Chemistry* 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al., 1996, *supra*; Perry-O'Keefe et al., 1996, *Proc. Natl. Acad. Sci. USA* 93: 14670-675.

PNAs can, for example, be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup, 1996, *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, 1996, *supra*, and Finn et al., 1996, *Nucleic Acids Res.* 24(17):3357-63. For example, a DNA chain can be synthesized on a support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al., 1989, *Nucleic Acids Res.* 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al., 1996, *Nucleic Acids Res.* 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al., 1975, *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the

cell membrane (*see, e.g.*, Letsinger et al., 1989, Proc. Natl. Acad. Sci. USA 86:6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. USA 84:648-652; International Publication No. W0 88/09810) or the blood-brain barrier (*see, e.g.*, International Publication No. W0 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (*see, e.g.*, Krol et al., 1988, Bio/Techniques 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988, Pharm. Res. 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

Antibody Compositions

10 In one embodiment, antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention are administered to an individual, in some embodiments a human, to prevent, treat, or ameliorate one or more colorectal pathologies. In another embodiment, any combination of antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention are administered to a subject, in some
15 embodiments a human, to prevent, treat, or ameliorate one or more colorectal pathologies. In a specific embodiment, one or more antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention are administered to a subject, in some embodiments a human, in combination with other types of therapies to prevent, treat, or ameliorate one or more colorectal pathologies.

20 One or more antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention can be administered to a subject, in some embodiments a human, using various delivery systems are known to those of skill in the art. For example, such antibodies can be administered by encapsulation in liposomes, microparticles or microcapsules. See, *e.g.*, U.S. Patent No. 5,762,904, U.S. Patent No. 6,004,534, and International Publication
25 No. WO 99/52563. In addition, such antibodies can be administered using recombinant cells capable of expressing the antibodies, or retroviral, other viral vectors or non-viral vectors capable of expressing the antibodies.

Antibodies that specifically bind one or more protein products of one or more biomarkers of the invention can be obtained from any known source. For example, **Table 5** provides a list
30 of commercially available antibodies specific for one or more of the protein products of the biomarkers of the invention. Alternatively, antibodies that specifically bind to one or more

protein products of one or more biomarkers of the invention can be produced by any method known in the art for the synthesis of antibodies, in particular, by chemical synthesis or in some embodiments, by recombinant expression techniques.

Antibodies include, but are not limited to, polyclonal antibodies, monoclonal antibodies, 5 bispecific antibodies, multispecific antibodies, human antibodies, humanized antibodies, camelised antibodies, chimeric antibodies, single-chain Fvs (scFv) (see *e.g.*, Bird *et al.* (1988) *Science* 242:423-426; and Huston *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:5879-5883), single chain antibodies, single domain antibodies, Fab fragments, F(ab') fragments, disulfide-linked Fvs (sdFv), and anti-idiotypic (anti-Id) antibodies (including, *e.g.*, anti-Id antibodies to 10 antibodies of the invention), and epitope-binding fragments of any of the above. The term "antibody", as used herein, refers to immunoglobulin molecules and immunologically active fragments of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site. Immunoglobulin molecules can be of any type (*e.g.*, IgG, IgE, IgM, IgD, IgA and IgY), class (*e.g.*, IgG₁, IgG₂, IgG₃, IgG₄, IgA₁ and IgA₂) or subclass. Examples of immunologically active 15 fragments of immunoglobulin molecules include F(ab) fragments (a monovalent fragment consisting of the VL, VH, CL and CH1 domains) and F(ab')₂ fragments (a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region) which can be generated by treating the antibody with an enzyme such as pepsin or papain. Immunologically active fragments also include, but are not limited to, Fd fragments (consisting of the VH and 20 CH1 domains), Fv fragments (consisting of the VL and VH domains of a single arm of an antibody), dAb fragments (consisting of a VH domain; Ward *et al.*, (1989) *Nature* 341:544-546), and isolated complementarity determining regions (CDRs). Antibodies that specifically bind to an antigen can be produced by any method known in the art for the synthesis of antibodies, in particular, by chemical synthesis or in some embodiments, by recombinant expression 25 techniques.

Polyclonal antibodies that specifically bind to an antigen can be produced by various procedures well-known in the art. For example, a human antigen can be administered to various host animals including, but not limited to, rabbits, mice, rats, etc. to induce the production of sera containing polyclonal antibodies specific for the human antigen. Various adjuvants may be used 30 to increase the immunological response, depending on the host species, and include but are not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide,

surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and corynebacterium parvum. Such adjuvants are also well known in the art.

5 The term “monospecific antibody” refers to an antibody that displays a single binding specificity and affinity for a particular target, *e.g.*, epitope. This term includes monoclonal antibodies. Monoclonal antibodies can be prepared using a wide variety of techniques known in the art including the use of hybridoma, recombinant, and phage display technologies, or a combination thereof. See, *e.g.*, U.S. Pat. Nos. RE 32,011, 4,902,614, 4,543,439, 4,411,993 and
10 4,196,265; Kennett et al (eds.), *Monoclonal Antibodies, Hybridomas: A New Dimension in Biological Analyses*, Plenum Press (1980); and Harlow and Lane (eds.), *Antibodies. A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1988), which are incorporated herein by reference. For example, monoclonal antibodies can be produced using hybridoma techniques including those known in the art and taught, for example, in Harlow et al., *Antibodies: A
15 Laboratory Manual*, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988); Hammerling, et al., in: *Monoclonal Antibodies and T-Cell Hybridomas* 563-681 (Elsevier, N.Y., 1981) (said references incorporated by reference in their entireties). Other techniques that enable the production of antibodies through recombinant techniques (*e.g.*, techniques described by William D. Huse et al., 1989, *Science*, 246: 1275-1281; L. Sastry et al., 1989, *Proc. Natl. Acad. Sci.
20 USA*, 86: 5728-5732; and Michelle Alting-Mees et al., *Strategies in Molecular Biology*, 3: 1-9 (1990) involving a commercial system available from Stratacyte, La Jolla, Calif.) may also be utilized to construct monoclonal antibodies. The term “monoclonal antibody” as used herein is not limited to antibodies produced through hybridoma technology. The term “monoclonal antibody” refers to an antibody that is derived from a single clone, including any eukaryotic,
25 prokaryotic, or phage clone, and not the method by which it is produced.

Methods for producing and screening for specific antibodies using hybridoma technology are routine and well known in the art. Briefly, mice can be immunized with a protein products of a biomarker of the invention, and once an immune response is detected, *e.g.*, antibodies specific for the protein are detected in the mouse serum, the mouse spleen is harvested and splenocytes
30 isolated. The splenocytes are then fused by well known techniques to any suitable myeloma cells, for example cells from cell line SP20 available from the ATCC. Hybridomas are selected

and cloned by limited dilution. Additionally, a RIMMS (repetitive immunization multiple sites) technique can be used to immunize an animal (Kilpatrick et al., 1997, Hybridoma 16:381-9, incorporated by reference in its entirety). The hybridoma clones are then assayed by methods known in the art for cells that secrete antibodies capable of binding a polypeptide of the invention. Ascites fluid, which generally contains high levels of antibodies, can be generated by immunizing mice with positive hybridoma clones.

Accordingly, the present invention provides methods of generating antibodies by culturing a hybridoma cell secreting an antibody of the invention wherein, in some embodiments, the hybridoma is generated by fusing splenocytes isolated from a mouse immunized with a protein products of a biomarker of the invention, with myeloma cells and then screening the hybridomas resulting from the fusion for hybridoma clones that secrete an antibody able to bind to the protein or protein fragment.

Antibody fragments which recognize specific epitopes of a protein products of a biomarker of the invention may be generated by any technique known to those of skill in the art. For example, Fab and F(ab')₂ fragments of the invention may be produced by proteolytic cleavage of immunoglobulin molecules, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')₂ fragments). F(ab')₂ fragments contain the variable region, the light chain constant region and the CH1 domain of the heavy chain. Further, the antibodies of the present invention can also be generated using various phage display methods known in the art.

In phage display methods, functional antibody domains are displayed on the surface of phage particles which carry the polynucleotide sequences encoding them. In particular, DNA sequences encoding VH and VL domains are amplified from animal cDNA libraries (e.g., human or murine cDNA libraries of affected tissues). The DNA encoding the VH and VL domains are recombined together with an scFv linker by PCR and cloned into a phagemid vector. The vector is electroporated in *E. coli* and the *E. coli* is infected with helper phage. Phage used in these methods are typically filamentous phage including fd and M13 and the VH and VL domains are usually recombinantly fused to either the phage gene III or gene VIII. Phage expressing an antigen binding domain that binds to a particular antigen can be selected or identified with antigen, e.g., using labeled antigen or antigen bound or captured to a solid surface or bead. Examples of phage display methods that can be used to make the antibodies of the present

invention include those disclosed in Brinkman et al., 1995, J. Immunol. Methods 182:41-50; Ames et al., 1995, J. Immunol. Methods 184:177-186; Kettleborough et al., 1994, Eur. J. Immunol. 24:952-958; Persic et al., 1997, Gene 187:9-18; Burton et al., 1994, Advances in Immunology 57:191-280; PCT Application No. PCT/GB91/O1 134; International Publication
5 Nos. WO 90/02809, WO 91/10737, WO 92/01047, WO 92/18619, WO 93/1 1236, WO 95/15982, WO 95/20401, and WO97/13844; and U.S. Patent Nos. 5,698,426, 5,223,409, 5,403,484, 5,580,717, 5,427,908, 5,750,753, 5,821,047, 5,571,698, 5,427,908, 5,516,637, 5,780,225, 5,658,727, 5,733,743 and 5,969,108; each of which is incorporated herein by reference in its entirety.

10 As described in the above references, after phage selection, the antibody coding regions from the phage can be isolated and used to generate whole antibodies, including human antibodies, or any other desired antigen binding fragment, and expressed in any desired host, including mammalian cells, insect cells, plant cells, yeast, and bacteria, *e.g.*, as described below. Techniques to recombinantly produce Fab, Fab' and F(ab')₂ fragments can also be employed
15 using methods known in the art such as those disclosed in International Publication No. WO 92/22324; Mullinax et al., 1992, BioTechniques 12(6):864-869; Sawai et al., 1995, AJRI 34:26-34; and Better et al., 1988, Science 240:1041-1043 (said references incorporated by reference in their entirety).

To generate whole antibodies, PCR primers including VH or VL nucleotide sequences, a
20 restriction site, and a flanking sequence to protect the restriction site can be used to amplify the VH or VL sequences in scFv clones. Utilizing cloning techniques known to those of skill in the art, the PCR amplified VH domains can be cloned into vectors expressing a VH constant region, *e.g.*, the human gamma 4 constant region, and the PCR amplified VL domains can be cloned into vectors expressing a VL constant region, *e.g.*, human kappa or lambda constant regions. In some
25 embodiments, the vectors for expressing the VH or VL domains comprise an EF-1a promoter, a secretion signal, a cloning site for the variable domain, constant domains, and a selection marker such as neomycin. The VH and VL domains may also be cloned into one vector expressing the necessary constant regions. The heavy chain conversion vectors and light chain conversion vectors are then co-transfected into cell lines to generate stable or transient cell lines that express
30 full-length antibodies, *e.g.*, IgG, using techniques known to those of skill in the art.

For some uses, including *in vivo* use of antibodies in humans and *in vitro* detection assays, it may be preferable to use human or chimeric antibodies. Completely human antibodies are particularly desirable for therapeutic treatment of human subjects. Human antibodies can be made by a variety of methods known in the art including phage display methods described above
5 using antibody libraries derived from human immunoglobulin sequences. See also U.S. Patent Nos. 4,444,887 and 4,716,111; and International Publication Nos. WO 98/46645, WO 98/50433, WO 98/24893, WO98/16654, WO 96/34096, WO 96/33735, and WO 91/10741; each of which is incorporated herein by reference in its entirety.

Antibodies can also be produced by a transgenic animal. In particular, human antibodies
10 can be produced using transgenic mice which are incapable of expressing functional endogenous immunoglobulins, but which can express human immunoglobulin genes. For example, the human heavy and light chain immunoglobulin gene complexes may be introduced randomly or by homologous recombination into mouse embryonic stem cells. Alternatively, the human variable region, constant region, and diversity region may be introduced into mouse embryonic
15 stem cells in addition to the human heavy and light chain genes. The mouse heavy and light chain immunoglobulin genes may be rendered non-functional separately or simultaneously with the introduction of human immunoglobulin loci by homologous recombination. In particular, homozygous deletion of the J_H region prevents endogenous antibody production. The modified embryonic stem cells are expanded and microinjected into blastocysts to produce chimeric mice.
20 The chimeric mice are then be bred to produce homozygous offspring which express human antibodies. The transgenic mice are immunized in the normal fashion with a selected antigen, *e.g.*, all or a portion of a polypeptide of the invention. Monoclonal antibodies directed against the antigen can be obtained from the immunized, transgenic mice using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange
25 during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA, IgM and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for
30 producing such antibodies, *see, e.g.*, International Publication Nos. WO 98/24893, WO 96/34096, and WO 96/33735; and U.S. Patent Nos. 5,413,923, 5,625,126, 5,633,425, 5,569,825,

5,661,016, 5,545,806, 5,814,318, and 5,939,598, which are incorporated by reference herein in their entirety. In addition, companies such as Abgenix, Inc. (Freemont, CA) and Genpharm (San Jose, CA) can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

5 U.S. Patent No. 5,849,992, for example, describes a method of expressing an antibody in the mammary gland of a transgenic mammal. A transgene is constructed that includes a milk-specific promoter and nucleic acids encoding the antibody of interest and a signal sequence for secretion. The milk produced by females of such transgenic mammals includes, secreted-therein, the antibody of interest. The antibody can be purified from the milk, or for some applications,
10 used directly.

A chimeric antibody is a molecule in which different portions of the antibody are derived from different immunoglobulin molecules. Methods for producing chimeric antibodies are known in the art. See *e.g.*, Morrison, 1985, Science 229:1202; Oi et al., 1986, BioTechniques 4:214; Gillies et al., 1989, J. Immunol. Methods 125:191-202; and U.S. Patent Nos. 5,807,715,
15 4,816,567, 4,816,397, and 6,331,415, which are incorporated herein by reference in their entirety.

A humanized antibody is an antibody or its variant or fragment thereof which is capable of binding to a predetermined antigen and which comprises a framework region having substantially the amino acid sequence of a human immunoglobulin and a CDR having
20 substantially the amino acid sequence of a non-human immunoglobulin. A humanized antibody comprises substantially all of at least one, and typically two, variable domains (Fab, Fab', F(ab')₂, Fabc, Fv) in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin (*i.e.*, donor antibody) and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. In some
25 embodiments, a humanized antibody also comprises at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin. Ordinarily, the antibody will contain both the light chain as well as at least the variable domain of a heavy chain. The antibody also may include the CH1, hinge, CH2, CH3, and CH4 regions of the heavy chain. The humanized antibody can be selected from any class of immunoglobulins, including IgM, IgG,
30 IgD, IgA and IgE, and any isotype, including IgG₁, IgG₂, IgG₃ and IgG₄. Usually the constant domain is a complement fixing constant domain where it is desired that the humanized antibody

exhibit cytotoxic activity, and the class is typically IgG₁. Where such cytotoxic activity is not desirable, the constant domain may be of the IgG₂ class. The humanized antibody may comprise sequences from more than one class or isotype, and selecting particular constant domains to optimize desired effector functions is within the ordinary skill in the art. The framework and

5 CDR regions of a humanized antibody need not correspond precisely to the parental sequences, *e.g.*, the donor CDR or the consensus framework may be mutagenized by substitution, insertion or deletion of at least one residue so that the CDR or framework residue at that site does not correspond to either the consensus or the import antibody. Such mutations, however, will not be extensive. Usually, at least 75% of the humanized antibody residues will correspond to those of

10 the parental FR and CDR sequences, more often 90%, and most often greater than 95%. Humanized antibody can be produced using variety of techniques known in the art, including but not limited to, CDR-grafting (European Patent No. EP 239,400; International Publication No. WO 91/09967; and U.S. Patent Nos. 5,225,539, 5,530,101, and 5,585,089), veneering or resurfacing (European Patent Nos. EP 592,106 and EP 519,596; Padlan, 1991, Molecular

15 Immunology 28(4/5):489-498; Studnicka et al., 1994, Protein Engineering 7(6):805-814; and Roguska et al. , 1994, PNAS 91:969-973), chain shuffling (U.S. Patent No. 5,565,332), and techniques disclosed in, *e.g.* , U.S. Pat. No. 6,407,213, U.S. Pat. No. 5,766,886, WO 9317105, Tan et al., 2002, J. Immunol. 169:1119-25, Caldas et al., 2000, Protein Eng. 13(5):353 - 60, Morea et al., 2000, Methods 20(3):267-79, Baca et al., 1997, J. Biol. Chem. 272(16):10678-84,

20 Roguska et al., 1996, Protein Eng. 9(10):895-904, Couto et al., 1995, Cancer Res. 55 (23 Supp):5973s - 5977s, Couto et al., 1995, Cancer Res. 55(8):1717-22, Sandhu JS, 1994, Gene 150(2):409-10, and Pedersen et al., 1994, J. Mol. Biol. 235(3):959-73. Often, framework residues in the framework regions will be substituted with the corresponding residue from the CDR donor antibody to alter, preferably improve, antigen binding. These framework

25 substitutions are identified by methods well known in the art, *e.g.*, by modeling of the interactions of the CDR and framework residues to identify framework residues important for antigen binding and sequence comparison to identify unusual framework residues at particular positions. (See, *e.g.*, Queen et al., U.S. Patent No. 5,585,089; and Riechmann et al., 1988, Nature 332:323, which are incorporated herein by reference in their entirety.)

30 Single domain antibodies, for example, antibodies lacking the light chains, can be produced by methods well-known in the art. See Riechmann et al., 1999, J. Immuno. 231:25-38;

Nuttall et al., 2000, Curr. Pharm. Biotechnol. 1(3):253-263; Muylderman, 2001, J. Biotechnol. 74(4):277302; U.S. Patent No. 6,005,079; and International Publication Nos. WO 94/04678, WO 94/25591, and WO 01/44301, each of which is incorporated herein by reference in its entirety.

Further, the antibodies that specifically bind to an antigen can, in turn, be utilized to generate anti-idiotypic antibodies that "mimic" an antigen using techniques well known to those skilled in the art. (See, *e.g.*, Greenspan & Bona, 1989, FASEB J. 7(5):437-444; and Nissinoff, 1991, J. Immunol. 147(8):2429-2438). Such antibodies can be used, alone or in combination with other therapies, in the prevention, treatment, or amelioration of one or more colorectal pathologies.

The invention encompasses polynucleotides comprising a nucleotide sequence encoding an antibody or fragment thereof that specifically binds to an antigen. The invention also encompasses polynucleotides that hybridize under high stringency, intermediate or lower stringency hybridization conditions to polynucleotides that encode an antibody of the invention.

The polynucleotides may be obtained, and the nucleotide sequence of the polynucleotides determined, by any method known in the art. The nucleotide sequences encoding known antibodies can be determined using methods well known in the art, *i.e.*, nucleotide codons known to encode particular amino acids are assembled in such a way to generate a nucleic acid that encodes the antibody. Such a polynucleotide encoding the antibody may be assembled from chemically synthesized oligonucleotides (*e.g.*, as described in Kutmeier et al., 1994, BioTechniques 17:242), which, briefly, involves the synthesis of overlapping oligonucleotides containing portions of the sequence encoding the antibody, fragments, or variants thereof, annealing and ligating of those oligonucleotides, and then amplification of the ligated oligonucleotides by PCR.

Alternatively, a polynucleotide encoding an antibody may be generated from nucleic acid from a suitable source. If a clone containing a nucleic acid encoding a particular antibody is not available, but the sequence of the antibody molecule is known, a nucleic acid encoding the immunoglobulin may be chemically synthesized or obtained from a suitable source (*e.g.*, an antibody cDNA library or a cDNA library generated from, or nucleic acid, in some embodiments poly A+ RNA, isolated from, any tissue or cells expressing the antibody, such as hybridoma cells selected to express an antibody of the invention) by PCR amplification using synthetic primers hybridizable to the 3' and 5' ends of the sequence or by cloning using an oligonucleotide probe

specific for the particular gene sequence to identify, *e.g.*, a cDNA clone from a cDNA library that encodes the antibody. Amplified nucleic acids generated by PCR may then be cloned into replicable cloning vectors using any method well known in the art.

Once the nucleotide sequence of the antibody is determined, the nucleotide sequence of
5 the antibody may be manipulated using methods well known in the art for the manipulation of nucleotide sequences, *e.g.*, recombinant DNA techniques, site directed mutagenesis, PCR, etc. (see, for example, the techniques described in Sambrook et al., 1990, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY and Ausubel et al., eds., 1998, Current Protocols in Molecular Biology, John Wiley & Sons, NY,
10 which are both incorporated by reference herein in their entireties), to generate antibodies having a different amino acid sequence, for example to create amino acid substitutions, deletions, and/or insertions.

Once a polynucleotide encoding an antibody molecule, heavy or light chain of an antibody, or fragment thereof (in some embodiments, but not necessarily, containing the heavy
15 or light chain variable domain) of the invention has been obtained, the vector for the production of the antibody molecule may be produced by recombinant DNA technology using techniques well-known in the art.

In one preferred embodiment, monoclonal antibodies are produced in mammalian cells. Preferred mammalian host cells for expressing the clone antibodies or antigen-binding fragments
20 thereof include Chinese Hamster Ovary (CHO cells) (including dhfr- CHO cells, described in Urlaub and Chasin (1980, Proc. Natl. Acad. Sci. USA 77:4216-4220), used with a DHFR selectable marker, *e.g.*, as described in Kaufman and Sharp (1982, Mol. Biol. 159:601-621), lymphocytic cell lines, *e.g.*, NS0 myeloma cells and SP2 cells, COS cells, and a cell from a transgenic animal, *e.g.*, a transgenic mammal. For example, the cell is a mammary epithelial
25 cell.

In addition to the nucleic acid sequence encoding the diversified immunoglobulin domain, the recombinant expression vectors may carry additional sequences, such as sequences that regulate replication of the vector in host cells (*e.g.*, origins of replication) and selectable marker genes. The selectable marker gene facilitates selection of host cells into which the vector
30 has been introduced (see *e.g.*, U.S. Patents Nos. 4,399,216, 4,634,665 and 5,179,017). For example, typically the selectable marker gene confers resistance to drugs, such as G418,

hygromycin or methotrexate, on a host cell into which the vector has been introduced. Preferred selectable marker genes include the dihydrofolate reductase (DHFR) gene (for use in *dhfr*⁻ host cells with methotrexate selection/amplification) and the *neo* gene (for G418 selection).

5 In an exemplary system for recombinant expression of an antibody, or antigen-binding portion thereof, of the invention, a recombinant expression vector encoding both the antibody heavy chain and the antibody light chain is introduced into *dhfr*⁻ CHO cells by calcium phosphate-mediated transfection. Within the recombinant expression vector, the antibody heavy and light chain genes are each operatively linked to enhancer/promoter regulatory elements (*e.g.*, derived from SV40, CMV, adenovirus and the like, such as a CMV enhancer/AdMLP promoter regulatory element or an SV40 enhancer/AdMLP promoter regulatory element) to drive high levels of transcription of the genes. The recombinant expression vector also carries a DHFR gene, which allows for selection of CHO cells that have been transfected with the vector using methotrexate selection/amplification. The selected transformant host cells are cultured to allow for expression of the antibody heavy and light chains and intact antibody is recovered from the culture medium. Standard molecular biology techniques are used to prepare the recombinant expression vector, transfect the host cells, select for transformants, culture the host cells and recover the antibody from the culture medium. For example, some antibodies can be isolated by affinity chromatography with a Protein A or Protein G.

For antibodies that include an Fc domain, the antibody production system preferably synthesizes antibodies in which the Fc region is glycosylated. For example, the Fc domain of IgG molecules is glycosylated at asparagine 297 in the CH2 domain. This asparagine is the site for modification with biantennary-type oligosaccharides. It has been demonstrated that this glycosylation is required for effector functions mediated by Fcγ receptors and complement C1q (Burton and Woof, 1992, Adv. Immunol. 51 :1-84 ; Jefferis et al., 1998, Immunol. Rev. 163:59-76). In a preferred embodiment, the Fc domain is produced in a mammalian expression system that appropriately glycosylates the residue corresponding to asparagine 297. The Fc domain can also include other eukaryotic post-translational modifications.

Once an antibody molecule has been produced by recombinant expression, it may be purified by any method known in the art for purification of an immunoglobulin molecule, for example, by chromatography (*e.g.*, ion exchange, affinity, particularly by affinity for the specific antigen after Protein A, and sizing column chromatography), centrifugation, differential

solubility, or by any other standard technique for the purification of proteins. Further, the antibodies or fragments thereof may be fused to heterologous polypeptide sequences known in the art to facilitate purification.

Gene Therapy Techniques

5 Gene therapy refers to therapy performed by the administration to a subject of an expressed or expressible nucleic acid. Any of the methods for gene therapy available in the art can be used according to the present invention. Exemplary methods are described below.

10 In specific embodiments, one or more antisense oligonucleotides for one or more biomarkers of the invention are administered to prevent, treat, or ameliorate one or more colorectal pathologies, by way of gene therapy. In other embodiments, one or more nucleic acid molecules comprising nucleotides encoding one or more antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention are administered to prevent, treat, or ameliorate one or more colorectal pathologies, by way of gene therapy. In other
15 embodiments, one or more nucleic acid molecules comprising nucleotides encoding protein products of one or more biomarkers of the invention or analogs, derivatives or fragments thereof, are administered to prevent, treat, or ameliorate one or more colorectal pathologies, by way of gene therapy. In yet other embodiments, one or more nucleic acid molecules comprising nucleotides encoding one or more dominant-negative polypeptides of one or more protein products of one or more biomarker of the invention are administered to prevent, treat, or
20 ameliorate one or more colorectal pathologies, by way of gene therapy.

 For general reviews of the methods of gene therapy, see Goldspiel et al., 1993, *Clinical Pharmacy* 12:488-505; Wu and Wu, 1991, *Biotherapy* 3:87-95; Tolstoshev, 1993, *Ann. Rev. Pharmacol. Toxicol.* 32:573-596; Mulligan, 1993, *Science* 260:926-932; and Morgan and Anderson, 1993, *Ann. Rev. Biochem.* 62:191-217; May, 1993, *TIBTECH* 11(5):155-215).
25 Methods commonly known in the art of recombinant DNA technology which can be used are described in Ausubel et al. (eds.), 1993, *Current Protocols in Molecular Biology*, John Wiley & Sons, NY; and Kriegler, 1990, *Gene Transfer and Expression, A Laboratory Manual*, Stockton Press, NY.

 In one aspect, a composition of the invention comprises nucleic acid sequences encoding
30 one or more antibodies that specifically bind to one or more protein products of one or more biomarkers of the invention, said nucleic acid sequences being part of expression vectors that

express one or more antibodies in a suitable host. In particular, such nucleic acid sequences have promoters operably linked to the antibodies, said promoter being inducible or constitutive, and, optionally, tissue-specific.

5 In another aspect, a composition of the invention comprises nucleic acid sequences encoding dominant-negative polypeptides of one or protein products of one or more biomarkers of the invention, said nucleic acid sequences being part of expression vectors that express dominant-negative polypeptides in a suitable host. In particular, such nucleic acid sequences have promoters operably linked to the dominant-negative polypeptides, said promoter being inducible or constitutive, and, optionally, tissue-specific. In another particular embodiment,
10 nucleic acid molecules are used in which the dominant-negative coding sequences and any other desired sequences are flanked by regions that promote homologous recombination at a desired site in the genome, thus providing for intrachromosomal expression of the dominant-negative nucleic acids (Koller and Smithies, 1989, Proc. Natl. Acad. Sci. USA 86:8932-8935; Zijlstra et al., 1989, Nature 342:435-438).

15 Delivery of the nucleic acids into a patient may be either direct, in which case the patient is directly exposed to the nucleic acid or nucleic acid-carrying vectors, or indirect, in which case, cells are first transformed with the nucleic acids *in vitro*, then transplanted into the patient. These two approaches are known, respectively, as *in vivo* or *ex vivo* gene therapy.

In a specific embodiment, the nucleic acid sequence is directly administered *in vivo*,
20 where it is expressed to produce the encoded product. This can be accomplished by any of numerous methods known in the art, *e.g.*, by constructing it as part of an appropriate nucleic acid expression vector and administering it so that they become intracellular, *e.g.*, by infection using defective or attenuated retrovirals or other viral vectors (see U.S. Patent No. 4,980,286), or by direct injection of naked DNA, or by use of microparticle bombardment (*e.g.*, a gene gun;
25 Biolistic, Dupont), or coating with lipids or cell-surface receptors or transfecting agents, encapsulation in liposomes, microparticles, or microcapsules, or by administering them in linkage to a peptide which is known to enter the nucleus, by administering it in linkage to a ligand subject to receptor-mediated endocytosis (see, *e.g.*, Wu and Wu, 1987, J. Biol. Chem. 262:4429-4432) (which can be used to target cell types specifically expressing the receptors),
30 etc. In another embodiment, nucleic acid-ligand complexes can be formed in which the ligand comprises a fusogenic viral peptide to disrupt endosomes, allowing the nucleic acid to avoid

lysosomal degradation. In yet another embodiment, the nucleic acid can be targeted *in vivo* for cell specific uptake and expression, by targeting a specific receptor (see, *e.g.*, International Publication Nos. WO 92/06180 dated April 16, 1992 (Wu et al.); WO 92/22635 dated December 23, 1992 (Wilson et al.); WO92/20316 dated November 26, 1992 (Findeis et al.); WO 93/14188
5 dated July 22, 1993 (Clarke et al.), WO 93/20221 dated October 14, 1993 (Young)). Alternatively, the nucleic acid can be introduced intracellularly and incorporated within host cell DNA for expression, by homologous recombination (Koller and Smithies, 1989, Proc. Natl. Acad. Sci. USA 86:8932-8935; Zijlstra et al., 1989, Nature 342:435-438).

For example, a retroviral vector can be used. These retroviral vectors have been modified
10 to delete retroviral sequences that are not necessary for packaging of the viral genome and integration into host cell DNA. The nucleic acid sequences encoding the antibodies of interest, or proteins of interest or fragments thereof to be used in gene therapy are cloned into one or more vectors, which facilitates delivery of the gene into a patient. More detail about retroviral vectors can be found in Boesen et al., 1994, Biotherapy 6:291-302, which describes the use of a
15 retroviral vector to deliver the *mdr1* gene to hematopoietic stem cells in order to make the stem cells more resistant to chemotherapy. Other references illustrating the use of retroviral vectors in gene therapy are: Clowes et al., 1994, J. Clin. Invest. 93:644-651; Kiem et al., 1994, Blood 83:1467-1473; Salmons and Gunzberg, 1993, Human Gene Therapy 4:129-141; and Grossman and Wilson, 1993, Curr. Opin. in Genetics and Devel. 3:110-114.

20 Adenoviruses are other viral vectors that can be used in gene therapy. Adenoviruses are especially attractive vehicles for delivering genes to respiratory epithelia. Adenoviruses naturally infect respiratory epithelia where they cause a mild disease. Other targets for adenovirus-based delivery systems are liver, the central nervous system, endothelial cells, and muscle. Adenoviruses have the advantage of being capable of infecting non-dividing cells.
25 Kozarsky and Wilson, 1993, Current Opinion in Genetics and Development 3:499-503 present a review of adenovirus-based gene therapy. Bout et al., 1994, Human Gene Therapy 5:3-10 demonstrated the use of adenovirus vectors to transfer genes to the respiratory epithelia of rhesus monkeys. Other instances of the use of adenoviruses in gene therapy can be found in Rosenfeld et al., 1991, Science 252:431-434; Rosenfeld et al., 1992, Cell 68:143-155; Mastrangeli et al.,
30 1993, J. Clin. Invest. 91:225-234; PCT Publication WO94/12649; and Wang, et al., 1995, Gene Therapy 2:775-783. In a preferred embodiment, adenovirus vectors are used.

Adeno-associated virus (AAV) has also been proposed for use in gene therapy (Walsh et al., 1993, Proc. Soc. Exp. Biol. Med. 204:289-300; U.S. Patent No. 5,436,146).

Another approach to gene therapy involves transferring a gene to cells in tissue culture by such methods as electroporation, lipofection, calcium phosphate mediated transfection, or viral infection. Usually, the method of transfer includes the transfer of a selectable marker to the cells. The cells are then placed under selection to isolate those cells that have taken up and are expressing the transferred gene. Those cells are then delivered to a patient.

In this embodiment, the nucleic acid is introduced into a cell prior to administration *in vivo* of the resulting recombinant cell. Such introduction can be carried out by any method known in the art, including but not limited to transfection, electroporation, microinjection, infection with a viral or bacteriophage vector containing the nucleic acid sequences, cell fusion, chromosome-mediated gene transfer, microcell-mediated gene transfer, spheroplast fusion, etc. Numerous techniques are known in the art for the introduction of foreign genes into cells (see, e.g., Loeffler and Behr, 1993, Meth. Enzymol. 217:599-618; Cohen et al., 1993, Meth. Enzymol. 217:618-644; Cline, 1985, Pharmac. Ther. 29:69-92) and may be used in accordance with the present invention, provided that the necessary developmental and physiological functions of the recipient cells are not disrupted. The technique should provide for the stable transfer of the nucleic acid to the cell, so that the nucleic acid is expressible by the cell and preferably heritable and expressible by its cell progeny.

The resulting recombinant cells can be delivered to a patient by various methods known in the art. Recombinant blood cells (e.g., hematopoietic stem or progenitor cells) and/or intestinal cells are preferably administered intravenously. The amount of cells envisioned for use depends on the desired effect, patient state, etc., and can be determined by one skilled in the art.

Cells into which a nucleic acid can be introduced for purposes of gene therapy encompass any desired, available cell type, and include but are not limited to epithelial cells, endothelial cells, keratinocytes, intestinal cells, fibroblasts, muscle cells, hepatocytes; blood cells such as T lymphocytes, B lymphocytes, monocytes, macrophages, neutrophils, eosinophils, megakaryocytes, granulocytes; various stem or progenitor cells, in particular hematopoietic stem or progenitor cells, e.g., as obtained from bone marrow, umbilical cord blood, peripheral blood, fetal liver, etc.

In a preferred embodiment, the cell used for gene therapy is autologous to the patient.

In one embodiment in which recombinant cells are used in gene therapy, nucleic acid sequences encoding antibodies of interest, or proteins of interest or fragments thereof are introduced into the cells such that they are expressible by the cells or their progeny, and the recombinant cells are then administered *in vivo* for therapeutic effect. In a specific embodiment, stem or progenitor cells are used. Any stem and/or progenitor cells which can be isolated and maintained *in vitro* can potentially be used in accordance with this embodiment of the present invention (see, *e.g.*, International Publication No. WO 94/08598, dated April 28, 1994; Stemple and Anderson, 1992, *Cell* 71:973-985; Rheinwald, 1980, *Meth. Cell Bio.* 21A:229; and Pittelkow and Scott, 1986, *Mayo Clinic Proc.* 61:771).

Promoters that may be used to control the expression of nucleic acid sequences encoding antibodies of interest, proteins of interest or fragments thereof may be constitutive, inducible or tissue-specific. Non-limiting examples include the SV40 early promoter region (Bernoist and Chambon, 1981, *Nature* 290:304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto, et al., 1980, *Cell* 22:787-797), the herpes thymidine kinase promoter (Wagner et al., 1981, *Proc. Natl. Acad. Sci. USA* 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, *Nature* 296:39-42); prokaryotic expression vectors such as the β -lactamase promoter (Villa-Komaroff et al., 1978, *Proc. Natl. Acad. Sci. USA* 75:3727-3731), or the *tac* promoter (DeBoer et al., 1983, *Proc. Natl. Acad. Sci. USA* 80:21-25); see also "Useful proteins from recombinant bacteria" in *Scientific American*, 1980, 242:74-94; plant expression vectors comprising the nopaline synthetase promoter region (Herrera-Estrella et al., *Nature* 303:209-213) or the cauliflower mosaic virus 35S RNA promoter (Gardner et al., 1981, *Nucl. Acids Res.* 9:2871), and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase (Herrera-Estrella et al., 1984, *Nature* 310:115-120); promoter elements from yeast or other fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, alkaline phosphatase promoter, and the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: elastase I gene control region which is active in pancreatic acinar cells (Swift et al., 1984, *Cell* 38:639-646; Ornitz et al., 1986, *Cold Spring Harbor Symp. Quant. Biol.* 50:399-409; MacDonald, 1987, *Hepatology* 7:425-515); insulin gene control region which is active in pancreatic beta cells (Hanahan, 1985, *Nature* 315:115-122), immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et al., 1984, *Cell* 38:647-658;

Adames et al., 1985, Nature 318:533-538; Alexander et al., 1987, Mol. Cell. Biol. 7:1436-1444), mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder et al., 1986, Cell 45:485-495), albumin gene control region which is active in liver (Pinkert et al., 1987, Genes and Devel. 1:268-276), alpha-fetoprotein gene control region
5 which is active in liver (Krumlauf et al., 1985, Mol. Cell. Biol. 5:1639-1648; Hammer et al., 1987, Science 235:53-58; alpha 1-antitrypsin gene control region which is active in the liver (Kelsey et al., 1987, Genes and Devel. 1:161-171), beta-globin gene control region which is active in myeloid cells (Mogam et al., 1985, Nature 315:338-340; Kollias et al., 1986, Cell 46:89-94; myelin basic protein gene control region which is active in oligodendrocyte cells in the
10 brain (Readhead et al., 1987, Cell 48:703-712); myosin light chain-2 gene control region which is active in skeletal muscle (Sani, 1985, Nature 314:283-286), and gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason et al., 1986, Science 234:1372-1378).

In a specific embodiment, the nucleic acid to be introduced for purposes of gene therapy
15 comprises an inducible promoter operably linked to the coding region, such that expression of the nucleic acid is controllable by controlling the presence or absence of the appropriate inducer of transcription.

(S) Pharmaceutical Compositions

Biologically active compounds identified using the methods of the invention or a
20 pharmaceutically acceptable salt thereof can be administered to a patient, in some embodiments a mammal, including a human, having one or more colorectal pathologies. In a specific embodiment, a compound or pharmaceutically acceptable salt thereof is administered to a patient, in some embodiments a mammal, including a human, having one or more colorectal pathologies. In another embodiment, a compound or a pharmaceutically acceptable salt thereof
25 is administered to a patient, in some embodiments a mammal, including a human, as a preventative measure against one or more colorectal pathologies. In accordance with these embodiments, the patient may be a child, an adult or elderly, wherein a "child" is a subject between the ages of 24 months of age and 18 years of age, an "adult" is a subject 18 years of age or older, and "elderly" is a subject 65 years of age or older.

30 When administered to a patient, the compound or a pharmaceutically acceptable salt thereof is administered as component of a composition that optionally comprises a

pharmaceutically acceptable vehicle. The composition can be administered orally, or by any other convenient route, for example, by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (*e.g.*, oral mucosa, rectal, and intestinal mucosa, *etc.*) and may be administered together with another biologically active agent. Administration can be systemic or local. Various delivery systems are known, *e.g.*, encapsulation in liposomes, microparticles, microcapsules, capsules, *etc.*, and can be used to administer the compound and pharmaceutically acceptable salts thereof.

Methods of administration include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, oral, sublingual, intranasal, intracerebral, intravaginal, transdermal, rectally, by inhalation, or topically, particularly to the ears, nose, eyes, or skin. The mode of administration is left to the discretion of the practitioner. In most instances, administration will result in the release of the compound or a pharmaceutically acceptable salt thereof into the bloodstream.

In specific embodiments, it may be desirable to administer the compound or a pharmaceutically acceptable salt thereof locally. This may be achieved, for example, and not by way of limitation, by local infusion during surgery, topical application, *e.g.*, in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers. In a specific embodiment, a compound is administered locally to one or more sections of the gastrointestinal system.

In certain embodiments, it may be desirable to introduce the compound or a pharmaceutically acceptable salt thereof into the gastrointestinal system by any suitable route, including intraventricular, intrathecal and epidural injection. Intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir.

Pulmonary administration can also be employed, *e.g.*, by use of an inhaler or nebulizer, and formulation with an aerosolizing agent, or via perfusion in a fluorocarbon or synthetic pulmonary surfactant. In certain embodiments, the compound and pharmaceutically acceptable salts thereof can be formulated as a suppository, with traditional binders and vehicles such as triglycerides.

In another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a vesicle, in particular a liposome (see Langer, 1990, Science 249:1527-1533; Treat et al., in Liposomes in the Therapy of Infectious Disease and Cancer, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see
5 generally *ibid.*).

In yet another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a controlled release system (see, *e.g.*, Goodson, in Medical Applications of Controlled Release, *supra*, vol. 2, pp. 115-138 (1984)). Other controlled-release systems discussed in the review by Langer, 1990, Science 249:1527-1533 may be used. In one
10 embodiment, a pump may be used (see Langer, *supra*; Sefton, 1987, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald et al., 1980, Surgery 88:507; Saudek et al., 1989, N. Engl. J. Med. 321:574). In another embodiment, polymeric materials can be used (see Medical Applications of Controlled Release, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); Controlled Drug Bioavailability, Drug Product Design and Performance, Smolen and Ball (eds.), Wiley,
15 New York (1984); Ranger and Peppas, 1983, J. Macromol. Sci. Rev. Macromol. Chem. 23:61; see also Levy et al., 1985, Science 228:190; During et al., 1989, Ann. Neurol. 25:351; Howard et al., 1989, J. Neurosurg. 71:105). In yet another embodiment, a controlled-release system can be placed in proximity of a target RNA of the compound or a pharmaceutically acceptable salt thereof, thus requiring only a fraction of the systemic dose.

20 The compounds described herein can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the active compound and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with
25 pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The invention includes methods for preparing pharmaceutical compositions for
30 modulating the expression or activity of a polypeptide or nucleic acid of interest. Such methods comprise formulating a pharmaceutically acceptable carrier with an agent that modulates

expression or activity of a polypeptide or nucleic acid of interest. Such compositions can further include additional active agents. Thus, the invention further includes methods for preparing a pharmaceutical composition by formulating a pharmaceutically acceptable carrier with an agent that modulates expression or activity of a polypeptide or nucleic acid of interest and one or more
5 additional active compounds.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Intravenous administration is preferred. Solutions or
10 suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates,
15 citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions
20 (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of
25 manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the
30 required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for

example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays
5 absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a polypeptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle
10 which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

15 Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and
20 expectorated or swallowed.

Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent
25 such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable propellant, *e.g.*, a gas
30 such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives.

5 Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal
10 delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid,
15 collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to
20 those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the individual to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired
25 therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

30 For antibodies, the preferred dosage is 0.1 mg/kg to 100 mg/kg of body weight (more preferably, 0.1 to 20 mg/kg, 0.1-10 mg/kg). Generally, partially human antibodies and fully

human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (*e.g.*, into the gastrointestinal system). A method for lipidation of antibodies is
5 described by Cruikshank et al. (1997, J. Acquired Immune Deficiency Syndromes and Human Retrovirology 14:193).

In a specific embodiment, an effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably
10 about 0.1 to 1.0 mg/kg, 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight.

The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat an individual, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the individual, and other diseases
15 present. Moreover, treatment of a individual with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

In addition to those compounds described above, the present invention encompasses the use of small molecules that modulate expression or activity of a nucleic acid or polypeptide of
20 interest. Non-limiting examples of small molecules include peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (*i.e.*, including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic
25 or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The
30 dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the individual or sample being treated, further depending upon the route by which

the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of individual or sample weight (*e.g.*, about 1 microgram per kilogram to about 500 milligrams per kilogram, 5 about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram). It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be 10 administered to an individual (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the 15 specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

20 (T) Kits

The present invention provides kits for measuring the expression of the protein and RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, or all or any combination of the biomarkers of the invention. Such kits comprise materials and reagents required for measuring the expression of 25 such protein and RNA products. In specific embodiments, the kits may further comprise one or more additional reagents employed in the various methods, such as: (1) reagents for purifying RNA from blood; (2) biomarker specific primer sets for generating test nucleic acids; (3) dNTPs and/or rNTPs (either premixed or separate), optionally with one or more uniquely labeled dNTPs and/or rNTPs (*e.g.*, biotinylated or Cy3 or Cy5 tagged dNTPs); (4) post synthesis labeling 30 reagents, such as chemically active derivatives of fluorescent dyes; (5) enzymes, such as reverse transcriptases, DNA polymerases, and the like; (6) various buffer mediums, *e.g.*, hybridization,

washing and/or enzymatic buffers; (7) labeled probe purification reagents and components, like spin columns, etc.; and (8) protein purification reagents; (9) signal generation and detection reagents, *e.g.*, streptavidin-alkaline phosphatase conjugate, chemifluorescent or chemiluminescent substrate, and the like. In particular embodiments, the kits comprise prelabeled
5 quality controlled protein and or RNA isolated from a sample (*e.g.*, blood) or synthesized for use as a control. In some embodiments kits can include a computer-readable medium which has a formula which uses data representing a level of products of at least one biomarker and generating an indication of the probability that a test subject has one or more colorectal pathologies including one or more polyps or one or more subtypes of polyps. The formula of the computer-
10 readable medium can be generated by using the methods outlined in section (G).

In some embodiments, the kits are PCR kits or Real time PCR kits and/or QRT-PCR kits. In other embodiments, the kits are nucleic acid arrays and protein arrays. Such kits according to the subject invention will at least comprise an array having associated protein or nucleic acid members of the invention and packaging means therefore. Alternatively the protein or nucleic
15 acid members of the invention may be prepackaged onto an array.

In one embodiment, the QRT-PCR kit includes the following: (a) two or more biomarker specific primer sets, each set used to amplify a biomarker within the combination of biomarkers of the invention; (b) buffers and enzymes including a reverse transcriptase; (c) one or more thermos table polymerases; and (d) Sybr® Green. In another embodiment, the kit of the
20 invention can include (a) a reference control RNA and/or(b) a spiked control RNA. In another embodiment, the kit also includes a computer readable medium which has a formula which uses data representing a level of products of at least one biomarker and generating an indication of the probability that a test subject has one or more colorectal pathologies including one or more polyps or one or more subtypes of polyps. The formula of the computer-readable medium can
25 be generated by using the methods outlined in section (G).

The invention provides kits that are useful for testing, detecting, screening, diagnosing, monitoring and prognosing one or more colorectal pathologies including one or more polyps or one or more subtypes of polyps. For example, in a particular embodiment of the invention a kit is comprised a forward and reverse primer wherein the forward and reverse primer are designed
30 to quantitate expression of all of the species of mRNA corresponding to a single distinct biomarker, where each of the distinct biomarkers is selected from the group identified in **Tables**

1, 2, 11, or 12. In certain embodiments, at least one of the primers of a primer set is designed to span an exon junction of a species of mRNA.

The invention includes kits that are useful for testing, detecting, screening, diagnosing, monitoring and prognosing one or more colorectal pathologies including one or more types or subtypes of polyps based upon the expression levels of protein or RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, or all or any combination of the biomarkers of the invention in a sample.

The invention includes kits useful for monitoring the efficacy of one or more therapies that an individual is undergoing based upon the expression of a protein or RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, or all or any combination of the biomarkers of the invention in a sample.

The invention includes kits using for determining whether an individual will be responsive to a therapy based upon the expression of a protein or RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, or all or any combination of the biomarkers of the invention in a sample.

The invention includes kits for measuring the expression of a RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, or all or any combination of the biomarkers of the invention in a sample. In a specific embodiment, such kits comprise materials and reagents that are necessary for measuring the expression of a RNA products of a biomarker of the invention. For example, a microarray or QRT-PCR kit may be produced for detecting one or more colon pathologies including polyps or one or more subtypes of polyps and contain only those reagents and materials necessary for measuring the levels of RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, or all or any combination of the biomarkers of the invention. Alternatively, in some embodiments, the kits can comprise materials and reagents that are not limited to those required to measure the levels of RNA products of 1, 2, 3, 4, 5, 6, 7, 8 or all or any combination of the biomarkers of the invention. For example, a microarray kit may contain reagents and materials necessary for measuring the levels of RNA products not necessarily associated with or indicative of one or more colorectal pathologies, in addition to reagents and materials necessary for measuring the levels of the RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, or all or any combination of the biomarkers of the invention. In a specific

embodiment, a microarray or QRT-PCR kit contains reagents and materials necessary for measuring the levels of RNA products of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, or all or any combination of the biomarkers of the invention, and 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, or more genes other than the biomarkers of the invention, or 1-10, 1-100, 1-150, 1-200, 1-300, 1-400, 1-500, 1-1000, 25-100, 25-200, 25-300, 25-400, 25-500, 25-1000, 100-150, 100-200, 100-300, 100-400, 100-500, 100-1000, 500-1000 other genes than the biomarkers of the invention.

For nucleic acid microarray kits, the kits generally comprise probes attached or localized to a support surface. The probes may be labeled with a detectable label. In a specific embodiment, the probes are specific for an exon(s), an intron(s), an exon junction(s), or an exon-intron junction(s), of RNA products of 1, 2, 3, 4, 5, 6, 7, 8, all or any combination of the biomarkers of the invention. The microarray kits may comprise instructions for performing the assay and methods for interpreting and analyzing the data resulting from the performance of the assay. In a specific embodiment, the kits comprise instructions for detecting or diagnosing one or more colorectal pathologies including one or more polyps or one or more subtypes of polyps. The kits may also comprise hybridization reagents and/or reagents necessary for detecting a signal produced when a probe hybridizes to a target nucleic acid sequence. Generally, the materials and reagents for the microarray kits are in one or more containers. Each component of the kit is generally in its own suitable container. In another embodiment, the kit also includes a computer readable medium which has a formula which uses data representing a level of products of at least one biomarker and generating an indication of the probability that a test subject has one or more colorectal pathologies or a subtype of colorectal pathology including a polyp or one or more subtypes of polyps. The formula of the computer-readable medium can be generated by using the methods outlined in section (G).

For QRT-PCR kits, the kits generally comprise pre-selected primers specific for particular RNA products (*e.g.*, an exon(s), an intron(s), an exon junction(s), and an exon-intron junction(s)) of 1, 2, 3, 4, 5, 6, 7, 8, or all or any combination of the biomarkers of the invention. The QRT-PCR kits may also comprise enzymes suitable for reverse transcribing and/or amplifying nucleic acids (*e.g.*, polymerases such as Taq), and deoxynucleotides and buffers needed for the reaction mixture for reverse transcription and amplification. The QRT-PCR kits

may also comprise probes specific for RNA products of 1,2, 3, 4, 5, 6, 7, 8, or all or any combination of the biomarkers of the invention. The probes may or may not be labeled with a detectable label (*e.g.*, a fluorescent label). Each component of the QRT-PCR kit is generally in its own suitable container. In another embodiment, the kit also includes a computer readable
5 medium which has a formula which uses data representing a level of products of at least one biomarker and generating an indication of the probability that a test subject has one or more colorectal pathologies or a subtype of colorectal pathology including a polyp or one or more subtypes of polyps. The formula of the computer-readable medium can be generated by using the methods outlined in section (G).

10 Thus, these kits generally comprise distinct containers suitable for each individual reagent, enzyme, primer and probe. Further, the QRT-PCR kits may comprise instructions for performing the assay and methods for interpreting and analyzing the data resulting from the performance of the assay. In a specific embodiment, the kits contain instructions for diagnosing or detecting one or more colorectal pathologies including one or more polyps or one or more
15 subtypes of polyps.

In a specific embodiment, the kit is a QRT-PCR kit. Such a kit may comprise a 96 well plate and reagents and materials necessary for SYBR Green detection. The kit may comprise reagents and materials so that beta-actin can be used to normalize the results. The kit may also comprise controls such as water, phosphate buffered saline, and phage MS2 RNA. Further, the kit
20 may comprise instructions for performing the assay and methods for interpreting and analyzing the data resulting from the performance of the assay. In a specific embodiment, the instructions state that the level of a RNA products of 1, 2, 3, 4, 5, 6, 7, 8, all or any combination of the biomarkers of the invention should be examined at two concentrations that differ by, *e.g.*, 5 fold to 10-fold.

25 For antibody based kits, the kit can comprise, for example: (1) a first antibody (which may or may not be attached to a support) which binds to protein of interest (*e.g.*, a protein products of 1, 2, 3, 4, 5, 6, 7, 8, all or any combination of the biomarkers of the invention); and, optionally, (2) a second, different antibody which binds to either the protein, or the first antibody and is conjugated to a detectable label (*e.g.*, a fluorescent label, radioactive isotope or enzyme).
30 The antibody-based kits may also comprise beads for conducting an immunoprecipitation. Each component of the antibody-based kits is generally in its own suitable container. Thus, these kits

generally comprise distinct containers suitable for each antibody. Further, the antibody-based kits may comprise instructions for performing the assay and methods for interpreting and analyzing the data resulting from the performance of the assay. In a specific embodiment, the kits contain instructions for diagnosing or detecting one or more colorectal pathologies including
5 one or more polyps or one or more subtypes of polyps. In another embodiment, the kit contains instructions for applying the data to a formula in the form of a computer readable medium which contains said instructions. Said computer readable medium can also contain instructions for interpreting the analyzing the data resulting from the performance of the assay.

(U) SNPs

10 A Single Nucleotide Polymorphism (SNP) is a single nucleotide variation at a specific location in the genome of different individuals. SNPs are found in both coding and non-coding regions of genomic DNA. In spite of the paucity of scorable phenotypes, SNPs are found in large numbers throughout the human genome (Cooper et al., Hum Genet 69:201-205, 1985). SNPs are stable genetic variations frequently found in genes, and contribute to the wide range of
15 phenotypic variations found in organisms. Single nucleotide polymorphisms (SNPs) can be of predictive value in identifying many genetic diseases, as well as phenotypic characteristics. It is known for example that certain SNPs result in disease-causing mutations such as the SNP correlated with heritable breast cancer (Cannon-Albright and Skolnick, Semin. Oncol. 23:1-5, 1996).

20 A SNP may be identified in the DNA of an organism by a number of methods well known to those of skill in the art, including but not limited to identifying the SNP by DNA sequencing, by amplifying a PCR product and sequencing the PCR product, by Oligonucleotide Ligation Assay (OLA), by Doublecode OLA, by Single Base Extension Assay, by allele specific primer extension, or by mismatch hybridization.

25 The instant invention offers a more focused and efficient method of screening SNPs to identify those SNPs which are specifically associated with one or more colorectal pathologies by having identified a selection of genes which are differentially expressed in blood from individuals having one or more colorectal pathologies as compared with individuals not having said one or more colorectal pathologies. In one aspect of the invention, a selection of SNPs to
30 be screened are those SNPs found in the genes listed in **Tables 2** and **6**. In another aspect of the

invention, novel SNPs can be identified in the disease-associated biomarkers using those methods listed above.

In particular, this invention focuses on methods for identifying those SNPs which are associated with one or more colorectal pathologies by screening only those SNPs in the biomarkers identified herein. Those SNPs which are identified using the methods disclosed herein will be convenient diagnostic markers.

More specifically a SNP is considered to be a polyp associated SNP, if those individuals having one or more colorectal pathologies have a different polymorphism at the SNP locus than those individuals not having the one or more colorectal pathologies. Further, a particular SNP is considered to be diagnostic for one or more colorectal pathologies if a particular polymorphism of the SNP is found to present at a statistically significant higher frequency in those individuals having one or more colorectal pathologies than in those individuals not having the one or more colorectal pathologies. Indices of statistical significance include $p < 0.05$, $p < .001$, $p < .01$, and $p < .10$. This invention includes methods of determining the diagnostic value of SNPs with respect to diagnosing or detecting one or more colorectal pathologies.

As would be understood, a preferred sample is blood, but these methods encompass any samples from which DNA can be obtained including epithelial cells, buccal cells, hair, saliva, tissue cells and the like. There are a variety of available methods for obtaining and storing tissue and/or blood samples. These alternatives allow tissue and blood samples to be stored and transported in a form suitable for the recovery of genomic DNA from the samples for genotype analysis. DNA samples can be collected and stored on a variety of solid mediums, including Whatmann paper, Guthrie cards, tubes, swabs, filter paper, slides, or other containers. When whole blood is collected on filter paper, for example, it can be dried and stored at room temperature.

The blood sample may be any one of various types of blood samples, including, for example, a sample of serum-depleted blood, a sample of erythrocyte-depleted blood, a sample of serum-depleted and erythrocyte-depleted blood, a sample of lysed blood, a blood sample which has not been fractionated into cell types and a sample of unfractionated cells of a lysed blood sample. Examples of blood samples are described in **Example 1** of the Examples section below.

In another aspect of the invention, polyp associated SNPs can be identified from RNA transcripts of the polyp biomarker genes, listed in **Tables 2** and **6**, instead of from genomic

DNA. In one embodiment, RNA is isolated from a sample such as blood, from individuals with and without the given disease or disorder, and transcripts encoded by these polyp biomarker genes are reversed transcribed into cDNA. The cDNA is amplified and analyzed to determine the presence of SNPs in the polyp biomarker genes. A polyp associated SNP, can be identified
5 by then comparing the distribution of each of the SNPs identified in the polyp associated biomarker gene(s) differentially expressed in those individuals having one or more colorectal pathologies and individuals who do not have one or more colorectal pathologies. In a further variation of this embodiment, instead analyzing cDNA for the presence of SNPs, the RNA transcripts of the disease specific biomarker genes, or their amplified products, are analyzed for
10 the presence of SNPs.

Analysis of genomic DNA comprising the polyp biomarker genes has the potential to identify SNPs in the coding region as well as in regulatory regions, the latter which may contribute to the change in expression levels of the gene. Analysis of cDNA encoded SNPs has the potential to identify only SNPs in the coding region of the polyp biomarker genes, which
15 may be instrumental in deciphering protein based mechanisms of polyp formation. Methods of analyzing cDNA encoded SNPs can be carried out by analyzing the cDNA generated in the reverse transcription PCR reactions described herein that are used to identify the level of the biomarker in samples from patients and non patients.

A polyp associated SNP may be identified in the DNA of the polyp biomarker genes by a
20 number of methods well known to those of skill in the art (see for example U.S. Pat. Nos. 6,221,592 and 5,679,524), including but not limited to identifying the SNP by PCR or DNA amplification, Oligonucleotide Ligation Assay (OLA) (Landegren et al., Science 241:1077, 1988), Doublecode OLA, mismatch hybridization, mass spectrometry, Single Base Extension Assay, (US6,638,722), RFLP detection based on allele-specific restriction-endonuclease
25 cleavage (Kan and Dozy, Lancet ii: 910-912, 1978), hybridization with allele-specific oligonucleotide probes (Wallace et al., Nucl Acids Res 6:3543-3557, 1978), including immobilized oligonucleotides (Saiki et al., Proc Natl Acad Sci USA 86:6230-6234, 1989) or oligonucleotide arrays (Maskos and Southern, Nucl Acids Res 21:2269-2270, 1993), allele-specific PCR (Newton et al., Nucl Acids Res 17:2503-16, 1989), mismatch-repair detection
30 (MRD) (Faham and Cox, Genome Res 5:474-482, 1995), binding of MutS protein (Wagner et al., Nucl Acids Res 23:3944-3948, 1995), single-strand-conformation-polymorphism detection

(Orita et al., Genomics 5:874-879, 1983), RNAase cleavage at mismatched base-pairs (Myers et al., Science 230:1242, 1985), chemical (Cotton et al., Proc Natl Acad Sci USA 85:4397-4401, 1988) or enzymatic (Youil et al., Proc Natl Acad Sci USA 92:87-91, 1995) cleavage of heteroduplex DNA, methods based on allele specific primer extension (Syvanen et al., Genomics 5 8:684-692, 1990), genetic bit analysis (GBA) (Nikiforov et al., Nuci Acids Res 22:4167-4175, 1994), and radioactive and/or fluorescent DNA sequencing using standard procedures well known in the art.

The instant methods of screening a subset of SNPs to identify polyp associated SNPs in polyp biomarker genes also encompass non-PCR methods of DNA. These methods include 10 ligase chain reaction ("LCR"), disclosed in European Patent Application No. 320,308, Qbeta Replicase, described in PCT Patent Application No. PCT/US87/00880, isothermal amplification methods, Walker et al. (Nucleic Acids Res 20(7):1691-6, 1992), Strand Displacement Amplification (SDA) described in U.S. Pat. Nos. 5,712,124, 5,648,211 and 5,455,166, Cyclic Probe Reaction, Transcription-Based Amplification, including nucleic acid sequence based 15 amplification (NASBA) and 3SR, Kwoh et al., Proc Natl Acad Sci USA, 86:1173-77, 1989; PCT Patent Application WO 88/10315 et al., 1989, other amplification methods, as described in British Patent Application No. GB 2,202,328, and in PCT Patent Application No. PCT/US89/01025, Davey et al., European Patent Application No. 329,822, Miller et al., PCT Patent Application WO 89/06700, "race and "one-sided PCR TM." described in Frohman, In: 20 PCR Protocols: A Guide To Methods And Applications, Academic Press, N.Y., 1990, methods based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide, described in Wu et al., Genomics 4:560-569, 1989.

While it is generally contemplated that the polymerase employed will be thermostable, non-thermostable polymerases may also be employed in the context of the present disclosure. 25 Exemplary polymerases and nucleic acid modifying enzymes that may be used in the context of the disclosure include the thermostable DNA Polymerases of OmniBase Sequencing Enzyme, Pfu DNA Polymerase, Taq DNA Polymerase, Taq DNA Polymerase, Sequencing Grade, TaqBead Hot Start Polymerase, AmpliTaq Gold, Vent DNA Polymerase, Tub DNA Polymerase, TaqPlus DNA Polymerase, Tfl DNA Polymerase, Tli DNA Polymerase, Tth DNA Polymerase; 30 the DNA Polymerases of DNA Polymerase I, Klenow Fragment, Exonuclease Minus, DNA Polymerase I, DNA Polymerase I Large (Klenow) Fragment, Terminal Deoxynucleotidyl

Transferase, T7 DNA Polymerase, T4 DNA Polymerase; the Reverse transcriptases of AMV Reverse Transcriptase and M-MLV Reverse Transcriptase; T4 DNA ligase and T4 polynucleotide kinase.

Recognition moieties incorporated into primers, incorporated into the amplified product
5 during amplification, or attached to probes are useful in the identification of the amplified molecules. A number of different labels may be used for this purpose such as, for example: fluorophores, chromophores, radio-isotopes, enzymatic tags, antibodies, chemiluminescence, electroluminescence, affinity labels, etc. One of skill in the art will recognize that these and other fluorophores not mentioned herein can also be used with success in this disclosure. Examples of
10 affinity labels include but are not limited to the following: an antibody, an antibody fragment, a receptor protein, a hormone, biotin, DNP, or any polypeptide/protein molecule that binds to an affinity label and may be used for separation of the amplified gene. Examples of enzyme tags include enzymes such as urease, alkaline phosphatase, or peroxidase. Additionally, colorimetric indicator substrates can be employed to provide a detection means visible to the human eye or
15 spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples. All these examples are generally known in the art and the skilled artisan will recognize that the present disclosure is not limited to the examples described above. The following fluorophores are specifically contemplated to be useful in the present disclosure: Alexa 350, Alexa 430, AMCA, BODIPY 630/650, BODIPY 650/665, BODIPY-FL, BODIPY-R6G,
20 BODIPY-TMR, BODIPY-TRX, Cascade Blue, Cy2, Cy3, Cy5, 6-FAM, Fluorescein, HEX, 6-JOE, Oregon Green 488, Oregon Green 500, Oregon Green 514, Pacific Blue, REG, Rhodamine Green, Rhodamine Red, ROX, TAMRA, TET, Tetramethylrhodamine, and Texas Red.

In the context of the present disclosure, it is specifically contemplated that the DNA amplification products of the disclosed methods may be analyzed using DNA chips or
25 microarrays in order to detect SNPs. The amplified DNA products may then be passed over a DNA chip or microarray encompassing oligonucleotide or polynucleotide probes. The ability or inability of the amplified DNA to hybridize to the microarray or DNA chip will facilitate the characterization of the SNPs present in the biomarker genes encoding the transcripts present in the sample.

30 The following non-limiting examples are illustrative of the present invention:

EXAMPLES

EXAMPLE 1**RNA isolation from unfractionated whole blood****(a) Centrifuged Lysed blood (Serum Reduced, Erythrocyte Reduced Blood Sample)**

Ten ml of peripheral whole blood was collected in EDTA Vacutainer tubes (Becton
5 Dickinson, Franklin Lakes, N.J.) and stored on ice until processing (within 6 hours). Upon
centrifugation, blood samples separated into plasma (including the buffy coat) and red blood cell
layers. The plasma was removed and a hypotonic buffer (1.6 mM EDTA, 10 mM KHCO₃, 153
mM NH₄Cl, pH 7.4) was added to lyse the red blood cells at a 3:1 volume ratio. The mixture
was centrifuged to yield a cell pellet, which was dissolved and homogenized into 1.0 ml of
10 TRIzol® Reagent (Invitrogen Corp., Carlsbad, CA) and 0.2 ml of chloroform according to the
manufacture's instructions. After centrifugation, isopropanol was added to the aqueous phase at a
1:1 ratio and allowed to precipitate at -20°C. Subsequent centrifugation yielded an RNA pellet
that was resuspended in water for experimental use. RNA quality was assessed on Agilent 2100
Bioanalyzer RNA 6000 Nano Chips as specified by the manufacturer, and RNA quantity was
15 determined by absorbance at 260 nm in a Beckman-Coulter DU640 Spectrophotometer.

(b) Lysed blood

10 ml whole blood is obtained in a Vacutainer or any smaller volume container desired.
Lysis Buffer is added directly to the blood sample (where the blood sample does not have the
serum removed) in a ratio of 3 parts Lysis Buffer to 1 part blood (Lysis Buffer (1L) 0.6g EDTA;
20 1.0g KHCO₂, 8.2g NH₄Cl adjusted to pH 7.4 (using NaOH)). Sample is mixed and placed on
ice for 5-10 minutes until transparent. Lysed sample is centrifuged at 1000 rpm for 10 minutes at
4°C, and supernatant is aspirated. Pellet is resuspended in 5ml Lysis Buffer, and centrifuged
again at 1000 rpm for 10 minutes at 4°C. Pelleted cells are homogenized using TRIzol
(GIBCO/BRL) in a ratio of approximately 6ml of TRIzol for every 10ml of the original blood
25 sample and vortexed well. Samples are left for 5 minutes at room temperature. RNA is
extracted using 1.2 ml of chloroform per 1 ml of TRIzol. Sample is centrifuged at 12,000 x g for
5 minutes at 4°C and upper layer is collected. To upper layer, isopropanol is added in ratio of
0.5 ml per 1 ml of TRIzol. Sample is left overnight at -20°C or for one hour at -20°C. RNA is
pelleted in accordance with known methods, RNA pellet air dried, and pellet resuspended in
30 DEPC treated ddH₂O. RNA samples can also be stored in 75% ethanol where the samples are
stable at room temperature for transportation.

(c) From serum reduced whole blood

10 ml whole blood is obtained in a Vacutainer and spun at 2,000 rpm (800g) for 5 min at 4°C and the plasma layer removed. The remaining blood sample is homogenized using TRIzol (GIBCO/BRL) in a ratio of approximately 6ml of TRIzol for every 10ml of the original blood sample and vortexed well. Samples are left for 5 minutes at room temperature. RNA is extracted using 1.2 ml of chloroform per 1 ml of TRIzol. Sample is centrifuged at 12,000 x g for 5 minutes at 4°C and upper layer is collected. To upper layer, isopropanol is added in ratio of 0.5 ml per 1 ml of TRIzol. Sample is left overnight at -20°C or for one hour at -20°C. RNA is pelleted in accordance with known methods, RNA pellet air dried, and pellet resuspended in DEPC treated ddH₂O. RNA samples can also be stored in 75% ethanol where the samples are stable at room temperature for transportation.

d) From whole blood

10 ml whole blood is obtained in a Vacutainer and the sample is homogenized using TRIzol (GIBCO/BRL) in a ratio of approximately 6ml of TRIzol for every 10ml of the blood sample and vortexed well. Samples are left for 5 minutes at room temperature. RNA is extracted using 1.2 ml of chloroform per 1 ml of TRIzol. Sample is centrifuged at 12,000 x g for 5 minutes at 4°C and upper layer is collected. To upper layer, isopropanol is added in ratio of 0.5 ml per 1 ml of TRIzol. Sample is left overnight at -20°C or for one hour at -20°C. RNA is pelleted in accordance with known methods, RNA pellet air dried, and pellet resuspended in DEPC treated ddH₂O. RNA samples can also be stored in 75% ethanol where the samples are stable at room temperature for transportation.

(e) From whole blood using PAXgene™

2.5 ml whole blood is collected into PAXgene™ Blood RNA Tubes and processed in accordance with the instructions of the PAXgene™ Blood RNA Kit protocol. In brief after storing the blood in the PAXgene™ tube for at least 2 hours, the blood sample is centrifuged and the supernatant discarded. To the remaining sample, 360 µl of the supplied Buffer BR1 is added and the sample is pipetted into the spin column and centrifuged and then washed with numerous wash steps and finally eluted and stored.

(f) From whole blood using PAXgene™ and subsequent globin reduction

RNA isolated from PAXGene™ as noted in (d) above is subsequently treated to selectively remove globin mRNA as is described in Affymetrix® technical note entitled "Globin

Reduction Protocol". Oligonucleotides specific for the alpha 1, alpha 2 and beta globin species are incubated with an oligonucleotide hybridization buffer and RNase H used to specifically target degradation of the globin mRNA and the cRNA clean up column from Affymetrix used to remove the globin mRNA.

5 EXAMPLE 2

Target nucleic acid preparation and hybridization

(a) Genes which are differentially expressed with the presence of One or more colorectal pathologies

Total RNA (5 µg) was labeled and hybridized onto Affymetrix U133Plus 2.0 GeneChips
10 (Affymetrix; Santa Clara, CA) along with other similarly prepared samples from individuals having or not having polyps and hybridized according to the manufacturer's instructions. Briefly, the five µg total RNA was used for cDNA synthesis with GeneChip T7-Oligo(dT) primer provided in the promoter primer kit (Affymetrix, P/N 900375). cDNA was cleaned up with cDNA Cleanup Spin Column and then subjected to synthesis of Biotin-Labeled cRNA with
15 Enzo®BioArray™High Yield™ RNA Transcript Labeling Kit. The labeled cRNA was further purified using the IVT cTNA Cleanup Spin Column and quantified using spectrophotometer with absorbance at 260 nm. 20 µg cRNA was then added into the hybridization cocktail and the cocktail was applied to the probe array cartridge. After approximately 16 hours hybridization, the array was washed with Affymetrix fluidics station 400. The array was then scanned with
20 Affymetrix®GeneChip®Scanner.

Hybridization signals were scaled in the Affymetrix GCOS software (version 1.1.1), using a scaling factor determined by adjusting the global trimmed mean signal intensity value to 500 for each array, and imported into GeneSpring version 7.2 (Silicon Genetics; Redwood City, CA). Signal intensities were then centered to the 50th percentile of each chip, and for each
25 individual gene, to the median intensity of the whole sample set. Only genes called present or marginal by the GCOS software in at least 80% of each group of samples were used for further analysis. Differentially expressed genes were identified using 1) the non-parametric Wilcoxon-Mann-Whitney non-parametric test ($P < 0.05$), 2) parametric t test ($P < 0.05$), and/or the 3) unsupervised analysis method (14). In the un-supervised analysis, the signal-intensity filtered
30 genes were used to select genes with at least 2-fold change (up or down) in expression level, away from the mean, in at least 15% of the samples. Hierarchical cluster analysis was performed

on each comparison to assess correlation analysis using Spearman correlation among samples for each identified gene set as the similarity measure with average centroid linkage in GeneSpring v6.0. Results from numerous experiments were analyzed and a compiled list of results provided in Table 1.

5 **(b) Genes which are differentially expressed with the presence of High Risk Polyps**

Total RNA was isolated from centrifuged lysed blood (ie serum reduced, erythrocyte reduced blood) as described in Example 1 from patients diagnosed with having high risk polyps. 1 µg of Oligo-dT primers were annealed to 10 µg of total RNA for each individual tested in a total volume of 10 µl, by heating to 70°C for 10 min, and cooled on ice. Individuals were
10 diagnosed as having one or more of the high risk polyp subtypes (colorectal pathology during a colonoscopy identified the polyp identified as one or more of the following types: Tubulovillous Adenoma; Villous Adenoma; Cancer; High Grade Dysplasia; and Tubular Adenoma wherein the diameter of the Tubular Adenoma is greater than 10mm). Procedures were otherwise carried out as described in Example 2(a) above.

15 Results from numerous experiments were analyzed and a compiled list of results provided in Table 11.

EXAMPLE 3

Quantitative real time PCR (QRT-PCR)

QRT-PCR was performed on a selection of the genes in Table 1 which are disclosed in
20 Table 2. QRT-PCR was done using either the SYBR® Green Kit from Qiagen (Product Number 204143) and/or using Applied Biosystems PCR reagent kits (Cat 4334973). Amplicons were detected in real time using a Prism 7500 instrument (Applied Biosystems).

Reverse transcription was first performed using the High-Capacity cDNA Archive Kit from Applied Biosystems (Product number 4322171) and following the protocol utilized therein.

25 More specifically purified RNA as described previously herein was incubated with reverse transcriptase buffer, dNTPs, random primers and reverse transcriptase and incubated for 25°C for 10 minutes and subsequently for 37 °C for two hours and the resulting mixture utilized as the starting product for quantitative PCR.

cDNA resulting from reverse transcription was incubated with the QuantiTect SYBR®
30 Green PCR Master Mix as provided and no adjustments were made for magnesium concentration. Uracil-N-Glycosylase was not added. 5µM of both forward primer and reverse

primer specific to the genes of the invention were added and the reaction was incubated and monitored in accordance with the standard protocol utilizing the ABI PRISM 7700/ABI GeneAmp 5700/iCycler/DNA Engine Opticon. Primers utilized are shown in **Table 6**. Other examples of primers which can be used are disclosed in **Table 4**. Forward and reverse primers
5 for the candidate biomarkers were designed using "PrimerQuest". A tool which is available from Integrated DNA Technologies, Coralville, IA. **Table 6** lists the primer sets for eight of the genes of **Table 2**, namely, MBTPS1 (membrane-bound transcription factor protease site 1), MGC45871, MKLN1 (muskelin 1), NIPBL (Nipped-B homolog (Drosophila)), APEH (acylpeptide hydrolase), FLJ23091, MGC40157, and PPP1R2 (protein phosphatase 1, regulatory
10 (inhibitor) subunit 2). Serial dilution measurements for the target genes and a housekeeping gene (beta-actin, ACTB) were assayed, to ensure that the values were within linear range and the amplification efficiency was approximately equal for the target and ACTB. ACTB was selected as a housekeeping gene because no statistical significant differences were observed between control and disease group in this study (data not shown). The melting temperature [T_m] in
15 thermal dissociation, and examination on agarose gels provided confirmation of specific PCR amplification and the lack of primer-dimer formation in each reaction well.

For individual target gene analysis, changes in Ct value between each gene and the ACTB house-keeping was calculated as $\Delta Ct = Ct(\text{target gene}) - Ct(\text{house-keeping gene})$.

EXAMPLE 4

20 **TaqMan**[®] QRT PCR can also be performed using the QuantiTect[™] Probe RT-PCR system from Qiagen (Product Number 204343) in conjunction with a TaqMan[®] dual labelled probe and primers corresponding to the gene of interest. The TaqMan[®] probe and primers can be ordered from Applied Biosystems Assays-On-Demand[™].

The dual labelled probe contains both a fluorophore and a quencher molecule. The
25 proximity of the fluorescent reporter with the quencher prevents the reporter from fluorescing, but during the PCR extension step, the 5' – 3' exonuclease activity of the Taq DNA polymerase releases the fluorophore which allows it to fluoresce. As such, the amount of fluorescence correlates with the amount of PCR product generated. Examples of TaqMan[®] probes which can be utilized with the genes disclosed in **Table 2** are shown in **Table 6** and/or 4.

30 **EXAMPLE 5**

Identification of combinations of biomarkers using logistic regression

A selection of eight genes from **Table 2** were chosen as follows: MBTPS1, MGC45871, MKLN1, NIPBL APEH, MGC40157, PPP1R2, FLJ23091. Combinations of pairs of the selected eight genes were tested to determine the ability of each combination of pairs to screen for one or more colorectal pathologies. Blood samples were drawn into into lavender-top BD
 5 Vacutainer tube prior to anaesthesia and prior to any surgery. RNA from whole blood was prepared by first removing the serum and subsequently treating the remaining sample with hypotonic Lysis Buffer (Lysis Buffer (1L) 0.6g EDTA; 1.0g KHCO₂, 8.2g NH₄Cl adjusted to pH 7.4 (using NaOH)) in a ratio of 3 parts Lysis Buffer to 1 part blood so as to preferentially lyse the red blood cells. The samples were centrifuged and RNA extracted from the unfractionated cells
 10 of the sample.

Real-time quantitative PCR was conducted on 68 patients diagnosed as having colorectal pathology (ie one or more subtypes of polyps) (n=68) and 110 control individuals where the control individuals were diagnosed as not having one or more colorectal pathologies (n=110). QRT-PCR was performed using a two step procedure whereby cDNA was first prepared before
 15 performing PCR using ABI reagents. In this example, a matrix containing the ΔC_t of "ratios" of the eight biomarkers was used to create a reference training data set (AJ36h). The ratios used to generate the matrix of ΔC_t s was further constrained by requiring each ratio to be comprised of one upregulated gene and a second downregulated gene where the ΔC_t was generated by subtracting the C_t of the downregulated gene from the C_t of the upregulated gene. MBTPS1,
 20 MGC45871, MKLN1, NIPBL were identified as upregulated genes (ie when comparing polyps vs non polyp individuals) (**FIG. 3**). APEH, MGC40157, PPP1R2, FLJ23091 were identified as down regulated genes (**FIG. 3**). The advantages of gene-pair ratio method include a) reducing inter-individual variability, b) permitting analysis of individual samples without references so as to ensure technical differences between plates are minimized, c) can use any reliable method
 25 (microarray, real-time PCR, etc), d) independent of the platform utilized for data acquisition, e) no housekeeping gene required (relatively independent of the input of sample amount), f) translating the strengths of micorarray expression profiling into simple clinical tests, g) highly precise in disease discrimination.

A reference (training) data set (AJ36h) was constructed containing ΔC_t values for each
 30 possible ratio as described and constrained as noted above for the above eight genes assayed against a total of 178 subjects including 110 subject without pathology (Female/male: 55/54 with

one missing information, age average 57 year ranging 23 to 83 years of age) and 68 subjects with diagnosed colorectal pathology (Female/male: 22/45 with one missing information, age average 57 ranging from 38 to 82 years). The types of pathology identified include 21 (31%) tubular adenomas, 18 (27%) hyperplastic and 7 (10%) high risk pathology (villous morphology) and 22 (32%) other minor polyp (see Table 7).

Logistic regression (15-18) was used to analyze the dependence of the binary variable Y (0=control (has no pathology), 1=disease (has pathology) on all possible combinations of the ΔCt values from the reference data training set AJ36h. If P = probability that a patient sample is diagnosed as “diseased”, then a function $X = \text{Logit}(p)$ can be defined as follows:

$$X = \text{Logit}(P) = \ln(P/(1-P)) = b_0 + b_1 \Delta Ct_1 + b_2 \Delta Ct_2 + \dots + b_n \Delta Ct_n \quad (\text{Eq 1})$$

If $X \geq \text{threshold}$ then $Y=1$ (diagnosis = “diseased”), and if $X < \text{threshold}$ then $Y=0$ (diagnosis = “control”). The (empirical) coefficients $\{b_i\}$ that define the relationship between X and the experimental measurements $\{\Delta Ct_i\}$, where i represents a sample} were obtained by a maximum-likelihood (ML) fitting method. Identical $\{b_i\}$ values were obtained using several different commercial software programs: MedCalc (MedCalc Software, Mariakerke, Belgium) and XL-Stat (Addinsoft Software, Paris, France). ROC curve analysis was then used to evaluate the discriminatory power of each combination of ratios wherein all combinations of ratios as described above were tested (19-21). Each combinations of ratios resulted in an equation in the form of Equation 1. The top 10 best equations that gave an ROC = 0.72 were used in a formula where each equation was given equal ranking to perform the subsequent blind test prediction.

Cross Validation

Cross-validation was performed on dataset AJ36h (the 178-sample training set) using WEKA (www.cs.waikato.ac.nz/~ml/weka/index.html) (22). Two different cross-validation schemes were used. The WEKA MetaAnalysis function was used to construct 100 bootstrap replicates of dataset AJ36h. Each of the new datasets was analyzed with the SimpleLogistic classifier. Each of the resultant 100 logistic equations was then analyzed by 10-fold cross-validation. The results for all equations were averaged (“bootstrap aggregating”).

Prospective (Blind) Test

A blind set of 80 clinical samples were tested. The test set consisted of 40 controls and 40 subjects with one or more colorectal pathologies (having one or more polyps of any type). None of the test subjects used in the blind test were used to generate a classifier for each possible

combination of the ratio of biomarkers “ratios” where each ratio is selected as a combination of an upregulated gene (when comparing colorectal pathology to no colorectal pathology) and a down regulated gene (when comparing colorectal pathology to no colorectal pathology). MBTPS1, MGC45871, MKLN1, NIPBL were identified as upregulated genes (ie when
 5 comparing polyps vs non polyp individuals (**FIG. 3**). APEH, MGC40157, PPP1R2, FLJ23091 were identified as down regulated genes (**FIG. 3**).

Medical information on subjects in this blind set, including age, gender, and pathologist’s report, is summarized in **Table 7**. Samples were run on a single 96-well plate for Q RT-PCR with each of the eight genes measured in triplicate.

10 The measured values for each blind sample were evaluated using the following algorithm, which consists of an initial calculation, a binary logic gate, and a “committee machine” vote. Initial Calculation: The logit function X_i is computed for each of the equations ($i = 1$ to N) that gave ROC Area = 0.72 against the reference data set AJ36h. Logic Gate: If $X_i <$ threshold of j (where j stands for an equation number), then the blind sample is given score $S_i = -$
 15 1 for Eqn # j . If $X_i \geq$ threshold of j , then it is given score $S_i = +1$ for Eqn # j . Vote by Committee Machine (23): A vote is taken over the scores from the 10 logit equations used for diagnosis. By definition, $\text{Vote} = \sum S_i$. If $\text{Vote} \leq 0$ then the sample is called “no pathology”, while if $\text{Vote} > 0$ then the sample is called “diseased or with pathology”.

20 The best equation gives an ROC area 0.72 as shown in **FIG. 4**. Ten top equations that gave ROC=0.72 were used for the blind test prediction. **Table 8** lists the parameters for the 10 equations.

Cross-validation was done with the SimpleLogistic function in WEKA. 100 bootstrap replicates of dataset AJ36h were constructed and analyzed each with the WEKA SimpleLogistic function. The resultant set of 100 logistic equations was subjected to 10-fold cross-validation.
 25 The results for all equations were then averaged which gave an ROC Area = 0.66, overall accuracy 65%, sensitivity (TPF) 41%, specificity (TNF) 83%.

Prediction of Blind Samples

A blind test was conducted on an additional set of 80 samples (40 samples without pathology and 40 samples with colorectal pathology). A committee vote was taken over the ten
 30 (10) best logit equations from Reference Dataset AJ36h, to predict a state of either “colorectal pathology present” or “colorectal pathology absent (control)”. For the set of blind samples the

REPLACEMENT SHEET

sensitivity (true positive fraction, TPF) is 43 % (17/40), and the specificity (true negative fraction, TNF) is 80 % (32/40), with an overall accuracy of 61 % (49/80, Table 9).

While the present invention has been described with reference to what are presently considered to be the preferred examples, it is to be understood that the invention is not limited to the disclosed examples. To the contrary, the invention is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

All publications, patents and patent applications are herein incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

10 **EXAMPLE 6**

Testing of combinations of biomarkers using logistic regression and application of said combinations to screen for one or more types of colorectal pathology or one or more pathologies.

A selection of seven genes were chosen including: LIM domain containing preferred translocation partner in lipoma (LPP) Gene ID 4026; cytidine deaminase (CDA) Gene ID 978; sarcoma antigen NY-SAR-48 (MGC20553) Gene ID 93323; serine (or cysteine) proteinase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 2 (SERPINE 2) Gene ID 5270; B-cell novel protein 1 (BCNP1) Gene ID 199786; hypothetical protein MGC45871 (MGC45871) Gene ID 359845; membrane-bound transcription factor protease, site 1 (MBTPS1) Gene ID 8720. Genes were chosen from either Table 1 or other similar experiments.

Clinical Question - Polyp v. No Pathology

A reference (training) data set was constructed containing ΔC_t values for the above seven genes assayed against a total of 185 subjects having any type of polyp and 239 subjects not having polyps.

Logistic regression was used to analyze the dependence of the binary variable Y (0=control, 1=disease) on the ΔC_t values from the reference data set. If P = probability that a patient sample is identified as "diseased", then a function $X = \text{Logit}(p)$ can be defined as follows:

30
$$X = \text{Logit}(P) = \ln(P/(1-P)) = b_0 + b_1 \Delta C_{t1} + b_2 \Delta C_{t2} + \dots + b_n \Delta C_{tn} \quad (\text{Eq 1})$$

If $X \geq \text{threshold}$ then $Y=1$ (diagnosis = “has polyps”), and if $X < \text{threshold}$ then $Y=0$ (diagnosis = does not have polyps). The (empirical) coefficients $\{b_i\}$ that define the relationship between X and the experimental measurements $\{\Delta Ct_i\}$, where i represents a sample, were obtained by a maximum-likelihood (ML) fitting method. Identical $\{b_i\}$ values were obtained using several different commercial software programs: MedCalc (MedCalc Software, Mariakerke, Belgium) and XL-Stat (Addinsoft Software, Paris, France). ROC curve analysis was then used to evaluate the discriminatory power of the combinations. The best equation used the following genes (LPP; CDA; MBTPS1; SERPINE2; BCNP1) and shown as follows:

$X = \text{Logit}(P) = \ln(P/(1-P)) = -4.9104 - 0.4278 \Delta Ct_{\text{MGC20553}} - 0.6164 \Delta Ct_{\text{CDA}}$
 $+ 0.8230 \Delta Ct_{\text{MBTPS1}} + 0.3961 \Delta Ct_{\text{SERPINE2}} + 0.1641 \Delta Ct_{\text{BCNP1}}$ gave an ROC = 0.73 with a sensitivity of 80% and a specificity of 50%.

Clinical Question - High Risk Pathology v. Other

Using the same selection of seven biomarkers, but this time tested against a reference (training) data set with ΔCt values for the above seven genes assayed against a total of 129 subjects having “high risk” polyp and 295 subjects either having a polyp not classified as high risk or having no pathology, where subjects were classified as having “high risk polyps” if they had one of the following categories of polyps: Tubulovillous Adenoma; Villous Adenoma; High Grade Dysplasia and Tubular Adenoma where the diameter of the Tubular Adenoma polyp is greater than 10mm and cancer

Logistic regression was used to analyze the dependence of the binary variable Y (0=control, 1=disease) on the ΔCt values from the reference data set. If P = probability that a patient sample is identified as “diseased”, then a function $X = \text{Logit}(p)$ can be defined as follows:

$$X = \text{Logit}(P) = \ln(P/(1-P)) = b_0 + b_1 \Delta Ct_1 + b_2 \Delta Ct_2 + \dots + b_n \Delta Ct_n \quad (\text{Eq 1})$$

If $X \geq \text{threshold}$ then $Y=1$ (diagnosis = “has high risk polyps”), and if $X < \text{threshold}$ then $Y=0$ (diagnosis = does not have high risk polyps). The (empirical) coefficients $\{b_i\}$ that define the relationship between X and the experimental measurements $\{\Delta Ct_i\}$, where i represents a sample, were obtained by a maximum-likelihood (ML) fitting method. Identical $\{b_i\}$ values were obtained using several different commercial software programs: MedCalc (MedCalc

Software, Mariakerke, Belgium) and XL-Stat (Addinsoft Software, Paris, France). ROC curve analysis was then used to evaluate the discriminatory power of the combinations. The best equation used the following genes (CDA; MGC20553; MBTPS1; SERPINE2; BCNP1;) and shown as follows:

$$X = \text{Logit}(P) = \ln(P/(1-P)) = -5.2981 - 0.5433\Delta\text{CtCDA} - 0.4958\Delta\text{CtMGC20553} + 0.8551\Delta\text{CtMBTPS1} + 0.3554\Delta\text{CtBCNP1} + 0.2438\Delta\text{CtSERPINE2}$$

gave an ROC = 0.74 with a sensitivity of 83% and a specificity of 46%.

Clinical Question— Cancer v. Other

Finally the same seven genes were tested in combinations using a reference (training) data set containing ΔCt values for the above seven genes assayed against a total of 80 subjects having cancerous polyps and 344 subjects having other types of polyps or having no pathology.

Logistic regression was used to analyze the dependence of the binary variable Y (0=control, 1=disease) on the ΔCt values from the reference data set. If P = probability that a patient sample is identified as “diseased”, then a function $X = \text{Logit}(p)$ can be defined as follows:

$$X = \text{Logit}(P) = \ln(P/(1-P)) = b_0 + b_1 \Delta\text{Ct}_1 + b_2 \Delta\text{Ct}_2 + \dots + b_n \Delta\text{Ct}_n \quad (\text{Eq 1})$$

If $X \geq \text{threshold}$ then $Y=1$ (diagnosis = “has cancerous polyps”), and if $X < \text{threshold}$ then $Y=0$ (diagnosis = does not have cancerous polyps). The (empirical) coefficients $\{b_i\}$ that define the relationship between X and the experimental measurements $\{\Delta\text{Ct}_i\}$, where i represents a sample} were obtained by a maximum-likelihood (ML) fitting method. Identical $\{b_i\}$ values were obtained using several different commercial software programs: MedCalc (MedCalc Software, Mariakerke, Belgium) and XL-Stat (Addinsoft Software, Paris, France). ROC curve analysis was then used to evaluate the discriminatory power of the combinations. The best equation used the following genes (MGC20553, CDA, MBTPS1, SERPINE2, MGC45871; BCNP1) and shown as follows:

$$X = \text{Logit}(P) = \ln(P/(1-P)) = -12.9149 - 0.5378\Delta\text{CtCDA} - 0.5398\Delta\text{CtMGC20553} + 1.0386\Delta\text{CtMBTPS1} + 0.7405\Delta\text{CtBCNP1} + 0.4002\Delta\text{CtMGC45871} + 0.2074\Delta\text{CtSERPINE2}$$

gave an ROC = 0.83 with a sensitivity of 90% and a specificity of 55%.

Clinical Question - High Risk Pathology v. Other

A reference (training) data set was constructed containing ΔC_t values for the above seven genes assayed against a total of 252 subjects having high risk polyps and 272 subjects having other types of polyps or having no pathology where by high risk polyps is meant Tubulovillous Adenoma, Villous Adenoma, High Grade Dysplasia and Tubular Adenoma – regardless of diameter of polyp.

Logistic regression was used to analyze the dependence of the binary variable Y ($0=\text{control}$, $1=\text{disease}$) on the ΔC_t values from the reference data set. If $P =$ probability that a patient sample is identified as “diseased”, then a function $X = \text{Logit}(p)$ can be defined as follows:

$$X = \text{Logit}(P) = \ln(P/(1-P)) = b_0 + b_1 \Delta C_{t1} + b_2 \Delta C_{t2} + \dots + b_n \Delta C_{tn} \quad (\text{Eq 1})$$

If $X \geq \text{threshold}$ then $Y=1$ (diagnosis = “has high risk polyps”), and if $X < \text{threshold}$ then $Y=0$ (diagnosis = does not have high risk polyps). The (empirical) coefficients $\{b_i\}$ that define the relationship between X and the experimental measurements $\{\Delta C_{ti}$, where i represents a sample} were obtained by a maximum-likelihood (ML) fitting method. Identical $\{b_i\}$ values were obtained using several different commercial software programs: MedCalc (MedCalc Software, Mariakerke, Belgium) and XL-Stat (Addinsoft Software, Paris, France). ROC curve analysis was then used to evaluate the discriminatory power of the combinations. The best equation used the following genes (MGC20553; CDA; MBTPS1; SERPINE2; BCNP1;) and shown as follows:

$$X = \text{Logit}(P) = \ln(P/(1-P)) = -3.5819 - 0.7625\Delta C_{tCDA} - 0.6117\Delta C_{tMGC20553} + 1.15\Delta C_{tMBTPS1} + 0.2174\Delta C_{tBCNP1} + 0.2439\Delta C_{tSERPINE2}$$

gave an ROC = 0.76 with a sensitivity of 82% and a specificity of 52%.

EXAMPLE 7

Biomarkers to screen for presence of colorectal cancer. Deriving classifiers to be used with combinations of biomarkers and application of said classifiers to determine presence of colorectal cancer

QRT-PCR was performed on a selection of the genes identified from one or more microarray analyses (data not shown) as being able to differentiate as between individuals having colorectal cancer and individuals not having colorectal cancer. Some of genes selected for QRT-PCR were selected from microarray data performed on samples collected from three regions of

China for over 593 samples including 61 samples from individuals having been diagnosed with colorectal cancer and 532 from individuals not having colorectal cancer (data not shown). Other genes were selected from other similar microarray experiments with individuals from North America and/or Asia. Among the individuals not having colorectal cancer was a mixture of
 5 individuals having breast cancer, kidney cancer, prostate cancer, bladder cancer, and individuals having other subtypes of colorectal pathology which were not colorectal cancer.

QRT-PCR was done on each individual gene across a population of individuals having colorectal cancer and a population of individuals not having colorectal cancer. QRT-PCR experiments were done using either the SYBR® Green Kit from Qiagen (Product Number
 10 204143) and/or using Applied Biosystems PCR reagent kits (Cat 4334973) and/or using TaqMan Assay using the QuantiTect® Probe PCR Kit (Qiagen, Cat. # 204345). TaqMan® probes were developed for each gene of interest and labelled with FAM and the Black Hole Quencher® from Biosearch Technologies. Beta-actin was used as a housekeeping gene in the duplexed assays and labelled using HEX and Black Hole Quencher®. Amplicons were detected in real time using a
 15 Prism 7500 instrument (Applied Biosystems). Results of the QRT-PCR for each gene across the population tested are shown in **Table 12**.

Rather than testing all possible combinations of the biomarkers noted in Table 12, in other embodiments, all possible combinations of biomarkers with a p value of less than 0.05 can be chosen and all or a portion of combinations of biomarkers tested. Discussed below is
 20 representative classifiers identified for selected combinations tested.

Classifiers were derived for all two gene combinations of the biomarkers identified in **Table 12** using QRT-PCR for 28 of the genes identified in **Table 12** across 58 individuals having colorectal cancer and 57 individuals not having colorectal cancer. The 28 genes utilized are represented by the following gene symbols and are described in more detail herein:
 25 **OSBPL10, LOC283130, BANK1, COBLL1, MGC24039, C9orf85, BLNK, BCNP1, PDE3B, AKAP13, WDFY1, CDA, AGTRAP, ACTR2, UTS2, MS4A1, SPAP1, ANK3, KIAA1559, GBP5, MGC20553, CEACAM1, HIST1H4H, PRG1, BRD2, LTBP3, MAP4K3, and NIPA2**
 Primers utilized for the real time RT-PCR are further described in Table 16. Classifiers derived for selection combinations of two genes are shown as follows:

30 **Table A:** Table showing Resulting Classifiers for a Selection of Two Gene Combinations
 Wherein the Combinations are in the Form of Ratios and the Biomarkers Are Selected from the

28 Genes Selected from **Table 12**. Shown is the resulting ROC area of the Classifiers, the Sensitivity (at the noted cutoff Sens) and Specificity (at the noted cutoff Spec) are shown. The constant, and the coefficient for the selected two gene ratio are noted.

Table A

	ROCarea	Sensitivity at 90% Specificity	Specificity at 90% Sensitivity	Constant	Coeffic for Ratio
BANK1/CDA	0.8621	46.55	68.42	-3.7808	1.256
BCNP1/CDA	0.8451	43.10	57.89	-2.7251	1.2554
CDA/MS4A	0.8430	43.10	63.16	-1.3171	-1.0944
C9orf85/CDA	0.8382	44.83	52.63	-16.482	2.2374
CDA/OSBPL10	0.8367	56.90	47.37	-5.9699	-1.2875
CDA/SPAP1	0.8364	48.28	59.65	-3.3047	-0.9384
BLNK/CDA	0.8361	50.00	68.42	-4.8603	1.2367
CDA/LOC283130	0.8339	67.24	52.63	-3.9388	-1.6368
CDA/COBLL1	0.8321	53.45	56.14	-6.9371	-1.3126
BANK1/PRG1	0.8306	55.17	54.39	-8.6282	1.2825
BANK1/CEACAM1	0.8291	56.9	56.14	-1.7783	1.1225
BCNP1/PRG1	0.8224	60.34	45.61	-7.4478	1.2631
BANK1/MGC20553	0.8197	44.83	59.65	2.2412	0.9436
BANK1/NIPA2	0.8176	58.62	50.88	-2.2946	1.6015
ACTR2/BANK1	0.8155	41.38	54.39	-9.1667	-1.4596
CDA/HIST1H4H	0.8134	51.72	47.37	1.6688	-1.5062
AKAP13/GBP5	0.7012	22.41	22.81	-0.7784	0.6743
BCNP1/CEACAM1	0.8131	44.83	52.63	-0.8916	1.1843
CEACAM1/SPAP1	0.8128	34.48	56.14	-1.8563	-0.88
CDA/MAP4K3	0.8128	39.66	45.61	-9.7566	-1.7304
BCNP1/GBP5	0.8107	36.21	57.89	-4.611	1.1411

5

As noted for each two gene combination the ROC area is provided in addition to the specificity (when sensitivity is set at 90%) and sensitivity (when specificity is set at 90%). The

cutoff utilized to generate the sensitivity and specificity as noted are provided in the two righthand most columns. The classifier for each of the two gene combinations can be generically described as follows:

$$X(\text{having colorectal cancer}) = \text{Logit}(P) = \ln(P/(1-P)) = b_0 + b_1(\Delta Ct_1 - \Delta Ct_2)$$

- 5 Where b_0 is identified as the coefficient and b_1 is noted as the coefficient for the ratio.

EXAMPLE 8

Classifiers were also derived for all possible combinations of a selection of nine of the biomarkers identified in **Table 12**. Data for use in generating the classifiers for the combinations were obtained using real time RT-PCR for each of the nine genes tested across the same 58
10 individuals having colorectal cancer and 57 individuals not having colorectal cancer described in **Example 7**. The 9 genes utilized are represented by the following gene symbols and are described in more detail herein: **CDA, MGC20553, MS4A1, BCNP1, BANK1, GBP5, OSBPL10, SPAP1, LOC283130**. Primers utilized for the real time RT-PCR are further described in **Table 16**. A selection of the resulting classifiers are described below in **Table B**
15 below:

Table B

#	ROC area	Constant	CD A	MGC205 53	MS4A 1	BCNP 1	BANK 1	GBP 5	OSBPL 10	SPAP 1	LOC2831 30
1	0.8364	-14.48				1.61					
1	0.8197	-10.07			1.25						
1	0.8091	-10.56								1.02	
2	0.8681	-21.35				1.39			0.77		
2	0.8648	-7.58	0.93				1.41				
2	0.8594	-21.80					1.27		0.81		
3	0.8839	-12.85	1.26			1.20			0.93		
3	0.8811	-13.04	1.25				1.06		0.97		
3	0.8757	-10.48	1.20			1.11					0.94
4	0.8963	-9.09	1.47			1.34		0.72	0.93		
4	0.8917	-7.25	1.29	-0.51		1.17			1.02		
4	0.8902	-5.32	1.59			1.11		0.88			1.14

5	0.898	-10.60	1.42			0.99	0.43	0.71	0.92		
5	0.898	-8.84	1.63			1.16		0.80	0.75		0.52
5	0.896	-9.34	1.65				0.94	0.72	0.69		0.75
6	0.903	-10.25	1.57			0.82	0.43	0.78	0.73		0.52
6	0.900	-10.91	1.42		-0.32	1.13	0.62	0.72	0.91		
6	0.900	-10.64	1.41			1.19	0.57	0.71	0.88	-0.26	
7	0.905	-10.60	1.57		-0.31	0.96	0.61	0.79	0.73		0.51
7	0.902	-10.31	1.55			0.98	0.53	0.77	0.71	-0.20	0.49
7	0.901	-11.32	1.59	0.14		0.82	0.46	0.87	0.71		0.52
8	0.905	-10.57	1.57		-0.27	0.99	0.62	0.79	0.72	-0.06	0.50
8	0.904	-11.30	1.59	0.09	-0.29	0.95	0.62	0.85	0.71		0.51
8	0.902	-11.17	1.57	0.11		0.96	0.55	0.85	0.70	-0.18	0.49
9	0.904	-11.26	1.58	0.09	-0.25	0.98	0.63	0.85	0.71	-0.06	0.50

The classifiers shown above can be generically described in the following equation.

$$X \text{ (possibility of having colorectal cancer)} = \text{Logit} (P) = \ln (P/(1-P)) = b_0 + b_1 \Delta Ct_1 + b_2 \Delta Ct_2 + \dots + b_n \Delta Ct_n$$

5 Noted above are selected classifiers for combinations of 1, 2, 3, 4, 5, 6, 7, 8 or all 9 of the selected genes. The area under the curve is noted (ROC Area). Also provided is the constant b_0 and the coefficient for each biomarker (e.g. b_n) required for the selected classifier to be applied to the delta Ct of the noted gene for a test sample. As would be understood, where no coefficient is noted, that biomarker is not required for the classifier. Additional sensitivity and
10 specificity determinations can be made for each classifier, and will vary depending upon the threshold set.

The results of all of the combinations of the 9 genes is graphically represented in **FIG. 5**.

EXAMPLE 9

15 All possible combinations of biomarkers identified in **Table 12** are tested by applying logistic regression to data corresponding to level of product of the biomarkers noted. QRT-PCR is conducted on each of the genes noted in **Table 12** for a population of individuals who have

been diagnosed as having colorectal cancer or not having colorectal cancer. A matrix containing the ΔCt for RNA corresponding to each gene for each individual of the two populations is created and classifiers derived for each possible combination of biomarkers of those listed in **Table 12** (whether individually statistically significant or not) using techniques as described
 5 herein. For each classifier the ROC curve is plotted and the area under the curve determined. Classifiers are chosen depending on the specific sensitivity and specificity requirements of the specific intended use of the biomarkers (for example, it may be desirable to have a high sensitivity and fewer false negatives so as to miss less colorectal cancers); (alternatively high specificity resulting in lower false positives is also desirable as it can decrease costs of additional
 10 unnecessary medical interventions). A blind test is conducted on one or more of the resulting classifiers so as to demonstrate the utility of the classifier to test for colorectal cancer in a test individual. One or more the classifiers is applied to test an individual to determine the likelihood of said test subject having colorectal cancer.

EXAMPLE 10

15 **Blind Testing of one of the Combinations of the Genes in Table 12 using the Derived Classifier.**

A five gene combination encompassing B-cell scaffold protein with ankyrin repeats (BANK1), B-cell novel protein 1 (BCNP1), cytidine deaminase (CDA) membrane-spanning 4-domains, subfamily A, member 1 (MS4A1) and FERM domain containing 3 (MGC20553 aka
 20 FRMD3) was selected to pursue further blind sample testing.

The reference (training) data set was constructed containing ΔCt values for the above five genes assayed against a total of 57 subjects having colorectal cancer and 58 subjects not having colorectal cancer as described in more detail in Example 7.

Logistic regression was used to analyze the dependence of the binary variable Y
 25 (0=control, 1=disease) on the ΔCt values from the reference data set. If P = probability that a patient sample is identified as "having colorectal cancer), then a function $X = \text{Logit}(p)$ can be defined as follows:

$$X = \text{Logit}(P) = \ln(P/(1-P)) = b_0 + b_1 \Delta Ct_1 + b_2 \Delta Ct_2 + \dots + b_n \Delta Ct_n \quad (\text{Eq 1})$$

If $X \geq \text{threshold}$ then $Y=1$ (diagnosis = "has colorectal cancer"), and if $X < \text{threshold}$ then
 30 $Y=0$ (diagnosis = does not have colorectal cancer). The (empirical) coefficients $\{b_i\}$ that define the relationship between X and the experimental measurements $\{\Delta Ct_i\}$, where i represents a

REPLACEMENT SHEET

sample} were obtained by a maximum-likelihood (ML) fitting method. Identical {bi} values were obtained using several different commercial software programs: MedCalc (MedCalc Software, Mariakerke, Belgium) and XL-Stat (Addinsoft Software, Paris, France). ROC curve analysis was then used to evaluate the discriminatory power of the combinations. The classifier
5 derived using the selected genes(BANK1, BCNP1, CDA, MS4A1, and MGC20553) and shown as follows:

$X = \text{Logit}(P) = \ln(P/(1-P)) = -5.1338 - 0.8399(\Delta Ct \text{ CDA}) - 0.3314(\Delta Ct \text{ MGC20553}) - 0.3245(\Delta Ct \text{ MS4A1}) + 1.0903(\Delta Ct \text{ BCNP1}) + 0.7842(\Delta Ct \text{ BANK1})$ gave an area under the curve of 0.883 ± 0.032 . One can adjust the sensitivity or specificity by adjusting the testing cut
10 off point as follows:

Cut-off	Sens. (95% C.I.)	Spec. (95% C.I.)
-0.53	89.7 (78.8- 96.1)	78.9 (66.1- 88.6)
-0.31	86.2 (74.6- 93.8)	80.7 (68.1- 89.9)
-0.04	81.0 (68.6- 90.1)	82.5 (70.1- 91.2)
0.47	72.4 (59.1- 83.3)	87.7 (76.3- 94.9)
0.59	67.2 (53.7- 79.0)	89.5 (78.5- 96.0)
1.46	43.1 (30.2- 56.8)	94.7 (85.4- 98.8)

Using a cutoff of -1.1 gives a sensitivity of 98.3% and a specificity 50.9%.

A first blind test was performed using a scoring population comprised of 15 individuals not having colorectal cancer and 6 individuals having colorectal cancer. This blind test utilized
15 individuals selected from a single site in Penang. The first blind test resulted in a sensitivity of 100% and a specificity of 43%. A second blind test was performed using a second scoring population (non overlapping with the first scoring population) of 31 non colorectal patients and 23 colorectal patients. Patient samples were collected from three different sites in Asia. The test resulted in a sensitivity of 100% (all samples with colorectal cancer properly identified) and a
20 specificity of 47% (almost half of the samples without colorectal cancer properly identified). A final blind test was performed utilizing samples obtained from two clinics in North America including 15 colorectal cancer patients and 16 non colorectal cancer patients resulting in a sensitivity of 88% and a specificity of 33%.

EXAMPLE 11

25 **Selection of 6 Biomarkers (BCNP1, CD163, CDA, MS4A1, BANK1, MGC20553) to derive classifiers which are particularly useful in screening for the presence of colorectal cancer**

and application of the classifiers with the combinations selected to determine presence of colorectal cancer

QRT-PCR was performed on the selection of genes BCNP1, CD163, CDA, MS4A1, BANK1, MCG20553 identified from one or more microarray analyses (selected from Table 12 and Table 11). QRT-PCR data was most recently collected on RNA samples from centrifuged lysed blood from 109 samples, 60 individuals having colorectal cancer and 59 individuals not having colorectal cancer. QRT-PCR was performed using a duplexed TaqMan Assay using the QuantiTect® Probe PCR Kit (Qiagen, Cat. # 204345). TaqMan® probes were developed for each gene of interest and labelled with FAM and the Black Hole Quencher® from Biosearch Technologies. Beta-actin was used as a housekeeping gene in the duplexed assays and labelled using HEX and Black Hole Quencher®. Δ Cts (Ct gene of interest – Ct Beta-actin) were calculated.

Results of the average QRT-PCR results for each gene across the population tested are shown in Table C below. The average Δ Ct (Ct gene-Ct Beta-actin) of each gene in both the control population and the population having colorectal cancer (CRC) are shown, as is the standard deviation (SD) as amongst the data from the population. The p value for each gene individually is shown. The change in Δ Ct as between the Δ Ct of the Control samples for each gene and the Δ Ct for the CRC samples for each gene are shown as $\Delta\Delta$ Ct and the standard deviation are shown. Also shown is the fold change for each individual marker.

Table C

Control		BANK1	BCNP1	CDA	MGC205		
					53	MS4A1	CD163
	Avg Δ Ct of Control (non CRC)						
		7.5998	6.5558	3.4814	9.7375	3.9940	5.2012
	SD	0.767	0.907	0.491	1.036	0.974	0.381
Colorectal Cancer (CRC)					MGC205		
		BANK1	BCNP1	CDA	53	MS4A1	CD163
	CRC Avg	8.3398	7.6368	3.2070	9.3773	5.1298	5.3896
	Standard Deviation	0.708	0.893	0.539	0.867	1.041	0.482
	p-value	2.6E-07	1.6E-09	4.4E-03	4.2E-02	1.1E-08	2.0E-02

(CRC v.
Control)

	BANK1	BCNP1	CDA	MGC205 53	MS4A1	CD163
$\Delta\Delta\text{Ct}(\Delta\text{CtCo}$ $\text{nt}-\Delta\text{CtCRC})$	-0.7400	-1.0810	0.2743	0.3602	-1.1359	-0.1884
Standard Deviation	1.044	1.273	0.729	1.351	1.425	0.614
	BANK1	BCNP1	CDA	MGC205 53	MS4A1	CD163
Regulation (direction of regulation comparing CRC v. Control)	Down regulat ed	Down Regulat ed	Up regulat ed	Up regulated	Down regulat ed	Down regulat ed
Fold change (CRC v. Control)	1.67	2.12	1.21	1.28	2.20	1.14
Standard Deviation (Fold change)	2.06	2.42	1.66	2.55	2.69	1.53

A matrix of the individual QRT-PCR data corresponding to each of the six biomarkers across a population comprising 60 individuals having colorectal cancer and 59 individuals not having colorectal cancer was generated using the same methods noted above. The matrix of data was used to generate classifiers for all possible combinations of the six biomarkers by applying

5 logistic regression to the data of each possible combination. A listing of all of the combinations and the biomarkers which make up each combination are noted in **Table D** below. The ROC curve for each resulting classifier was measured to determine the ability of each derived classifier to properly identify a test patient as having colorectal cancer. The given sensitivity (at a pre-defined specificity) and the specificity (at a pre-defined sensitivity) can be determined from

10 the ROC curve. The resulting classifiers from this analysis is shown in **Table E** and the results of these classifiers are shown graphically in **FIG. 5**.

Table D Each combination is noted as a unique combination number ranging from 1 to 63. Presence of the number 1 in a column below a noted Gene indicates the presence of that gene within the generated classifier.

Table D

Combination #	BANK1 Gene01	BCNP1 Gene02	CDA Gene03	MS4A1 Gene04	MGC20553 Gene05	CD163 Gene06
1	1					
2		1				
3	1	1				
4			1			
5	1		1			
6		1	1			
7	1	1	1			
8				1		
9	1			1		
10		1		1		
11	1	1		1		
12			1	1		
13	1		1	1		
14		1	1	1		
15	1	1	1	1		
16					1	
17	1				1	
18		1			1	
19	1	1			1	
20			1		1	
21	1		1		1	
22		1	1		1	
23	1	1	1		1	
24				1	1	
25	1			1	1	
26		1		1	1	
27	1	1		1	1	
28			1	1	1	
29	1		1	1	1	
30		1	1	1	1	
31	1	1	1	1	1	
32						1
33	1					1
34		1				1
35	1	1				1
36			1			1
37	1		1			1
38		1	1			1
39	1	1	1			1
40				1		1
41	1			1		1

42		1		1		1
43	1	1		1		1
44			1	1		1
45	1		1	1		1
46		1	1	1		1
47	1	1	1	1		1
48					1	1
49	1				1	1
50		1			1	1
51	1	1			1	1
52			1		1	1
53	1		1		1	1
54		1	1		1	1
55	1	1	1		1	1
56				1	1	1
57	1			1	1	1
58		1		1	1	1
59	1	1		1	1	1
60			1	1	1	1
61	1		1	1	1	1
62		1	1	1	1	1
63	1	1	1	1	1	1

In preferred embodiments, data representing levels of products of any combination of two of the six biomarkers in a sample isolated or derived from a test subject are input to a formula as further described herein, for the purpose of providing a probability that the test subject has one or more colorectal pathologies. In certain other embodiments, data representing levels of products of any combination of three, four, or five of the six biomarkers are input to a formula as further described herein. It is also consistent with the methods described herein that data representing levels of products of all six of the six biomarkers are input to a formula to provide a probability that a test subject has one or more colorectal pathologies.

Table E Classifiers resulting from each possible combination of the six biomarkers are noted. The Combination number corresponds to the combinations as noted in **Table D**. The number of genes contributing to the combination is noted, and the coefficient for that biomarker are noted within the row. Where 0 is noted as a coefficient, this biomarker does not contribute to the resulting classifier. The area under the curve (ROC area) is noted as is the sensitivity at 90% specificity and the specificity at 90% sensitivity.

Table E

Comb #	# of genes	ROCarea	Sensitivity @ 90% spec	Specificity @ 90% sens	Constant	BANK1	BCNP1	CDA	MGC20553	MS4A1	CD163
2	1	0.8065	48.33	50.85	-9.4336	0	1.3378	0	0	0	0
16	1	0.8062	33.33	40.68	-5.1508	0	0	0	0	1.1456	0
1	1	0.7664	25.00	40.68	-10.646	1.3395	0	0	0	0	0
8	1	0.6531	11.67	18.64	3.9001	0	0	0	-0.40595	0	0
4	1	0.6421	26.67	15.25	3.5585	0	0	-1.0592	0	0	0
32	1	0.6232	18.33	18.64	-5.4467	0	0	0	0	0	1.0322
48	2	0.8285	36.67	61.02	-10.284	0	0	0	0	1.1534	0.96391
34	2	0.8184	48.33	54.24	-14.22	0	1.3294	0	0	0	0.91273
6	2	0.8155	33.33	52.54	-6.6582	0	1.2559	0.65891	0	0	0
18	2	0.8130	48.33	50.85	-8.6118	0	1	0	0	0.3475	0
20	2	0.8102	35.00	42.37	-2.2241	0	0	0.78667	0	1.0765	0
3	2	0.8093	48.33	47.46	-10.046	0.18289	1.2184	0	0	0	0
10	2	0.8068	51.67	52.54	-11.01	0	1.3885	0	0.12681	0	0
24	2	0.8065	31.67	42.37	-4.3207	0	0	0	-7.84E-02	1.1282	0
17	2	0.8048	28.33	42.37	-3.938	0.24835	0	0	0	1.3149	0
33	2	0.7921	36.67	49.15	-16.106	1.3541	0	0	0	0	1.0097
5	2	0.7754	30.00	54.24	-7.6711	1.2227	0	0.61344	0	0	0
9	2	0.7718	35.00	52.54	-8.8519	1.2937	0	0	-0.14952	0	0
36	2	0.7113	20.00	27.12	-2.8691	0	0	-1.391	0	0	1.4243
40	2	0.6983	15.00	25.42	-1.5982	0	0	0	-0.47483	0	1.1646
12	2	0.6961	18.33	18.64	7.3191	0	0	-1.0488	-0.39666	0	0
52	3	0.8319	26.67	61.02	-8.1557	0	0	-1.0734	0	1.0665	1.3105
49	3	0.8299	38.33	64.41	-9.4457	0.16237	0	0	0	1.2638	0.95548
38	3	0.8291	51.67	52.54	-11.787	0	1.2127	0.90702	0	0	1.1775
50	3	0.8288	43.33	55.93	-13.395	0	0.96662	0	0	0.38068	0.91813

WO 2007/048074

PCT/US2006/041600

56	3	0.8285	40.00	59.32	-8.8677	0	0	0	-0.16768	1.1205	1.0282
35	3	0.8226	50.00	52.54	-15.171	0.2494	1.1715	0	0	0	0.92838
22	3	0.8192	38.33	50.85	-5.7441	0	0.90176	0.66762	0	0.35836	0
21	3	0.8192	25.00	42.37	1.8168	0.71915	0	-0.9327	0	1.5568	0
42	3	0.8178	48.33	52.54	-14.643	0	1.3458	0	4.17E-02	0	0.89513
14	3	0.8158	38.33	52.54	-8.2509	0	1.3096	0.66151	0.1273	0	0
7	3	0.8155	33.33	52.54	-6.6677	2.33E-03	1.2544	0.65841	0	0	0
19	3	0.8147	45.00	52.54	-7.7027	0.18144	0.99459	0	0	0.47486	0
28	3	0.8124	38.33	42.37	-1.6155	0	0	0.78053	-5.87E-02	1.0615	0
26	3	0.8116	48.33	52.54	-9.8432	0	1.0675	0	9.36E-02	0.31563	0
11	3	0.8085	51.67	47.46	-11.398	0.15545	1.2828	0	0.11619	0	0
25	3	0.8068	28.33	47.46	-3.1168	0.24726	0	0	-7.82E-02	1.2971	0
37	3	0.8020	25.00	57.63	-13.206	1.1841	0	0.93208	0	0	1.3051
41	3	0.8017	38.33	55.93	-13.824	1.2908	0	0	-0.23193	0	1.0935
13	3	0.7788	20.00	52.54	-5.7872	1.1714	0	0.61629	-0.15341	0	0
44	3	0.7500	23.33	49.15	0.93831	0	0	-1.3775	-0.46832	0	1.5416
53	4	0.8421	28.33	62.71	-4.0963	0.72824	0	-1.2159	0	1.5557	1.3112
54	4	0.8359	43.33	54.24	-10.864	0	0.79246	0.93889	0	0.43542	1.215
60	4	0.8345	26.67	61.02	-6.7979	0	0	-1.0645	-0.16209	1.0289	1.3733
57	4	0.8322	41.67	64.41	-8.0548	0.15788	0	0	-0.16741	1.2282	1.0198
39	4	0.8294	51.67	52.54	-11.969	4.28E-02	1.1857	0.89948	0	0	1.179
58	4	0.8288	43.33	55.93	-13.371	0	0.96496	0	-2.24E-03	0.38154	0.91909
51	4	0.8285	41.67	55.93	-12.968	-7.93E-02	0.9637	0	0	0.43635	0.91323
46	4	0.8285	51.67	50.85	-12.048	0	1.2235	-	2.49E-02	0	1.1669

WO 2007/048074

PCT/US2006/041600

23	4	0.8237	41.67	44.07	-1.9249	0.65485	0.87804	0.81254	0	0.81182	0	0
43	4	0.8212	50.00	52.54	-15.382	0.2432	1.1845	0	2.32E-02	0	0.91822	0
30	4	0.8195	43.33	47.46	-6.9954	0	0.97268	-0.669	9.40E-02	0.32588	0	0
29	4	0.8192	28.33	47.46	2.3538	0.71624	0	0.92591	-5.37E-02	1.5416	0	0
15	4	0.8164	38.33	52.54	-8.1437	-3.34E-02	1.3322	0.66871	0.12965	0	0	0
27	4	0.8124	45.00	54.24	-8.9424	-0.1764	1.0611	0	9.25E-02	0.43969	0	0
45	4	0.8014	30.00	67.80	-10.66	1.1055	0	0.93995	-0.24668	0	1.3918	0
55	5	0.8410	30.00	57.63	-6.9559	0.67682	0.76786	-1.0879	0	0.9056	1.2208	0
61	5	0.8393	28.33	64.41	-2.722	0.73359	0	-1.2079	-0.16199	1.523	1.3754	0
62	5	0.8379	43.33	55.93	-10.555	0	0.76996	0.94059	-2.76E-02	0.44682	1.228	0
59	5	0.8285	41.67	55.93	-12.939	-7.94E-02	0.96171	0	-2.68E-03	0.43746	0.91437	0
47	5	0.8277	51.67	50.85	-12.173	3.61E-02	1.1995	0.90005	2.22E-02	0	1.1693	0
31	5	0.8212	41.67	40.68	-3.1633	0.65384	0.94851	0.81521	9.32E-02	0.77795	0	0
63	6	0.8407	31.67	61.02	-6.5754	0.67929	0.74043	-1.0901	-3.29E-02	0.92168	1.2365	0

Primers and probes which were utilized for the QRT-PCR) are further described in Table 17.

EXAMPLE 12

5 Testing a Test Subject for One or More Colorectal Pathologies Using The 6 Biomarkers (BCNP1, CD163, CDA, MS4A1, BANK1, MGC20553).

A reference population of 200 individuals is used to generate a formula which will be used to test the test subject. 100 of said individuals are confirmed to have colorectal cancer by colonoscopy. The remaining 100 individuals are screened for colorectal cancer by colonoscopy and are confirmed to be negative for colorectal cancer. Blood Samples are obtained for each of
10 the 200 individuals and total RNA isolated from each isolated blood sample. RNA from each sample is reverse transcribed using an oligo dT primer and cDNA corresponding to the total mRNA is obtained. QRT-PCR is performed on each of the six genes in each of the samples derived from the blood sample and a ΔC_t generated for each gene in reference to a Beta-actin control. A data matrix is generated of all of the data across the population. A classifier is
15 developed using each of the following methods (a) logistic regression, (b) linear regression (c) neural networks and (d) principle component analysis. A formula consisting of each of the classifiers, wherein each classifier itself is given a weighting of equal value is generated (ie the results of each classifier when applied to a test subject will give an indication of whether the test subject has a colorectal pathology, and then the results of each classifier are tallied such that if,
20 for example, 3 of the 4 classifiers indicate the test subject has colorectal pathology, the results of the formula indicate the test subject has colorectal pathology).

A blood sample is isolated from a test subject. Total RNA from the blood sample is isolated and cDNA derived using an oligo dT primer. QRT-PCR is performed in each of the six genes in the sample and a ΔC_t generated for each gene in reference to a Beta-actin control. The
25 data from the test subject's sample is input into the formula consisting of the four classifiers and a result of each classifier determined, along with the results of the sum of the four classifiers to provide an indication of whether said test subject has colorectal pathology, and in particular colorectal cancer.

30 FULL CITATIONS FOR REFERENCES REFERRED TO BY NUMBER IN THE SPECIFICATION

1] Ogawa M. Differentiation and proliferation of hematopoietic stem cells. Blood 1993;81:2844-53.

- 2] Liew, CC. Method for the detection of gene transcripts in blood and uses thereof. U.S. Patent Application No. 2002000268730.
- 5 3] Ma J, Liew CC. Gene profiling identifies secreted protein transcripts from peripheral blood cells in coronary artery disease. J Mol Cell Cardiol 2003;35:993-8.
- 10 4] Tsuang MT, Nossova N, Yager TD, Tsuang MM, Guo SC, Shyu KG, *et al.* Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: A preliminary report. Am J Med Genet B Neuropsychiatr Genet 2005;133B:1-5.
- 15 5] K. Wayne Marshal, Hongwei Zhang, Thomas D. Yager, Nadine Nossova, Adam Dempsey, Run Zheng, Mark Han, Hongchang Tang, Samuel Chao., C.C. Liew. Blood-Based Biomarkers for Detecting Mild Osteoarthritis in the Human Knee. Osteoarthritis and Cartilage 2005; 13: 861-871..
- 6] Whistler T, Unger ER, Nisenbaum R, Vernon SD. Integration of gene expression, clinical, and epidemiologic data to characterize Chronic Fatigue Syndrome. J Transl Med. 2003;1:10.
- 20 7] Bennett L, Palucka AK, Arce E., Cantrell V, Borvak J, Banchereau J, *et al.* Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J Exp Med 2003;197:711-23.
- 25 8] Zhang HQ, Lu H, Enosawa S, Takahara S, Sakamoto K, Nakajima T, *et al.* Microarray analysis of gene expression in peripheral blood mononuclear cells derived from long-surviving renal recipients. Transplantation Proceedings 2002;34:1757-9.
- 9] Tang Y, Nee AC, Lu A, Ran R, Sharp FR. Blood genomic expression profile for neuronal injury. J Cereb Blood Flow Metab. 2003;23:310-9.
- 30 10] Tang Y, Gilbert DL, Glauser TA, Hershey AD, Sharp FR. Blood Gene Expression Profiling of Neurologic Diseases: A Pilot Microarray Study. Arch Neurol. 2005;62:210-5.

- 11] Imperiale TF, Wagner DR, Lin CY, Larkin GN, Rogge JD, Ransohoff DF. Results of screening colonoscopy among persons 40 to 49 years of age. *N Engl J Med*. 2002 Jun 6;346(23):1781-5.
- 5 12] Ransohoff DF. Colorectal cancer screening in 2005: status and challenges. *Gastroenterology*. 2005 May;128(6):1685-95
- 13] Ahluwalia IB, Mack KA, Murphy W, Mokdad AH, Bales VS. State-specific prevalence of selected chronic disease-related characteristics--Behavioral Risk Factor Surveillance System, 2001. *MMWR Surveill Summ*. 2003 Aug 22;52(8):1-80.
- 10 14] Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, *et al*. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci U S A*. 2003;100:1896-901.
- 15 15] Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford, England: Oxford University Press; 2003.
- 16] Dupont WD. *Statistical Modeling for Biomedical Researchers*. Cambridge, England: Cambridge University Press; 2002.
- 20 17] Pampel FC. *Logistic regression: A Primer*. Publication # 07-132, Sage Publications: Thousand Oaks, CA. 2000.
- 25 18] King EN, Ryan TP. A preliminary investigation of maximum likelihood logistic regression versus exact logistic regression. *Am Statistician* 2002;56:163-170.
- 19] Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;8:283-98.
- 30 20] Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285-93.

- 21] Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993;39:561-77.
- 22] Witten IH, Frank Eibe. Data Mining: Practical Machine Learning Tools and Techniques
5 (second edition). Morgan Kaufman 2005.
- 23] Deutsch JM. Evolutionary algorithms for finding optimal gene sets in microarray prediction. Bioinformatics 2003;19:45-52.
- 10 24] Citarda F, Tomaselli G, Capocaccia R, Barcherini S, Crespi M; Italian Multicentre Study Group. Efficacy in standard clinical practice of colonoscopic polypectomy in reducing colorectal cancer incidence. Gut. 2001 Jun;48(6):812-5.
- 15 25] Niels Landwehr, Mark Hall and Eibe Frank (2003) Logistic Model Trees . pp 241-252 in Machine Learning: ECML 2003: 14th European Conference on Machine Learning, Cavtat-Dubrovnik, Croatia, September 22-26, 2003, Proceedings .Publisher: Springer-Verlag GmbH, ISSN: 0302-9743
- 20 26] Tonkin, E. T.; Wang, T.-J.; Lisgo, S.; Bamshad, M. J.; Strachan, T. NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nature Genet.* 36: 636-641, 2004. PubMed ID 15146185.
- 25 27] Krantz, I. D.; McCallum, J.; DeScipio, C.; Kaur, M.; Gillis, L. A.; Yaeger, D.; Jukofsky, L.; Wasserman, N.; Bottani, A.; Morris, C. A.; Nowaczyk, M. J. M.; Toriello, H.; and 9 others. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nature Genet.* 36: 631-635, 2004. PubMed ID 15146186.
- 30 28] Nadel MR, Shapiro JA, Klabunde CN, Seeff LC, Uhler R, Smith RA, Ransohoff DF. A national survey of primary care physicians' methods for screening for fecal occult blood. Ann Intern Med. 2005 Jan 18;142(2):86-94.

- 29] Scaloni A, Jones W, Pospischil M, Sassa S, Schneewind O, Popowicz AM, Bossa F, Graziano SL, Manning JM. Deficiency of acylpeptide hydrolase in small-cell lung carcinoma cell lines. *J Lab Clin Med*. 1992 Oct;120(4):546-52.
- 5 30] Erlandsson R, Boldog F, Persson B, Zabarovsky ER, Allikmets RL, Sumegi J, Klein G, Jornvall H. The gene from the short arm of chromosome 3, at D3F15S2, frequently deleted in renal cell carcinoma, encodes acylpeptide hydrolase. *Oncogene*. 1991 Jul;6(7):1293-5.
- 31] Brown MS, Goldstein JL. A proteolytic pathway that controls the cholesterol content of
10 membranes, cells, and blood. *Proc Natl Acad Sci U S A*. 1999 Sep 28;96(20):11041-8.
- 32] Adams JC, Seed B, Lawler J. Muskelin, a novel intracellular mediator of cell adhesive and cytoskeletal responses to thrombospondin-1. *EMBO J*. 1998 Sep 1;17(17):4964-74.
- 15 33] Gillis LA, McCallum J, Kaur M, DeScipio C, Yaeger D, Mariani A, Kline AD, Li HH, Devoto M, Jackson LG, Krantz ID. NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. *Am J Hum Genet*. 2004 Oct;75(4):610-23. Epub 2004 Aug 18.
- 20 34] Takakura S, Kohno T, Manda R, Okamoto A, Tanaka T, Yokota J. Genetic alterations and expression of the protein phosphatase 1 genes in human cancers. *Int J Oncol*. 2001 Apr;18(4):817-24.
- 35] Periale TF, Ransohoff DF, Itzkowitz SH, Turnbull BA, Ross ME; Colorectal Cancer Study
25 Group. Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N Engl J Med*. 2004 Dec 23;351(26):2704-14.

All patents, patent applications, and published references cited herein are hereby incorporated by reference in their entirety.

One skilled in the art will readily appreciate that the present invention is well adapted to
30 carry out the objects and obtain the ends and advantages mentioned, as well as those objects, ends and advantages inherent herein. The present examples, along with the methods, procedures,

treatments, molecules, and specific compounds described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention as defined by the scope of the claims.

5 While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

10

TABLES

Table 1 is found in the instant specification after Table 17, and before the sequence listing and claims.

TABLE 2

Gene Symbol	AffySpot	p-value(MW)	p-value (t-test)	Fold change	Direction	Gene Description
APEH	201284 s at		6.0e-03	0.91	downregulated	N-acylaminoacyl-pept
C1orf22	220342 x at	<0.001	3.15049390173162e-03	1.07	upregulated	chromosome 1 open re
ESR1	215551 at	0.058	0.03	0.87	downregulated	estrogen receptor 1
ETS1	214447 at	0.132	0.15	0.93	downregulated	v-ets erythroblastos
FLJ14624	225666 at		0.01	0.76	downregulated	hypothetical protein
FLJ20701			0.01	0.76	downregulated	hypothetical protein
FLJ23091	221958 s at		0.05	0.83	downregulated	putative NFkB activa
G2	234784 at	0.182	0.18	0.78	downregulated	G2 protein
ICOS	210439 at		0.07	0.88	downregulated	inducible T-cell co-
ITCH	235057 at	<0.001	1.73049856446377e-06	1.13	upregulated	itchy homolog E3 ubi
MBTPS1	217543 s at		8.0e-03	1.21	upregulated	membrane-bound trans
MGC40157	225065 x at	0.301	0.04	1.19	downregulated	hypothetical protein
MGC45871	226905 at		0.11	1.35	upregulated	hypothetical protein
MKLN1	242984 at	0.003	0.04	1.06	upregulated	muskelin 1, intracel
MMP9	203936 s at	0.006	7.80154116715436e-03	1.24	upregulated	matrix metalloprotei
NIPBL	1560474 at	0.005	0.01	1.07	upregulated	Nipped-B homolog (Dr
RPS24	1555878 at		0.05	0.88	downregulated	ribosomal protein S2
SMARCA1	203874 s at	0.077	0.06	0.82	downregulated	SWI/SNF related, mat

TABLE 3

Gene Symbol	RNA Accession No.	Protein Accession No.	Description
FLJ14624	NM_032813	NP_116202	Homo sapiens hypothetical protein FLJ14624 (FLJ14624), mRNA
FLJ14624	NM_032813	NP_116202	Homo sapiens hypothetical protein FLJ14624 (FLJ14624), mRNA
ETS1	NM_005238	NP_005229	Homo sapiens v-ets erythroblastosis virus E26 oncogene homolog 1 (avian) (ETS1), mRNA
RPS24	NM_033022	NP_148982	Homo sapiens ribosomal protein S24 (RPS24), transcript variant 1, mRNA
RPS24	NM_001026	NP_001017	Homo sapiens ribosomal protein S24 (RPS24), transcript variant 2, mRNA
FLJ20701	NM_017933	NP_060403	Homo sapiens hypothetical protein FLJ20701 (FLJ20701), mRNA
MKLN1	NM_013255	NP_037387	Homo sapiens muskelin 1, intracellular mediator containing kelch motifs (MKLN1), mRNA
NIPBL	NM_133433	NP_597677	Homo sapiens Nipped-B homolog (Drosophila) (NIPBL), transcript variant A, mRNA
NIPBL	NM_015384	NP_056199	Homo sapiens Nipped-B homolog (Drosophila) (NIPBL), transcript variant B, mRNA
FLJ20701	NM_017933	NP_060403	Homo sapiens hypothetical protein FLJ20701 (FLJ20701), mRNA
RPS24	NM_033022	NP_148982	Homo sapiens ribosomal protein S24 (RPS24), transcript variant 1, mRNA
RPS24	NM_001026	NP_001017	Homo sapiens ribosomal protein S24 (RPS24), transcript variant 2, mRNA
APEH	NM_001640	NP_001631	Homo sapiens N-acylaminoacyl-peptide hydrolase (APEH), mRNA
MBTPS1	NM_003791	NP_003782	Homo sapiens membrane-bound transcription factor protease, site 1 (MBTPS1), transcrip
MBTPS1	NM_201268	NP_957720	Homo sapiens membrane-bound transcription factor protease, site 1 (MBTPS1), transcrip
SMARCA1	NM_003069	NP_003060	Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin
SMARCA1	NM_139035	NP_620604	Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin
SMARCA1	NM_003069	NP_003060	Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin
SMARCA1	NM_139035	NP_620604	Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin
MMP9	NM_004994	NP_004985	Homo sapiens matrix metalloproteinase 9 (gelatinase B, 92kDa gelatinase, 92kDa type I
MKLN1	NM_013255	NP_037387	Homo sapiens muskelin 1, intracellular mediator containing kelch motifs (MKLN1), mRNA
ESR1	NM_000125	NP_000116	Homo sapiens estrogen receptor 1 (ESR1), mRNA
NIPBL	NM_133433	NP_597677	Homo sapiens Nipped-B homolog (Drosophila) (NIPBL), transcript variant A, mRNA
NIPBL	NM_015384	NP_056199	Homo sapiens Nipped-B homolog (Drosophila) (NIPBL), transcript variant B, mRNA
ITCH	NM_031483	NP_113671	Homo sapiens itchy homolog E3 ubiquitin protein ligase (mouse) (ITCH), mRNA
ITCH	NM_031483	NP_113671	Homo sapiens itchy homolog E3 ubiquitin protein ligase (mouse) (ITCH), mRNA
ICOS	NM_012092	NP_036224	Homo sapiens inducible T-cell co-stimulator (ICOS), mRNA
ESR1	NM_000125	NP_000116	Homo sapiens estrogen receptor 1 (ESR1), mRNA
ESR1	NM_000125	NP_000116	Homo sapiens estrogen receptor 1 (ESR1), mRNA
ESR1	NM_000125	NP_000116	Homo sapiens estrogen receptor 1 (ESR1), mRNA
ESR1	NM_000125	NP_000116	Homo sapiens estrogen receptor 1 (ESR1), mRNA
NIPBL	NM_133433	NP_597677	Homo sapiens Nipped-B homolog (Drosophila) (NIPBL), transcript variant A, mRNA

NIPBL	NM_015384	NP_056199	Homo sapiens Nipped-B homolog (Drosophila) (NIPBL), transcript variant B, mRNA
NIPBL	NM_133433	NP_597677	Homo sapiens Nipped-B homolog (Drosophila) (NIPBL), transcript variant A, mRNA
NIPBL	NM_015384	NP_056199	Homo sapiens Nipped-B homolog (Drosophila) (NIPBL), transcript variant B, mRNA
NIPBL	NM_133433	NP_597677	Homo sapiens Nipped-B homolog (Drosophila) (NIPBL), transcript variant A, mRNA
NIPBL	NM_015384	NP_056199	Homo sapiens Nipped-B homolog (Drosophila) (NIPBL), transcript variant B, mRNA
ETS1 mRNA	NM_005238	NP_005229	Homo sapiens v-ets erythroblastosis virus E26 oncogene homolog 1 (avian) (ETS1), mRNA
SMARCA1	NM_003069	NP_003060	Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin
SMARCA1	NM_139035	NP_620604	Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin
ESR1	NM_000125	NP_000116	Homo sapiens estrogen receptor 1 (ESR1), mRNA
ESR1	NM_000125	NP_000116	Homo sapiens estrogen receptor 1 (ESR1), mRNA
ITCH	NM_031483	NP_113671	Homo sapiens itchy homolog E3 ubiquitin protein ligase (mouse) (ITCH), mRNA
ESR1	NM_000125	NP_000116	Homo sapiens estrogen receptor 1 (ESR1), mRNA
ESR1	NM_000125	NP_000116	Homo sapiens estrogen receptor 1 (ESR1), mRNA
MBTPS1	NM_003791	NP_003782	Homo sapiens membrane-bound transcription factor protease, site 1 (MBTPS1), transcript
MBTPS1	NM_201268	NP_957720	Homo sapiens membrane-bound transcription factor protease, site 1 (MBTPS1), transcript
FLJ20701	NM_017933	NP_060403	Homo sapiens hypothetical protein FLJ20701 (FLJ20701), mRNA
C1orf22	NM_025191	NP_079467	Homo sapiens chromosome 1 open reading frame 22 (C1orf22), mRNA
C1orf22	NM_025191	NP_079467	Homo sapiens chromosome 1 open reading frame 22 (C1orf22), mRNA
FLJ23091	NM_024911	NP_079187	Homo sapiens putative NFkB activating protein 373 (FLJ23091), transcript variant 1, m
FLJ23091	NM_001002292	NP_001002292	Homo sapiens putative NFkB activating protein 373 (FLJ23091), transcript variant 2, m
C1orf22	NM_025191	NP_079467	Homo sapiens chromosome 1 open reading frame 22 (C1orf22), mRNA
ETS1 mRNA	NM_005238	NP_005229	Homo sapiens v-ets erythroblastosis virus E26 oncogene homolog 1 (avian) (ETS1), mRNA
MGC40157	NM_152350	NP_689563	Homo sapiens hypothetical protein MGC40157 (MGC40157), mRNA
MKLN1 mRNA	NM_013255	NP_037387	Homo sapiens muskellin 1, intracellular mediator containing kelch motifs (MKLN1), mRNA
FLJ14624	NM_032813	NP_116202	Homo sapiens hypothetical protein FLJ14624 (FLJ14624), mRNA
MGC45871	NM_182705	NP_874364	Homo sapiens hypothetical protein MGC45871 (MGC45871), mRNA
MGC45871	NM_182705	NP_874364	Homo sapiens hypothetical protein MGC45871 (MGC45871), mRNA
FLJ23091	NM_024911	NP_079187	Homo sapiens putative NFkB activating protein 373 (FLJ23091), transcript variant 1, m
FLJ23091	NM_001002292	NP_001002292	Homo sapiens putative NFkB activating protein 373 (FLJ23091), transcript variant 2, m
FLJ23091	NM_024911	NP_079187	Homo sapiens putative NFkB activating protein 373 (FLJ23091), transcript variant 1, m
FLJ23091	NM_001002292	NP_001002292	Homo sapiens putative NFkB activating protein 373 (FLJ23091), transcript variant 2, m
MKLN1 mRNA	NM_013255	NP_037387	Homo sapiens muskellin 1, intracellular mediator containing kelch motifs (MKLN1), mRNA
ESR1	NM_000125	NP_000116	Homo sapiens estrogen receptor 1 (ESR1), mRNA
ITCH	NM_031483	NP_113671	Homo sapiens itchy homolog E3 ubiquitin protein ligase (mouse) (ITCH), mRNA

MKLN1 mRNA	NM_013255	NP_037387	Homo sapiens muskelin 1, intracellular mediator containing kelch motifs (MKLN1),
FLJ20701	NM_017933	NP_060403	Homo sapiens hypothetical protein FLJ20701 (FLJ20701), mRNA
FLJ20701	NM_017933	NP_060403	Homo sapiens hypothetical protein FLJ20701 (FLJ20701), mRNA
ITCH	NM_031483	NP_113671	Homo sapiens itchy homolog E3 ubiquitin protein ligase (mouse) (ITCH), mRNA
NIPBL	NM_133433	NP_597677	Homo sapiens Nipped-B homolog (Drosophila) (NIPBL), transcript variant A, mRNA
NIPBL	NM_015384	NP_056199	Homo sapiens Nipped-B homolog (Drosophila) (NIPBL), transcript variant B, mRNA
MKLN1 mRNA	NM_013255	NP_037387	Homo sapiens muskelin 1, intracellular mediator containing kelch motifs (MKLN1),
MKLN1 mRNA	NM_013255	NP_037387	Homo sapiens muskelin 1, intracellular mediator containing kelch motifs (MKLN1),

TABLE 4

Gene Symbol	SensePrimer	SEQ ID NO	AntisensePrimer	SEQ ID NO	TaqManProbe	SEQ ID NO
APEH	GCCCTGTATTATGTGGACCT	54	AGATGGGTACTGCAGGTAGA	55	CCGGCTGAGCCAGACCAAT	56
APEH	CACTCGGAGACACACTTGT	57	CTTGGTCTGGCTTCTTCAG	58	CATCGCTGGCACTGACGTCCA	59
APEH	ACTCGGAGACACACTTGTG	60	CTTGGTCTGGCTTCTTCAG	61	CATCGCTGGCACTGACGTCCA	62
APEH	AGTGGTGGTAGATGTTGTGC	63	AGACCACTCTCTGGCTGTC	64	TGCAGCCTTCTGCCTTTGGGA	65
APEH	CACCTTGTGTATGTGGCAGA	66	GCCTGGCTATCTCATCATC	67		
C1orf22	TAGGGAGGAGAAACAGAAAGC	68	GGCCTCTAACTCGACCTCTA	69	TGGAACATGCTTACCCTGCTGATGA	70
C1orf22	TGTGGTGGATAAGAGCTGTC	71	CCCATCTTCTTCAGGATTC	72	TGGCCATGAATACTCTGGCTCTCA	73
C1orf22	GAGGAGAGTTTTCAGGAGTGG	74	CTCCCATCTTTGAGGTGAAT	75	TGGCCATGAATACTCTGGCTCTCA	76
C1orf22	TTGCTTGGAGATGACAGTTT	77	AGCAATTGGTTTGTGGATATG	78		
ESR1	GGCACATCTTCTGTCTTCTG	79	CTGTGAAGAGCTACGGGAAT	80	TGGAATCCCTTTTGGCTGTTOCC	81
ESR1	TGGCACATCTTCTGTCTTCT	82	CTGTGAAGAGCTACGGGAAT	83	TGGAATCCCTTTTGGCTGTTOCC	84
ESR1	CCCTACTACCTGGAGAACGA	85	ATTGGTACTGGCCAAATCTTT	86	CTGCCAACCTTGGCGTGGATT	87
ESR1	CACCATTCCTCAAGTTAATCC	88	GAAATGCAGTTTGGAAACAGA	89	TGGGACCAAAAGTTTCATTTGCTCCA	90
ESR1	TGCCCTACTACCTGGAGAAC	91	ATTGGTACTGGCCAAATCTTT	92	AGCCAGCGGCTACACGGTG	93
ESR1	GTGCCCTACTACCTGGAGAA	94	ATTGGTACTGGCCAAATCTTT	95	AGCCAGCGGCTACACGGTG	96
ESR1	TGTGCAATGACTATGCTTCA	97	TTATCAATGGTGCACTGGTT	98		
ESR1	TGTGCAATGACTATGCTTCA	99	TTTATCAATGGTGCACTGGT	100		
ETS1	TGGGTGGTTTATACACTGGA	101	ATAAGGGTTTTCACCCAGCTA	102	CCAGATTTGCCCATCCTTCCCTCTG	103
ETS1	GAGCTCCTCTCCCTCAAGTA	104	GGTGACGACTTCTTGTGTTGA	105	CCTCGGTCACTTCTCCGAGACCC	106
ETS1	GGATGTCAGGTGAGACTGTG	107	GCCTTCAAGTCAATTCCTCTC	108	CCCTGGCATCACCTGTGCCA	109
ETS1	TCAAACAAGAAAGTCGTCACC	110	GCGATCACAACATACTGTAGC	111	CCCGAGTTTACCACGACTGGTCCTC	112
ETS1	CCGCTATACCTCGGATTACT	113	GCGTCTGATAGGACTCTGTG	114	CCCAGTGTGTTCACCATCGGA	115
FLJ14624	ACAGCTGCCATCAGAAACTA	116	GCTTCCTGTAGCTCATTCCT	117	CCGGGAAGCTGTAAAGATTAAATCCCAA	118
FLJ14624	TGATAAAGGCAACCAGACAG	119	AGCTTCCTGTAGCTCATTC	120	TGCCATCAGAAACTACCGGGAAGC	121
FLJ14624	GGATTTCCTCGTTATCCCAT	122	TCTTGGTATGTTTGTCTCAGG	123	AGGACACGCTCCGCGACCAC	124
FLJ14624	CAAATACAGCCAGACTTTGC	125	TTTAAATTGCTGTCCGGTAA	126	TGCTGCTTCAAAACCGTTTCAGGC	127
FLJ14624	CAAATACAGCCAGACTTTGC	128	GTTTAAATTGCTGTCCGGTAA	129	TGCTGCTTCAAAACCGTTTCAGGC	130
FLJ14624	CTTGGTGATAGGCAAAATTCA	131	AGCAGGGTCAATTCCTGAAGAG	132	CCCGTTCTGAGCATGCCGA	133

PCT/US2006/041600

WO 2007/048074

PCT/US2006/041600

FLJ14624	TTTGGCTGTGCTTTATCATC	134	CAGCAGACCGTAATTCTCCT	135	TGTTTCTTGCCAAAGTCTAGATGTCCC	136
FLJ20701	ATTGGAGGACAAGAGCAGAT	137	CCATCGCTCTCTAGATTGG	138	TTCTTAGGTGCCGCAAGTGCCC	139
FLJ20701	CCTTCAGGAAGACTTTCAC	140	CTCAAAGTTTCATTCAGCCATC	141	ACGGGCGGATCCACAGCAAC	142
FLJ20701	TGATCAGCAAGTGAAACAC	143	CTCGGTGATAGCAAAATCAGA	144	TGCATCCTTGATGGCAAAGCTTCA	145
FLJ20701	GTGATCAGCAAGTGAAACACA	146	CTCGGTGATAGCAAAATCAGA	147	TGCATCCTTGATGGCAAAGCTTCA	148
FLJ20701	TTAATGGTCTGTCTGATGC	149	GGGTCTCTAGACAAAGCCAAG	150		
FLJ20701	AGATTGTGACAGACACAGAAGC	151	TTCTAACATCAGGTGGTTGC	152		
FLJ23091	TACAACTCACGAATCCCTTCT	153	ACAGGAAGTAGAGGCAGAGG	154		
FLJ23091	CAACTCACGAATCCCTTCTAC	155	ACAGGAAGTAGAGGCAGAGG	156		
FLJ23091	ACAACTCACGAATCCCTTCTA	157	ACAGGAAGTAGAGGCAGAGG	158		
FLJ23091	AACTCACGAATCCCTTCTACA	159	ACAGGAAGTAGAGGCAGAGG	160		
FLJ23091	GGGATTTCCATGACCTTTAT	161	ATCCAGAAAGGACAGAAAGCAT	162	TGCCCTGTGGATGTCAACA	163
FLJ23091	ATTGAAGAGGCAATTCCAAG	164	TTTAGCTTGAAGGCAATGTC	165	CCACATGGAGATGAGTCCTTGGTTCC	166
ICOS	CATGTGTAATGCTGGATGTG	167	AAACAACTCAGGGAACACCT	168	TGGACAACCTGACTGGCTTTGCA	169
ICOS	CAGGCCTCTGGTATTCTTT	170	ATTTGTACACCTCCGTTGTG	171	TTGGCAGAAACCATTTGATTTCTCCTGTT	172
ICOS	AAACATGAAGTCAGGCCCTCT	173	GTACACCTCCGTTGTGAAAT	174	TTGGCAGAAACCATTTGATTTCTCCTGTT	175
ICOS	GTCAGGCCTCTGGTATTCT	176	ATTTGTACACCTCCGTTGTG	177	TTGGCAGAAACCATTTGATTTCTCCTGTT	178
ICOS	GAAGTACAGGCCTCTGGTATT	179	ATTTGTACACCTCCGTTGTG	180	TTGGCAGAAACCATTTGATTTCTCCTGTT	181
ITCH	CAGATCCAAGGATGAAACAA	182	ACCACCATTTGAGAGTGATG	183	CACCAGCTCCTGCATCTTCAGGG	184
ITCH	TCAATCCAGATCACCTGAAA	185	AATCCTTGAGTCCCACTGGT	186	CCCATGGAACAGAGCCATGGC	187
ITCH	AAGCTGTGTTTGCCATAGA	188	CAGAGAAAGGACAAACATTGC	189	TGCCCATTCATGGTGCAAGTTCTC	190
ITCH	AAGCTGTGTTTGCCATAGA	191	GCAGAGAAAGGACAAACATTG	192	TGCCCATTCATGGTGCAAGTTCTC	193
ITCH	GCTGTGTTTGCCATAGAAG	194	CAGAGAAAGGACAAACATTGC	195	TGCCCATTCATGGTGCAAGTTCTC	196
ITCH	GCTGTGTTTGCCATAGAAG	197	GCAGAGAAAGGACAAACATTG	198	TGCCCATTCATGGTGCAAGTTCTC	199
MBTPS1	CAATGACGGACCTCTTTATG	200	GGTAGCTCCAGGTAGTCAT	201	TGCCGCCTACTCCAATCACATCC	202
MBTPS1	TCTGTGGGAAGAAACATCTG	203	TGATGAGAAATTCACCTTCA	204		
MBTPS1	AATCCATCCAGTGACTACCC	205	ACTTGAGGGAAACGAAAGACT	206	AACATCAAAACGGGTACGCCC	207
MBTPS1	CAGCCAAAGCTAGAAATTCA	208	TAGGGTAGTCACCTGGATGGA	209	TTCACTGCTCTTTCAGGGCAGCTTGAA	210
MBTPS1	GCAATGACGGACCTCTTTAT	211	GGTAGCTCCAGGTAGTCAT	212	TGCCGCCTACTCCAATCACATCC	213
MGC40157	AGAATCAGCATCATGTTTGG	214	ATAACCTTCTCTTGGGCTGA	215	CCTCATGGCAGGCTCCTGGC	216
MGC40157	TCAGCCCAAGAGAAAGGTTAT	217	TGAGCATGTCTCTGTATACA	218	TCCCAAGGACCAAGTAGCTGCCA	219
MGC40157	GACTACAGCTCACAGCACAC	220	AAAGCTACAACTTGGCCTGT	221	TGCCCAGGCTGGTCTCAGGC	222
MGC40157	TCAGCCCAAGAGAAAGGTTAT	223	AGGCAAGCATGTTTCTACAC	224	TCCCAAGGACCAAGTAGCTGCCA	225
MGC40157	ACTACAGCTCACAGCACACC	226	AAAGCTACAACTTGGCCTGT	227	TGCCCAGGCTGGTCTCAGGC	228

PCT/US2006/041600

WO 2007/048074

PCT/US2006/041600

MGC40157	TCAGCCCAAGAGAAAGGTTAT	229	CATAAGGCAAGCATGTTCT	230	TCCAAGGACCAGTAGTGCCA	231
MGC40157	AGGCTCATGGATCACTCTTT	232	GGTACGCAATCCAGTTCTCT	233	CCGGCCTTCGCAGACTCCAG	234
MCG45871	GACCTCTCTGATGAATGCTG	235	AATGACGTGAAGGTAAGGT	236	ACCGGCTCTCCCGCTGTCT	237
MCG45872	GACCTCTCTGATGAATGCTG	238	GAATGACGTGAAGGTAAGG	239	ACCGGCTCTCCCGCTGTCT	240
MCG45873	GACCTCTCTGATGAATGCTG	241	GAATGACGTGAAGGTAAGGT	242	ACCGGCTCTCCCGCTGTCT	243
MCG45874	GACCTCTCTGATGAATGCTG	244	GGAATGACGTGAAGGTAAG	245	ACCGGCTCTCCCGCTGTCT	246
MCG45875	CCTCTCTGATGAATGCTGAC	247	GAATGACGTGAAGGTAAGG	248	ACCGGCTCTCCCGCTGTCT	249
MCG45876	ACCTCTCTGATGAATGCTGA	250	GAATGACGTGAAGGTAAGG	251	ACCGGCTCTCCCGCTGTCT	252
MKLN1	CCAGTGAACCAACAATTCAGT	253	ATGCAGTGTCTTATTCGAGA	254	TGGATGTCTCAGGCCCAGCA	255
MKLN1	GGATCACACCTATGCTCAAA	256	CCATTCTGTGTCCAGTGAC	257	TGACAGCATGACTCTCTCTAAAGGCA	258
MKLN1	ATTTGAGAGAGGAGGCTGAG	259	AATGAAATGCCCTGTCAGTTG	260	CCACTGGACAAACCAAAACCATTTCTC	261
MKLN1	GACTTGTAATGGCAGCGTAG	262	CTCGAAGAAGTTTCCAGGTT	263	TGACAGCAGAGCCAGTGAACCACA	264
MKLN1	ACTTGTAATGGCAGCGTAGA	265	CTCGAAGAAGTTTCCAGGTT	266	TGACAGCAGAGCCAGTGAACCACA	267
MMP9	ACTTTGACAGCGACAAAGAAAT	268	GCGGTACATAGGGTACATGA	269	CGCCGCCACGAGGAACAAAC	270
MMP9	AACTTTGACAGCGACAAAGAA	271	GAAGCGGTACATAGGGTACAT	272	CGCCGCCACGAGGAACAAAC	273
MMP9	CAGTACCACGGCCAACTAC	274	TGGAAGATGAATGGAAACTG	275	CCCATCAGCATTGCCGTCCC	276
MMP9	CCACTACTGTGCCCTTTGAGT	277	GTACTTCCCATCCTTGAACA	278	TTCCCAAATCTCCGCGATGGC	279
MMP9	CCACTACTGTGCCCTTTGAGT	280	CTTCCCATCCTTGAACAA	281	TTCCCAAATCTCCGCGATGGC	282
NIPBL	CACGAATAGCAGAAAGAGGTG	283	GTATGTCACCTTCTGGGTCA	284	CAGCTTGTCCATAGCCTCAACCAGG	285
NIPBL	CCCATCCTTCAAGTTACACA	286	ATTAGCTGAATTGCCAGACA	287	TCCACAGATGCAACAAGCATCGG	288
NIPBL	GGGATTGCTAGTCTCACAGA	289	TCTTCTGCTATTCTGTGCATT	290	TCCTGAACCAAGCTGCCTCTTCCA	291
NIPBL	CCCATCCTTCAAGTTACACA	292	AGACAACGGACCAGAAACTT	293	TCCACAGATGCAACAAGCATCGG	294
NIPBL	GCACAGGCTAAGTAGTGACG	295	AGGTGGTCTTTGAGCCTTTAG	296	TTTCCCTTGAGATCTCCACAGCCA	297
NIPBL	CACGAATAGCAGAAAGAGGTG	298	GTATGTCACCTTCTGGGTCA	299	CAGCTTGTCCATAGCCTCAACCAGG	300
NIPBL	CCCATCCTTCAAGTTACACA	301	ATTAGCTGAATTGCCAGACA	302	TCCACAGATGCAACAAGCATCGG	303
NIPBL	GGGATTGCTAGTCTCACAGA	304	TCTTCTGCTATTCTGTGCATT	305	TCCTGAACCAAGCTGCCTCTTCCA	306
NIPBL	CCCATCCTTCAAGTTACACA	307	AGACAACGGACCAGAAACTT	308	TCCACAGATGCAACAAGCATCGG	309
NIPBL	GCACAGGCTAAGTAGTGACG	310	AGGTGGTCTTTGAGCCTTTAG	311	TTTCCCTTGAGATCTCCACAGCCA	312
RPS24	CACCGTAACTATCCGCACTA	313	CGGTGTGGTCTTTGTACATTT	314	AGGCACTGTGCGCTTCCCGG	315
RPS24	AAGAAAAGCAACGAAAGGAAC	316	TCATTGTCAGCACCTTTTACTC	317	TGCCAGCACCAACAATTGGCC	318
RPS24	AGACATGGCCTGTATGAGAA	319	ATCCAATCTCCAGCTCACTT	320	TGCCAGCACCAACAATTGGCC	321
RPS24	AAGACATGGCCTGTATGAGA	322	ATCCAATCTCCAGCTCACTT	323	TGCCAGCACCAACAATTGGCC	324
RPS24	GACATGGCCTGTATGAGAAAG	325	ATCCAATCTCCAGCTCACTT	326	TGCCAGCACCAACAATTGGCC	327
RPS24	AAGGAACGCAAGAAACAGAAT	328	ATTGCAGCACCTTTTACTCT	329	TGCCAGCACCAACAATTGGCC	330

SMARCA1	TACCTGGTCATTGATGAAGC	331	CAAAGGTGTTCCAGTTAGGA	332	CGTGAGTTCAAGTCGACTAACCGCTTG	333
SMARCA1	TACCTGGTCATTGATGAAGC	334	TTATTCTGCAAAAGGTGTTCC	335	TTCAAGTCGACTAACCGCTTGCTCC	336
SMARCA1	TGGACCCAGAAATATGAAGAG	337	ATGTTCAAGTGGAGATGTTGG	338	TCTCTTTGCTCGGTCGGCTTTCA	339
SMARCA1	CTTTGCTTGGTTACCTGAAA	340	AAACAAATGACACGGAGAGA	341	CCGAAATATTCTCTGGACCCTCACATGG	342
SMARCA1	GGAAATGGACCCAGAAATATG	343	GTTGGAGATTCTGTGCTGA	344	TCTCTTTGCTCGGTCGGCTTTCA	345
SMARCA1	AAGGAAATGGACCCAGAAATA	346	CTGAAAGGCTGAATGAAATGT	347	TCTCTTTGCTCGGTCGGCTTTCA	348
SMARCA1	AGTGGGATGTTTGCGTTACT	349	CACGAACAATCTCTGAAAGC	350	TCATCAATGACCAGGTATCGCCAGTG	351

TABLE 5

Gene Symbol	Description	Commercially Available Antibody Reference	Scientific Reference
APEH	N-acylaminoacyl-peptide hydrolase		
C1orf22	chromosome 1 open reading frame 22		
ESR1	estrogen receptor 1	Ab2746 (Abcam)	Shevde NK & Pike JW Estrogen modulates the recruitment of myelopoietic cell progenitors in rat through a stromal cell-independent mechanism involving apoptosis. <i>Blood</i> 87:2683-92 (1996). Yang NN <i>et al.</i> Identification of an estrogen response element activated by metabolites of 17beta-estradiol and raloxifene. <i>Science</i> 273:1222-5 (1996).
ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)	Ab10936 (Abcam)	Pande P <i>et al.</i> Ets-1: a plausible marker of invasive potential and lymph node metastasis in human oral squamous cell carcinomas. <i>J Pathol</i> 189:40-5 (1999). Nakayama T <i>et al.</i> Expression of the Ets-1 proto-oncogene in human gastric carcinoma: correlation with tumor invasion. <i>Am J Pathol</i> 149:1931-9 (1996). Bories JC <i>et al.</i> Increased T-cell apoptosis and terminal B-cell differentiation induced by inactivation of the Ets-1 proto-oncogene. <i>Nature</i> 377:635-8 (1995). Wernert N <i>et al.</i> Stromal expression of c-Ets1 transcription factor correlates with tumor invasion. <i>Cancer Res</i> 54:5683-8 (1994).
FLJ14624			
FLJ20701	putative NFkB activating protein 373		
FLJ23091	inducible T-cell co-stimulator		
G2	G2 protein		
ICOS	inducible T-cell co-stimulator	Ab3744 (Abcam)	

ITCH	itchy homolog E3 ubiquitin protein ligase (mouse)		
MBTPS1	membrane-bound transcription factor protease, site 1		
MGC40157	hypothetical protein MGC40157		
MGC45871	hypothetical protein MGC45871		
MKLN1	muskelin 1, intracellular mediator containing kelch motifs		
MMP9	matrix metalloproteinase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	Ab5707	
NIPBL	Nipped-B homolog (Drosophila)		
RPS24	ribosomal protein S24		
SMARCA1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1	Ab21924 (Abcam)	Lazzaro MA & Picketts DJ Cloning and characterization of the murine Imitation Switch (ISWI) genes: differential expression patterns suggest distinct developmental roles for Snf2h and Snf2l. <i>J Neurochem</i> 77:1145-56 (2001).

TABLE 6

Symbol	Ref. ID	5' Primer	Position	3' Primer	Position	Product Length
		Primer Sequence		Primer Sequence		
APEH	NM_001640	AAGGATGTCCAGTTTG CAGTGG	2081	TGGCAGGAAATGAAG CCACCAT	2184	104
FLJ23091	NM_024911	ACAGGCATCTATGGGA TGTGGA	1694	AGATCGCCATTGGAC TGGTCTT	1794	101
MBTPS1	NM_003791	TCGGTACTCCAAGGTT CTGGA	3220	TGTTTCCAAAGGTTAC TGGGCG	3348	129
MGC40157	NM_152350	ACCCGAGAGAACTGGA TTGCGT	359	GCTCCAATACTCAGCT GCCAAA	469	111
MGC45871	NM_182705	CCACCAAAGGAAGTAA GGTACAC	362	TAGTTGCGCCACGTGC CATT	507	146
MKLN1	NM_013255	GAGGGCCGAAATTGGT GTTTGA	1169	ATTGTGGTTCACTGGC TCTGCT	1297	129
NIPBL	NM_015384	AGTGTACGCCACTTTG CCCTAA	6886	ATCAGCCTTGTTCCGC ATAGCA	7017	132
PPP1R2	NM_006241	AAGATGCCTGTAGTGA CACCGA	250	ATCCGATACTTTGGCT CCAAGC	349	100

TABLE 7

	AJ36h		Blind test	
	No Polyp	Polyp	No Polyp	Polyp
Sample Size	110	68	40	40
Gender (F/M)	55/54*	22/45*	21/19	13/27
Age mean (range)	57 (23~83)	57 (38~82)	57 (40~79)	60 (38~76)
Polyp Subtype				
Tubular adenoma		21 (31%)		14 (35%)
Hyperplastic		18 (27%)		15 (38%)
High risk pathology		7 (10%)		3 (7.5%)
Others		22 (32%)		8 (20%)

* one sample missing information

PCT/US2006/041600

TABLE 8

Equ #	# ratios	ROC area	Constant	MGC45871-APEH	MKLN1-APEH	MGC45871-FLJ23091	MKLN1-FLJ23091	NIPBL-FLJ23091	MGC45871-MGC40157	MKLN1-MGC40157	NIPBL-MGC40157	MGC45871-PPP1R2	MKLN1-PPP1R2	Thresh
26692	5	0.718	-2.483	0	0.173	0	-0.291	0	0	0	-0.173	-0.863	-0.929	-0.481
26690	5	0.718	-2.483	0.173	0	0	-0.291	0	0	0	-0.173	-1.036	-0.756	-0.481
26658	5	0.718	-2.483	0.173	0	-0.291	0	0	0	0	-0.173	-0.745	-1.047	-0.481
26660	5	0.718	-2.483	0	0.173	-0.291	0	0	0	0	-0.173	-0.572	-1.220	-0.481
25732	5	0.718	-3.284	0	0.441	0	0	-0.230	0	-0.514	0	-0.793	-0.916	-0.014
25218	5	0.718	-3.285	0.441	0	0	0	-0.230	-0.514	0	0	-0.720	-0.989	-0.014
25730	5	0.718	-3.284	0.441	0	0	0	-0.230	0	-0.514	0	-1.234	-0.475	-0.014
25220	5	0.718	-3.285	0	0.441	0	0	-0.230	-0.514	0	0	-0.279	-1.430	-0.014
17030	5	0.718	-3.285	-0.279	0.720	0	0	-0.230	-0.514	0	0	0	-1.709	-0.014
18052	5	0.718	-3.284	0	0.441	0	0	-0.230	-0.793	0.279	0	0	-1.709	-0.014

TABLE 9

Parameter	Blind
Number of Samples	80
Number of Equations	10
Sensitivity (TPF)	43 % (17/40)
Specificity (TNF)	80 % (32/40)
Overall accuracy	61 % (49/80)

TABLE 10

Reporter Gene	Protein Activity & Measurement
CAT (chloramphenicol acetyltransferase)	Transfers radioactive acetyl groups to chloramphenicol or detection by thin layer chromatography and autoradiography
GAL (beta-galactosidase)	Hydrolyzes colorless galactosides to yield colored products.
GUS (beta-glucuronidase)	Hydrolyzes colorless glucuronides to yield colored products.
LUC (luciferase)	Oxidizes luciferin, emitting photons
GFP (green fluorescent protein)	Fluorescent protein without substrate
SEAP (secreted alkaline phosphatase)	Luminescence reaction with suitable substrates or with substrates that generate chromophores
HRP (horseradish peroxidase)	In the presence of hydrogen oxide, oxidation of 3,3',5,5'-tetramethylbenzidine to form a colored complex
AP (alkaline phosphatase)	Luminescence reaction with suitable substrates or with substrates that generate chromophores

PCT/US2006/041600

Table 11 is found in the instant specification after Tables 17 and 1, and before the sequence listing and claims.

5

TABLE 12

AffySpotID	Gene	GENE ID	p value	CCa/Ctrl	direction	DEFAULTGENEDESCRIPTION
219073 s at	OSBPL10	114884	3.8499E-08	0.451209844	downregulated	oxysterol binding protein-like 10
1563498 s at	LOC283130	283130	8.1224E-06	0.628849872	downregulated	hypothetical protein LOC283130
219667 s at	BANK1	55024	1.0668E-10	0.430259207	downregulated	B-cell scaffold protein with ankyrin repeats 1
205547 s at	TAGLN	6876	0.40991641	0.889953014	downregulated	transgelin
203642 s at	COBLL1	22837	5.3941E-07	0.474077388	downregulated	COBL-like 1
228551 at	MGC24039	160518	0.05283113	0.695785002	downregulated	hypothetical protein MGC24039
235866 at	C9orf85	138241	4.2656E-06	0.649259289	downregulated	chromosome 9 open reading frame 85
207655 s at	BLNK	29760	4.7146E-09	0.508654531	downregulated	B-cell linker
230983 at	BCNP1	199786	1.8615E-11	0.422871727	downregulated	B-cell novel protein 1
208591 s at	PDE3B	5140	0.05166031	0.826542405	downregulated	phosphodiesterase 3B, cGMP-inhibited
201340 s at	ENC1	8507	0.62995039	0.93326798	downregulated	ectodermal-neural cortex (with BTB-like domain)
208325 s at	AKAP13	11214	0.00033172	0.80202517	downregulated	A kinase (PRKA) anchor protein 13
233559 s at	WDFY1	57590	0.00920709	0.845905759	downregulated	WD repeat and FYVE domain containing 1
205627 at	CDA	978	9.2103E-05	1.343547864	upregulated	cytidine deaminase
1555736 a at	AGTRAP	57085	0.09637521	0.906895763	downregulated	angiotensin II receptor-associated protein
1554390 s at	ACTR2	10097	0.03306275	0.892740407	downregulated	ARP2 actin-related protein 2 homolog (yeast)
201554 x at	GYG	2992	0.51537492	1.055438457	upregulated	glycogenin 1
1559051 s at	C6orf150	115004	0.97824175	1.002133453	upregulated	chromosome 6 open reading frame 150
207795 s at	KLRD1	3824	0.19904885	0.861882517	downregulated	killer cell lectin-like receptor subfamily D, member 1

220784_s at	UTS2	10911	0.0424344 6	0.627549503	downregulated	urotensin 2
217418_x at	MS4A1	931	1.3139E- 09	0.423617526	downregulated	membrane-spanning 4-domains, subfamily A, member 1
1563674 at	SPAP1	79368	9.3602E- 09	0.400736546	downregulated	Fc receptor-like 2
1570259 at	LIMS1	3987	0.9227008 8	0.990561902	downregulated	LIM and senescent cell antigen- like domains 1
228988 at	ZNF6	7552	0.2656937 8	0.691253833	downregulated	zinc finger protein 6 (CMPX1)
215314 at	ANK3	288	0.0345854 7	0.655571614	downregulated	ankyrin 3, node of Ranvier (ankyrin G)
232911 at	KIAA155 9	57677	2.7111E- 07	0.640258188	downregulated	mouse zinc finger protein 14-like
200661 at	PPGB	5476	0.6034369	1.039669304	upregulated	protective protein for beta- galactosidase (galactosialidosis)
238581 at	GBP5	115362	0.0326383 5	1.323760479	upregulated	guanylate binding protein 5
212859_x at	MT1E	4493	0.4624285 9	0.892029229	downregulated	metallothionein 1E (functional)
229893 at	MGC205 53 (FRMD3)	257019	0.0232284 4	1.294641476	upregulated	FERM domain containing 3
204415 at	G1P3	2537	0.9856335 6	1.003662309	upregulated	interferon, alpha-inducible protein (clone IFI-6-16)
211889_x at	CEACA M1	634	0.1218161 9	1.155086057	upregulated	carcinoembryonic antigen- related cell adhesion molecule 1 (biliary glycoprotein)
1558733 at	FLJ35036	253461	0.9989125	1.000149237	3/4	zinc finger and BTB domain containing 38
235885 at	P2RY12	64805	0.6452796 3	1.061974664	upregulated	purinergic receptor P2Y, G- protein coupled, 12
208180_s at	HIST1H4 H	8365	1.275E-05	0.673893846	downregulated	histone 1, H4h
204627_s at	ITGB3	3690	0.2785639 1	1.238694854	upregulated	integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)
1554676 at	PRG1	5552	0.1482295 4	1.10791216	upregulated	proteoglycan 1, secretory granule
208685_x at	BRD2	6046	0.0141117 7	0.840822417	downregulated	bromodomain containing 2
219922_s at	LTBP3	4054	0.0002801 1	0.553407885	downregulated	latent transforming growth factor beta binding protein 3

218311 at	MAP4K3	8491	5.0346E-06	0.680583549	downregulated	mitogen-activated protein kinase kinase kinase 3
212129 at	NIPA2	81614	0.00333649	0.8287373	downregulated	non imprinted in Prader-Willi/Angelman syndrome 2
229039 at	SYN2	6854	0.3194711	1.139160086	upregulated	synapsin II
203645_s at	CD163	9332	0.76	0.98	downregulated	CD163 molecule
221557_s at	LEF1	51176	0.28	0.87	downregulated	lymphoid enhancer-binding factor 1
1552772 at	CLEC4D	362432	0.01	1.31	upregulated	C-type lectin domain family 4, member d
220001 at	PADI4	23569	0.06	1.18	upregulated	peptidyl arginine deiminase, type IV
206026_s a	TNFAIP6	460097	0.03	1.29	upregulated	tumor necrosis factor, alpha-induced protein 6
202431_s at	MYC	4609	0.06	0.84	downregulated	v-myc myelocytomatosis viral oncogene homolog (avian)

TABLE 13

AffySpotID	GeneSymbolCG	RNA Sequence	RNA Acc'n	Protein Acc'n
1554390_s_at	ACTR2	SEQ ID NO 1	NM_001005386	NP_001005386
1554390_s_at	ACTR2	SEQ ID NO 2	NM_005722	NP_005713
1555736_a_at	AGTRAP	SEQ ID NO 3	NM_020350	NP_065083
1559051_s_at	C6orf150	SEQ ID NO 4	NM_138441	NP_612450
1563674_at	SPAP1	SEQ ID NO 5	NM_030764	NP_110391
200661_at	PPGB	SEQ ID NO 6	NM_000308	NP_000299
201340_s_at	ENC1	SEQ ID NO 7	NM_003633	NP_003624
201554_x_at	GYG	SEQ ID NO 8	NM_004130	NP_004121
203642_s_at	COBLL1	SEQ ID NO 9	NM_014900	NP_055715
203645_s_at	CD163	SEQ ID NO 10	NM_004244	NP_004235
203645_s_at	CD163	SEQ ID NO 11	NM_203416	NP_981961
204415_at	G1P3	SEQ ID NO 12	NM_002038	NP_002029
204415_at	G1P3	SEQ ID NO 13	NM_022872	NP_075010
204415_at	G1P3	SEQ ID NO 14	NM_022873	NP_075011
204627_s_at	ITGB3	SEQ ID NO 15	NM_000212	NP_000203
205547_s_at	TAGLN	SEQ ID NO 16	NM_001001522	NP_001001522
205547_s_at	TAGLN	SEQ ID NO 17	NM_003186	NP_003177
205627_at	CDA	SEQ ID NO 18	NM_001785	NP_001776
207655_s_at	BLNK	SEQ ID NO 19	NM_013314	NP_037446
207795_s_at	KLRD1	SEQ ID NO 20	NM_002262	NP_002253
207795_s_at	KLRD1	SEQ ID NO 21	NM_007334	NP_031360

208180_s_at	HIST1H4H	SEQ ID NO 22	NM_003543	NP_003534
208325_s_at	AKAP13	SEQ ID NO 23	NM_006738	NP_006729
208325_s_at	AKAP13	SEQ ID NO 24	NM_007200	NP_009131
208325_s_at	AKAP13	SEQ ID NO 25	NM_144767	NP_658913
208591_s_at	PDE3B	SEQ ID NO 26	NM_000922	NP_000913
208685_x_at	BRD2	SEQ ID NO 27	NM_005104	NP_005095
211889_x_at	CEACAM1	SEQ ID NO 28	NM_001712	NP_001703
212129_at	NIPA2	SEQ ID NO 29	NM_001008860	NP_001008860
212129_at	NIPA2	SEQ ID NO 30	NM_001008892	NP_001008892
212129_at	NIPA2	SEQ ID NO 31	NM_001008894	NP_001008894
212129_at	NIPA2	SEQ ID NO 32	NM_030922	NP_112184
217418_x_at	MS4A1	SEQ ID NO 33	NM_021950	NP_068769
217418_x_at	MS4A1	SEQ ID NO 34	NM_152866	NP_690605
218311_at	MAP4K3	SEQ ID NO 35	NM_003618	NP_003609
219073_s_at	OSBPL10	SEQ ID NO 36	NM_017784	NP_060254
219667_s_at	BANK1	SEQ ID NO 37	NM_017935	NP_060405
219922_s_at	LTBP3	SEQ ID NO 38	NM_021070	NP_066548
220784_s_at	UTS2	SEQ ID NO 39	NM_006786	NP_006777
220784_s_at	UTS2	SEQ ID NO 40	NM_021995	NP_068835
228988_at	ZNF6	SEQ ID NO 41	NM_021998	NP_068838
229039_at	SYN2	SEQ ID NO 42	NM_003178	NP_003169
230983_at	BCNP1	SEQ ID NO 43	NM_173544	NP_775815
232911_at	KIAA1559	SEQ ID NO 44	NM_020917	NP_065968
233559_s_at	WDFY1	SEQ ID	NM_020830	NP_065881

WO 2007/048074

PCT/US2006/041600

		NO 45		
235885_at	P2RY12	SEQ ID NO 46	NM_022788	NP_073625
235885_at	P2RY12	SEQ ID NO 47	NM_176876	NP_795345
229893_at	MGC20553 (FRMD3)	SEQ ID NO 48	NM_174938	NP_777598
221557_s_at	LEF1	SEQ ID NO 49	NM_016269	NP_057353
1552772_at	CLEC4D	SEQ ID NO 50	NM_080387	NP_525126
220001_at	PADI4	SEQ ID NO 51	NM_012387	NP_036519
206026_s_a	TNFAIP6	SEQ ID NO 52	NM_007115	NP_009046
202431_s_at	MYC	SEQ ID NO 53	NM_002467	NP_002458

TABLE 14

Gene	GeneID	RNA Acc'n	Sense Primer (5' Primer)	Antisense Primer (3' Primer)	TaqMan Probe
OSBPL10	114884	NM_017784	GCTGGTGGTGTACTCTGCTA	CATTTCATGTGGTATTTGG	CAAGCTCGAAGCTGAGTCACCCA
OSBPL10	114884	NM_017784	GTGCTCAACTTGTACAGC	CTGCTCTGTGGAATGTGACT	CAGCCAGCCAGAAAGCCAGG
OSBPL10	114884	NM_017784	TCTATGGAGGGAAGTCCAC	TGACTTTGGTTTCTCCATTG	CCGCAGAAAGTGAAGCACAAACCC
LOC283130	283130	NM_182556	TGGTCTTGATCTCCTGACTT	AGGGACTCTCAGAAAGTGGAG	
LOC283130	283130	NM_182556	TTCTTCAAGGGAATGAGCTT	GGAAGATGTGCATGTAGCTG	CTGCTGGTGTCTACGGCCAC
LOC283130	283130	NM_182556	TCAACTCTGTCTCCTCTGCT	TTACAGTGTCAAAACGGGTGT	CCAGTTCCCGCAAGCCACAGA
BANK1	55024	AK091523	ACAGGCATCACTACTTCCAA	ACCATCTTCACCATATGCAA	TGAGGATCCCACATCTGAGATCAA
BANK1	55024	AK091523	TGAAAGCAGGAAGACATACG	ACCTCTAGTGGGCTGTGTTT	CATTTGCCTCAGCTCCATCTGCA
BANK1	55024	AK091523	TATCCAGCTTCACTTTCTGC	TGTCAAGTACAGAGCCCCATT	CCCATGGTCAACTGCCATCTGAA
TAGLN	6876	NM_003186	AACCCAGACACAAGTCTTCA	TCCAGCTCCTCGTCATACTT	CTGCACCTTCGCGGCTCATGC
TAGLN	6876	NM_003186	AGCAATGGTAACTGCACCT	CCAGGGAGGAGACAGTAGAG	
TAGLN	6876	NM_003186	TGACATGTTCCAGACTGTTG	GCGCTTCTTTCATAAACCA	CTGCCAAGCTGCCCAAAGCC
COBLL1	22837	NM_014900	ATGGCTAGATGTCCCAAGTT	AATGACACAGCCAGTAGTGC	
COBLL1	22837	NM_014900	GCTTGGTGTGTCTGATAAGG	ACTCTGTGCGCATCTGTCTTG	ACCAACTCCAAGTGTATGGCCCA
COBLL1	22837	NM_014900	CAGAACAGATGCGACAGAGT	AATGGCTGAGTCTTGACCCTT	
MGC24039	160518	AK125323	TTGACAAAGCTTCCCTTCTG	TTTCTCTTCCCACTGAGACA	CCAGCCTGAGCCTTACCTGCCA
MGC24039	160518	AK125323	ACCTTTGAGTGCCAGAACTT	TCATTTCTGACCATGACACA	
MGC24039	160518	AK125323	GGACACAAACGGACAAATAA	GCTGTGCAGATTTGCTTACT	CCAGGGAAGATGGTTCTCGCACC
C9orf85	138241	NM_182505	GCACCAGAATACGTTTAGCTT	TTTACACGCCCACTCAAGAAC	CAGCGCTGACATACCTCCATCATGAAG
C9orf85	138241	NM_182505	CTGGGAGTAGTTCGTTGGTT	CGTATCTCTGGTCTTCTGAG	TCTGGAACGAGCCACCGTTGCC
C9orf85	138241	NM_182505	GGAGATTGAAGTGAGCTGAG	AAGGTGCTCAGAAATTTACACA	CGTGCCATTACACTCCAGCCTGG
BLNK	29760	NM_013314	CTTGAGACCAGAGGCTTACC	AAGCTTGCTCATTCTGTTTG	CACGTACAGCAGTTCTCTGGCCC
BLNK	29760	NM_013314	GCTATTGAAGTGTCATCCA	CATGATAACTCAACCTCACCA	CAGGTCAGGCAATAGAACAAAGTCCACA
BLNK	29760	NM_013314	CCGTGGAAGATAATGATGAA	GAATTTGGCTTGGTTGATCT	TCCCACAGAAAGCAGTTTACCTCCA
BCNP1	199786	NM_173544	ACTGGATATTGGCAGCTTCT	TCTCCAGGGAGTGTGAAAG	CCCACTCCGGAAGCAGCTGC
BCNP1	199786	NM_173544	GCAGCTCCAAATCTTAACCTTG	AGAAGCTGCCAATATCCAGT	CCGACTTCTCTGCTTGCCAGC
BCNP1	199786	NM_173544	TTTCTCCCATCTCTGCTG	AGCAAATGTATTTTCGGAAGG	TGCCCTTTCCTCCTATTTCCCTCCA
PDE3B	5140	NM_000922	AATGGCTATCGAGACATTCC	TTCAATTTCTGTTCACACAAC	CAACACGGCCAGTTCTTGGC
PDE3B	5140	NM_000922	CTTCCACCACAAAGTCATTTT	CTCAGCTGGGTCTCTCTATTT	TGCTTTCTCAGGTTCTCTGTAGGCCA
PDE3B	5140	NM_000922	AGGTGGGATCGTAATAATGG	TGCTTCGGGATAGTCAGTAG	AAAGGCCTCACCAAGAATTTGGCA
ENC1	8507	NM_003633	TGCCGTCGTAGGTATTAGTG	GTGAGAAACATGGACGAAAG	TCGAGTTGCAAACTTTGGTCTTCCC
ENC1	8507	NM_003633	TGTGCCGTCGTAGGTATTAG	GTGAGAAACATGGACGAAAG	TCGAGTTGCAAACTTTGGTCTTCCC
ENC1	8507	NM_003633	CAAAACCATCAGGAAGATGA	TCCAGTTAATTGCAGACTCG	CAAGCCTTTTCATCCTCTGTCTCCAGC
AKAP13	11214	NM_006738	CTCCTGGTTCATGATTGTTGT	ATCCTGGACGAAATGAAAGT	CCACAAACACGGGAAGGCCCC

WO 2007/048074

PCT/US2006/041600

AKAP13	11214	NM_006738	AGGTTCTGGTGGACAAAGTT	GGGAAGAGAGATGACAAAGG	CAGGAACGGTCTGTATTCTCCTCCTCA
AKAP13	11214	NM_006738	GTTCACAGACTGAAAGTCC	TCCAGAGTCCACAATAGGTG	TGCCTCCCGAGCATGTTCCC
WDFY1	57590	NM_020830	GACAAAGTGTGTAGCTGGAT	TGATCTGCCCCAGAAATATCA	TCGCCCTGCCCTCTGGTGGAC
WDFY1	57590	NM_020830	GCTGAAGCTTGAACAGAACAA	TCAGATGCTCCTGAGAAAGAG	TTCCGCCGCTCCGAGGAACAG
WDFY1	57590	NM_020830	CTGTGCTACCTTCAGCTCAC	ACTTCCAAACCACTGAGGAG	ACCCGCTGGCATCTGTGCT
CDA	978	NM_001785	CAAAGGGTGCAACATAGAAA	TTGCCCTGAAATCCTTGTA	TGCTATCGCCAGTGACATGCAAGA
CDA	978	NM_001785	TACAAGGATTTGAGGGCAAT	AGTTGGTGCCAAACTCTCTC	TCCCTTCCCTTCCCTGTGGGCC
CDA	978	NM_001785	TGCCCTGGGACTTAGAACAC	CAGGATAGAAACCTTGGGAAG	CCAGCTCAGGGATTGCCTGA
AGTRAP	57085	NM_020350	GTTCTAGGGATGCTCCTGAC	GGAGTGTGATGTGTAAGT	CCGCCCTCCCGGTACATGTG
AGTRAP	57085	NM_020350	CTCCTGCTGCTTCTGCTAC	CCTCTGCTGAGTCAATCGT	CTACCCGCGGTGAGCCTCA
AGTRAP	57085	NM_020350	CACCATCTTCCCTGGACATC	CTCCCGGTACATGTGCTAG	ATCCTGGCCTGCCATCACGG
ACTR2	10097	NM_005722	GTGCTTCTGGAGGGTCTAC	TCTTCAATGCGGATCTTAAA	TCCCTCCCAGCAATATCCAGTCTCC
ACTR2	10097	NM_005722	AGACTCTGGAGATGGTGTGA	CTCGCAACAGAAAGTAGCTTG	CCCAACTTTGATGATACGTCCATCTGG
ACTR2	10097	NM_005722	AGAGCAGAAACTGGCCTTAG	GGCTGAAATAAAGCTTCTGG	TCACTCTGGCAGCACTGGGCA
GYG1	2992	NM_004130	TGTGTACCTTTCACGAGACC	TGGTTAATGCTATTGTTG	TGAGGCCCATGATCCCAACATG
GYG1	2992	NM_004130	CCAAAGTTGTGCAATTTCT	GGTAAAGATGTTCCACACACA	TCCAGCTATGGCACAGCCG
GYG1	2992	NM_004130	TGTGGCTTCTGTAGAAAGGA	TCTGCTCCCATATAATCAGC	AGCACCCAAAGAGGCCTGCG
C6orf150	115004	NM_138441	TATAACCCTGGCTTTGGAAT	ATGGCTTAGTCGTAGTTGCT	AGCACCCAAAGAGGCCTGCG
C6orf150	115004	NM_138441	GATATAACCCTGGCTTTGGA	ATGCTTGGGTACAAAGTAAA	CCAGGCTCTTTGGGTCCCA
C6orf150	115004	NM_138441	ACTGCCCTTCTTTCACGTATG	CCTGAGGCACTGAAGAAAG	CGTGCCCTTCTCTACTTCCGCTCTTGA
KLRD1	3824	AF498040	ACAATTCAACGCTGTTCTTT	TTATCCCTAAGGTCCCAGAA	TGTGCTTCTCAGAAATCCAGCCTGC
KLRD1	3824	NM_007334	CGGTGCAACTGTACTTTCAT	AATTGTTGACTGGAGCTCAT	CGCTCTTGGAAACATAAATTTCTCATGCG
KLRD1	3824	NM_007334	CACATCGTGCCTTCTCTACT	TCTATGTTGGTCCCTGGAGT	TCCACGTCTCTTTGCTTTGGCCA
UTS2	10911	NM_021995	TTATGCTCTGCGTCACTTCT	GAGGTGCTGAGAGTTGAAAG	TCCACGTCTCTTTGCTTTGGCCA
UTS2	10911	NM_021995	CACCTCTGCTCGGACTCATA	GTCTTCATGAGGTGCTGAGA	
UTS2	10911	NM_021995	CITTCAACTCTCAGCACCTC	TGAGTCTGCTTTCCCTGAGAA	
MS4A1	931	NM_021950	TCGTTGAGAAATGAATGAAA	ATGTTTCAGTTAGCCCAACC	CCGGATCACTCCTGGCAGCA
MS4A1	931	NM_021950	ACCCATCTGTGTGACTGTGT	CTTTGACCAACACTTCCCTG	TCACATTTCTGAAGCACTCATTTCTGCCT
MS4A1	931	NM_021950	GTGTTGTCACGCTTCTTCTT	AAAGCCTATCCAAGGAACAG	CCCTGGCTTCAAGTTGCATGAGC
SPATA1	64173	NM_022354	AAGTCACAGCATCAATGGAG	AGTCTGTGATGTTTGGCCAGA	TGGGATCTCTTCCAAAGTTCTCCTTGA
SPATA1	64173	NM_022354	CTGATGGAACAATCCACAGA	CCAGCTTCTTCCCTGATTCTT	CCTCCTCCATTGATGCTGTGACTTTC
SPATA1	64173	NM_022354	CGATATCATGCCCTACAAATGG	CTTCAAGTTGATGAGCAGT	TGCGCTCAGTGCTTCCAGCA
LIMS1	3987	NM_004987	AGCTGTACCATGAGCAGTGT	GGCAAAGAGCATCTGAAAG	TGGGAGACACCTGTGTGCGCC
LIMS1	3987	NM_004987	TATCGGGTTTGTCAAGAATG	TGAATATCAGAGGCTGCTCA	AGCAGTTACGCCACCAGGCC
LIMS1	3987	NM_004987	GTGATGTGGTCTCTGCTCTT	TACAGACTGGCTTTCATGTCA	TTCAACATCATCTTCAGCCTCCGC
ZNF6	7552	NM_021998	GATGGGTTTGGTCTGAAAGT	CTCCAGCTACTGAATGTCCA	TGCCGAGCCCAAGCAACAG
ZNF6	7552	NM_021998	TCAGCTTGAGGACTCTGATG	TTTGAAGGTACTGTGCTGCT	

ZNF6	7552	NM_021998	AAGACCCATACTGGAAGGAA	CCTTCTTGCAGAAATTCACAC	TCACATGTCGTTTAAAGCCAGATGCA
ANK3	288	NM_001149	TTGATCATCCAAGCCTAGTG	GAATCATCACCCCAATTCCTT	CGGTCAATTGCATCTTCACTCTGCG
ANK3	288	NM_001149	AATTCACCGAGAAGTGTGTG	TGAAAGTGATTTCCATGCTC	CGCTGCTTCCACACATTAATGGCA
ANK3	288	NM_001149	ATGACCTCCAGTCGTAAC	CTGGCTCTCATCTACATCCA	TGCCCTTAAACAGAAATGCCTGAAGCA
KIAA1559	57677	NM_020917.1	AAAATTAGCCAGGCATGGTG	GCGATCTCAACTACCCACAA	TCCGCCCTCCTGGATTCAAGTGA
KIAA1559	57677	NM_020917.1	ACCAAATTTACAGGTCACCA	CTCAAACTCCTGGCCTCAAG	CCCTGTGCCAGGCCAAATTCA
KIAA1559	57677	NM_020917.1	ACCAAATTTACAGGTCACCA	CTCAAGTGATCCACCCACCT	CCCTGTGCCAGGCCAAATTCA
PPGB	5476	NM_000308	GAGAGCTATGCTGGCATCTA	GACCAGGGAGTTGTCAATCT	TCTTGCATGACCAGCACCGGC
PPGB	5476	NM_000308	ACAGGCTTTGGTCTTCTCTC	TTGTAGATGTTGAGGCCAGA	CACGATCGGGCCACTTCCT
PPGB	5476	NM_000308	AACAAGCAGCCATACTGATG	GGCACTTTGCTTAGAAGAGG	CAGCTCCACGGCCTGATGCA
GBP5	115362	NM_052942	CCTACCTGATGAACAAGCTG	GATGAGGCACACACCATATC	CGGTGCAGTCTCACACCAAGGG
GBP5	115362	NM_052942	ACCAAGGGAATTTGGATATG	AGTAAGAGTGCCAGTGCAAA	CTCCCAGGCCCTCGGTGTC
GBP5	115362	NM_052942	GAAGAAAGAGGCACAAGTGA	TGTTTGGCTATCTCCATTG	TCTGCCCTTTGAATCGCCGCC
MT1E	4493	NM_175617	AAGTCTACTGCCACCTCTCAC	AAGAGCTGTTCCACATCA	CCCTGGGCACACTTGGCACA
MT1E	4493	NM_175617	CCTATGTTTCAGAACAGAGC	AGTGAGAGGTGGCAGTAGAC	CGGGCACCTCCCTGCCCTAA
MT1E	4493	NM_175617	ACTGCCACCTCTCACCTCTC	AGAGCTGTTCCACATCAG	CCCTGGGCACACTTGGCACA
FRMD3	257019	NM_174938	GAACCCACTGGTCAAGAGTT	ACTGCTCAAACCTCTGGTGC	CCAGGCCACCAACAAGGAGC
FRMD3	257019	NM_174938	TTGTGGTCTTTCAGGGAAT	AGTATTGAATGCCAACATGG	TCTCCTTCTGGGTGCCAATCACA
FRMD3	257019	NM_174938	AACATTTCTGCTCCCTTGAT	GGCACTTGGGTTGTACACTA	CTCCCGGGCTGCCCTTCACTG
G1P3	2537	NM_022873	GGTAATATTGGTGCCCTGA	TTAGGCCAAGAAGGAAGAAG	TGGGCTACGCCACCCACAAG
G1P3	2537	NM_022873	TATTGTCCAGGCTAGAGTGC	CTACTTGGGAGGTTGAGACA	TGCAGCCTCCAACTCCTAGCCTCA
G1P3	2537	NM_022873	TCTTCTCTCCTCCAAGTCT	TCTTACCTGCATCCTTACCC	TACCGCCTTCTGCCGCGATGG
CEACAM1	634	NM_001712	CAACCTACCTGTGGTGATA	TTCACACTCATAGGGTCTCTG	TCCAGGCTGCAGCTGTCCA
CEACAM1	634	NM_001712	CACCTCACAGAGTGC GTGA	GAATGGCATGGATTCAAGTAG	CTTCTGGAACCCGCCACCA
CEACAM1	634	NM_001712	AGGTTCTTCTCCTTGTCAC	GGTAGCTTGTGAGTTCCCT	CGGTTGCCATCCACTCTTTCCC
ZBTB38	253461	NM_172341.5	AGGGACCTCAAGGACGACTT	TTGAAGCTGCCAGAACATTG	CCCGAATGCGCTGCTCATTTAA
ZBTB38	253461	NM_172341.5	ACCAACGCTGAATTTCCAAG	TCCATTTTCAATGCCCTCCIC	CAACAGTCCAGCCATCCCATTGG
ZBTB38	253461	NM_172341.5	ATGGAGGAGGTTACACACCAG	GGATTATGGGGCTCATGCTA	TCATGCTGCACCCCTGACCCGG
P2RY12	64805	NM_176876	TGGTAACACCCAGTCTGTGC	GCCATTTGTGATAAGTCCAA	CACCCAGTCTCTTCCCCTGTC
P2RY12	64805	NM_022788	GCTGCAGAACAGAACACTTT	TGCCAAACCTCTTTGTGATA	TGCTCTTGTAACTCTGACCCCTGGACATG
P2RY12	64805	NM_022788	ACATTCAAACCCCTCCAGAAT	GGTGCACAGACTGGTGTAC	CGACAACCTCACCTCTGCGCC
HIST1H4H	8365	NM_003543	GTGTTCTGAAGGTGTTCTCTG	GTAAAGAGTGGTCCCTGTC	CGCCAAACGCAAGACCGTGA
HIST1H4H	8365	NM_003543	TAACATCCAGGGCATCACTA	GAACACCTTCAGAACACCAC	TATCCGGCGCCTTGCTCGTC
HIST1H4H	8365	NM_003543	ATAACATCCAGGGCATCACT	AGAACACCCAGAGTCTCCTC	TATCCGGCGCCTTGCTCGTC
ITGB3	3690	NM_000212	GACTCAAGATTGGAGACACG	AGGCACAGTCACAATCAAAG	TGTCCCTGAAGCCCAAGGGC
ITGB3	3690	NM_000212	TTACCACTGATGCCAAGACT	TGGAGGCAGAGTAATGATTG	CCTGCCAGCCTTCCGTCCAA
ITGB3	3690	NM_000212	AACCTTCAGATTTCCTTA	CAGTGAGGGTGTGGAATTAG	TGTCACCCCTTAGGCCAGCAACCA

PRG1	5552	NM_002727	CTTCCACITTTCTGAGGACT	ATCCTGTTCCATTTCGGTTA	CGGCTCCGGCTCTGGATCAG
PRG1	5552	NM_002727	GGTTCTGGAATCCTCAGTTC	TTGGATTACCTTGGAAAGTAG	TCTGGATTGCAGCGCACCCA
PRG1	5552	NM_002727	TCGAACTACTTCCAGGTGAA	GTCCTCAGAAAGTGGGAAGA	TCCAAAGACGAGAATCCAGGACTTGA
BRD2	6046	NM_005104	GATTCAGAGGAGGAGGAAGA	TGGGTTTGAATCACGTAAG	TCCAGGCTCAGCTGCCGCTT
BRD2	6046	NM_005104	ATGCCAGATGAACCACACTAGA	ATCTTCCTCCTCCTCTTCCT	CCCTGGCTTGGCCAAATCGTC
BRD2	6046	NM_005104	ACGGCTTATGTTCTCCAAC	CCCTGGTTCTAGTGGTTTCA	CGTGCCATTGCCACAACATCG
LTP3	4054	AF318354	ACCTTCAGGGCTCCTATGT	CACGTGTGTCATCGAAGTTCA	CCACCCAGGACCAGCACCGGT
LTP3	4054	AF318354	GGACAACAACATCGTCAACT	GCITGCAGTAGCACTCGTAG	CCAGCCCACCGTGACATCGA
LTP3	4054	AF318354	CTCTGGCTACCATCTGTCC	TGTAGGAGCCATTGGTATTG	TCCTCGCAGTGGCTCCGGTC
MAP4K3	8491	NM_003618	TGAAATTTGATCCACCCTTA	TGGTGTCTTGTCCATATTC	CCACATCATGAACCTCCCGACAGTG
MAP4K3	8491	NM_003618	TAAAGTGCAATATGGGTGCAT	TCGGCACCAATATCAAGTA	CACGTGTCATCATCATGGATAAACCCGA
MAP4K3	8491	NM_003618	CTGGAATATGGACAAGGACA	CTTCTAAATGTGCGACGTTG	CCCTCGCTGATGCAATTCCTTCAACA
NIPA2	81614	NM_030922	TATCAAGGAGCTGTTTGAG	TGGAGTCACAATGGAAGTGT	AAGCCTGTGCTCGGCATCC
NIPA2	81614	NM_030922	AACACTTCCATTGTGACTCC	GCACCTCAAAGTACCAATGA	CATCGTCAACAGGCATATCTTGCCA
NIPA2	81614	NM_030922	CTCACAAAGCTAGGTGATCCA	CCGATTACAGAGCAGATTGT	CCATGGCGAGGACCCACCAC
SYN2	6854	NM_003178	GAACTGGAAGACGAACACTG	TACAGCTTTGACAGCACAGA	CCAGGCCGCCAAACATCTCA
SYN2	6854	NM_003178	CATAATGGGAGACCAATCA	GGAAAGAGAAAGCTCCTCAA	CCACAGCCCTCCAGGGTCCA
SYN2	6854	NM_003178	TCTGTGCTGTCAAAGCTGTA	CTAGTTCGGTGATGAGTTGC	TGGCATGTCTACAGTCCATGACCTCA
LEF1	51176	NM_016269	GATGTCAACTCCAACAAGGC	AGTGAGGATGGTAGGGTTG	TTATCCCTTGTCTCCGGGTGGT
LEF1	51176	NM_016269	TGGATTCAAGGCAACCCCTAC	TGTGGGGATGTTCTGTTTG	ATGTCCAGGTTTTCCCATCA
LEF1	51176	NM_016269	AGGCTGGTCTGCAAGAGACA	GAATGAGCTTCGTTTTCCACC	TGCATCAGGTACAGGTCCAAGA
CLEC4D	362432	NM_080387	TTCACGCTGTAAAGAGAGGCAC	AGTTGGACTGGAAGGCTCTC	GAACTGAAAGTGTCTGAAGGGA
CLEC4D	362432	NM_080387	GAACTGAAAGTGTGAAGGG	TCCTTTCACTCTCAGCCAC	AGCACCTGGAACCTGTTGTCT
CLEC4D	362432	NM_080387	GGGCTGAGAGTGAAAGGAAC	CCACTGACCTTTGGCATTTC	ATGACCATCAGCACGGGAAGC
PADI4	23569	NM_012387	CTCCACCAGTCAAAGCTCTA	TCCAGGTCTCTGTGATCTTTC	CTACATCCAAGCCCCCACACAAAAC
PADI4	23569	NM_012387	ATGCAGGATGAAATGGAGAT	GTTTGATGGGAAACTCCTTC	CTACATCCAAGCCCCCACACAAA
PADI4	23569	NM_012387	GAAAGATCAGAGGACCTGGA	TCAAGCACTTCATCATCTCTC	TACATCCAAGCCCCCACACAAA
TNFAIP6	23569	NM_007115	CAGGTTGCTTGGCTGATTATG	TTGATTTGGAAACCTCCAGC	TGGCTTTGTGGGAAGATACTGTGG
TNFAIP6	23569	NM_007115	CATTAGACTCAAGTATGTCAGCG	TCCACAGTATCTTCCCACAAAG	CAGGTTGCTTGGCTGATTATGT
TNFAIP6	23569	NM_007115	TACCACAGAGAAGCACGGTC	ATCCATCCAGCAGCACAGA	GCAGAAGCTAAGGCGGTGTGTGAA
MYC	4609	NM_002467	GAGGCTATTCTGCCCAATTG	TCCTGTGTTGTTGAAGCTAACG	TTTCGGGTAGTGGAAAACCA
MYC	4609	NM_002467	TTTCGGGTAGTGGAAAACCA	CGTCGCAGTAGAAATACGGCT	CTATGACCTCGACTACGACTCGGT
MYC	4609	NM_002467	AGTGGAACACAGCAGCCTC	CGTCGCAGTAGAAATACGGCT	ATGACCTCGACTACGACTCGGT
CD163	9332	NM_203416	TGAGTCTTCTTGTGGGATTGT	CGACCTGTTGTGGCTTTT	TTCAGTGCAGAAAACCCCCACA
CD163	9332	NM_203416	TCTTCTTGTGGGATTGTCC	AGGATGACTGACGGGATGA	CAGTGCAGAAAACCCCCACAAA
CD163	9332	NM_203416	CCCACAAAAGCCACAACA	CTTGAGGAAACTGCAAGCC	TCATCCCGTCAGTCATCCTTTATTGC

TABLE 15

Gene Symbol	Protein Description	Commercially Available Antibody Reference	Scientific Reference
OSBPL10	oxysterol binding protein-like 10		
LOC283130	hypothetical protein LOC283130		
BANK1	B-cell scaffold protein with ankyrin repeats 1		
TAGLN	Transgelin (alternative name smooth muscle 22 Protein SM22)	Abcam Ab14106 AntiHuman Rabbit Polyclonal Antibody	Nishida W <i>et al.</i> Gene cloning and nucleotide sequence of SM22 alpha from the chicken gizzard smooth muscle. <i>Biochem Int</i> 23:663-8 (1991).
COBLL1	COBL-like 1		
MGC24039	hypothetical protein MGC24039		
C9orf85	chromosome 9 open reading frame 85		
BLNK	B-cell linker	Abcam Ab4474 AntiHuman Rabbit Polyclonal Antibody	Kabak S <i>et al.</i> The direct recruitment of BLNK to immunoglobulin alpha couples the B-cell antigen receptor to distal signaling pathways. <i>Mol Cell Biol</i> 22:2524-35 (2002).
BCNP1	B-cell novel protein 1		
PDE3B	phosphodiesterase 3B, cGMP-inhibited	SantaCruz Biotechnology AntiHuman Rabbit Polyclonal Antibody	Liu H, and Maurel D.H. 1998 Expression of cyclic GMP-Inhibited phosphodiesterase 3A and 3B (PDE3A and PDE 3B) in rat tissues: differential subcellular localization and regulated expression Br. J. Pharmacol 125 1501:1510.
ENC1	ectodermal-neural cortex (with BTB-like domain)		

AKAP13	A kinase (PRKA) anchor protein 13		Other Family Antibodies Available including: Abcam Ab10346 AKAP12 AntiRat Rabbit Polyclonal Antibody Abcam Ab 25805 AKAP 9 AntiHuman Rabbit Polyclonal Antibody Abcam AB14096 AKAP 3 Goat Polyclonal Antibody
WDFY1	WD repeat and FYVE domain containing 1 (aka FENS-1)	Abcam Ab21695 AntiHuman Goat Polyclonal Antibody	
CDA	cytidine deaminase	Abcam Ab5197 Antihuman Rabbit Polyclonal Antibody	Duquette ML <i>et al.</i> AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. <i>Oncogene</i> 24:5791-8 (2005).
AGTRAP	angiotensin II receptor-associated protein		
ACTR2	ARP2 actin-related protein 2 homolog (yeast)		
GYG	glycogenin 1		
C6orf150	chromosome 6 open reading frame 150		
KLRD1	killer cell lectin-like receptor subfamily D, member 1	Abcam Ab19740 CD94 Ab (aka KLRD1 Ab) AntiHuman Mouse Monoclonal Antibody	Moretta A <i>et al.</i> Human natural killer cell receptors for HLA-class I molecules, Evidence that the Kp43 (CD94) molecule functions as receptor for HLA-B alleles. <i>J Exp Med</i> 180:545-55 (1994).
UTS2	urotensin 2	Abcam Ab14200 Antihuman Rabbit Polyclonal Antibody	
MS4A1	membrane-spanning 4-domains, subfamily A, member 1	Abcam Ab9475 AntiHuman Mouse Monoclonal Antibody	Mason DY <i>et al.</i> Antibody L26 recognizes an intracellular epitope on the B-cell-associated CD20 antigen. <i>Am J Pathol</i> 136:1215-22 (1990).
SPAP1	Fc receptor-like 2		
LIMS1	LIM and senescent cell antigen-like domains 1		
ZNF6	zinc finger protein 6 (CMPX1)		

WO 2007/048074

PCT/US2006/041600

ANK3	ankyrin 3, node of Ranvier (ankyrin G)		
KIAA1559	mouse zinc finger protein 14-like		
PPGB	protective protein for beta- galactosidase (galactosialido sis)		
GBP5	guanylate binding protein 5		
MT1E	metallothionei n 1E (functional)		
MGC20553 (FRMD3)	FERM domain containing 3		
G1P3	interferon, alpha- inducible protein (clone IFI-6-16)		
CEACAM1	carcinoembryo nic antigen- related cell adhesion molecule 1 (biliary glycoprotein)	Abcam Ab26279 Antihuman mouse monoclonal Antibody	
FLJ35036 (ZBTB38)	zinc finger and BTB domain containing 38		
P2RY12	purinergic receptor P2Y, G-protein coupled, 12	Alamone Labs #APR-012 AntiHuman Mouse Polyclonal Antibody	Queiroz, G. <i>et al.</i> (2003) <i>J. Pharmacol. Exp. Ther.</i> 307, 809.
HIST1H4H	histone 1, H4h		
ITGB3	integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)	Abcam Ab7167 AntiHumanMouse Monoclonal Antibody	Pittier R <i>et al.</i> Neurite extension and in vitro myelination within three-dimensional modified fibrin matrices. <i>J Neurobiol</i> 63:1-14 (2005).; Soldi R <i>et al.</i> Role of alphavbeta3 integrin in the activation of vascular endothelial growth factor receptor-2. <i>EMBO J</i> 18:882-92 (1999).
PRG1	proteoglycan 1, secretory granule		

BRD2	bromodomain containing 2	Abcam Ab19276 Antihuman Rabbit PolyClonal Antibody	
LTBP3	latent transforming growth factor beta binding protein 3	Abcam Ab21621 Antibovine Rabbit Polyclonal	
MAP4K3	mitogen-activated protein kinase kinase kinase 3		
NIPA2	non imprinted in Prader-Willi/Angelman syndrome 2		
SYN2	synapsin II	Abcam Ab12240 Antihuman Rabbit Monoclonal	
LEF1	Lymphoid enhancer-binding factor 1	Abcam Ab22884 Antihuman Rabbit Polyclonal; Ab12037 Antihuman Mouse Monoclonal;	
CLEC4D	C-type lectin domain family 4, member d		
PADI4	peptidyl arginine deiminase, type IV	Abcam AB26071 Antihuman goat polyclonal	
TNFAIP6	tumor necrosis factor, alpha-induced protein 6	Abcam Ab36380 Antihuman chicken polyclonal	
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	Abcam Ab1383 Antihuman rabbit polyclonal	

TABLE 16

Symbol	5' Primer (Sense Primer)		3' Primer (Antisense Primer)		Product
	Primer Sequence		Primer Sequence	Position	
LOC283130	AAGGCCAGAATCCCAGCTCAG	934	ATCCATCTGCATCCGGGACTTGAT	1044	111
MGC24039	AACAAGGGATCGCCTGCTCC	3717	ATAAGGGAGTTGACAGTCATGCGG	3833	117
ACTR2	ACCGGGTTTGTGAAGTGTGGAT	109	CACTTGCCTCATCACCAACCAT	253	145
AGTRAP	CATGGCCATCCTCAGCTTGC	325	AAGACCCAAGGAAACCAGTGTGGA	434	110
AKAP13	TCTCAGCCCGGTGATGGTC	8562	TGTAAGAGACTTGTGCACGCGG	8710	149
ANK3	ATACGCCATTACATCAAGCAGCAC		CCAAGGGCAGTATTCCCATTCA		
BANK1	TGCTGAAAGGCATGGTCACAAAG	1303	GCTGGGTTCTGTGTGGAAGGAATA	1449	147
BCNP1	GGCGCGTGCTGAAGAAATTCAA	1664	TTTGCAGCCTGGCTCGAGTTG	1782	119
BLNK	TTTCAGAACAGGAAGCTGGCGT	1172	GTTGTTTGAATCATGGCCAGAGC	1315	144
BRD2	TGCCTATGCTTGGCCTTTCT	2799	ATCTTCCGCTTGACAGTGCTGA	2909	111
C6orf150	CAAGAAGGCCTGCGCATTCAA	1098	AGCCGCCATGTTTCTTCTTGGA	1225	128
C9orf85	TTATTCCGTTGAATAAAGAAACAGA	545	TTTGATGGTCTCCTCCTGTG	694	150
CDA	ATCGCCAGTGACATGCAAGA	376	TACCATCCGGCTTGGTCATGTA	484	109
CEACAM1	ATTGGAGTAGTGGCCCTGGTTG		ATTGGAGTGGTCCTGAGTGTGGT		
COBLL1	AGGAAGAGTGAGGGCAGGTTCA	375	GCTGTAAGGCAGTCACACGACTAT	523	149
ENC1	CATGAGCTCACTCCATCACTCGAT	2196	AGCATTTACAAGGTGCAGCAGAT	2325	130
FLJ35036	CTCCGAGTTGTCTTGAAGTGAGG		TTGGCAAAGATTGGGCAGCAAG		
G1P3	CCTCCAAGGTCTAGTGACGGA	68	CCCCTGCAAGTGAAGAGCA	167	100
GBP5	CTGACTCTGCGAGCTTCTTCC		GATCACTACCTTGCTTTGGCCTT		
GYG	GGGACCAAGGCATACTGAACACA	538	TGGCACTTGCACCAAACACTTT	678	141
HIST1H4H	CACTTACACAGAGCACGCCAAA		TTAGCCACCGAAGCCGTAAAGA		
ITGB3	AGCTCATTGTTGATGCTTATGGGA	1118	ATACAAGACTTGAGGCCAGGGA	1232	
KIAA1559	TGTGGGAGAACTACAGCAAC	95	TGGACTCCAAATCAGGGCAGTA	238	144
KLRD1	cccagtatctatcttccatcatttg	670	tctctgccccagaacatt	819	150
LIMS1	AGGTGATGTGGTCTCTGCTCTT		CAGACTGGCTTCATGTCAAACCTCC		
LTBP3	TCTGCATCAACTTTCCCGGTCA	2024	TTGTTCTCGCATTTGCCATCCG	2162	139
MAP4K3	TGAACCTCCCGACAGTGATGGT	1141	AACCACCTTGGTGTCTTGT	1250	110
MGC20553	AGGTGCACAGAGCCAACATTAC		AATGGAACACCCCTCACCCAGA		
MS4A1	AAAGAACGTGCTCCAGACCC		TTCAGTTAGCCCAACCACTTCTTC		
MT1E	CTTGTTTCGTCTCACTGGTGT	1	ACTCTTCTTGCAAGGAGTGCAT	142	142
NIPA2	CTGGACTGCTGTCAATGGGA	471	ACTTACTAGCACGCTGAGAGC	577	107
OSBPL10	ATGGAGTCCAGGAACCTCTGG	2467	TTTGGGCTTCCATGGTGTGC	2616	150
P2RY12	GAACACTTTCTCATGTCCAGGGT		CCTGCAGAGTGGCATCTGGTATTT		
PDE3B	TGAGCAGGGAGATGAAGAAGCAA	3171	GCAAACCAGCAGCATCATAGGAG	3316	146
PPGB	GACACTGTTGTGGTCCAGGATTTG		TGGAAGCAGCTGTTGTGTTGG		
PRG1	TACTCAAATGCAGTCGGCTTGTCC		ACCCATTGGTACCTGGCTCTCT		
SPAP1	AGCCAGTGTATGTCAATGTGGG		GGAGTCCTTGTTCTCCAGAAGTGT		
SYN2	ACAGCTCAACAAGTCGCAGT	1470	AAAGAGGCTGGCAAAGGACT	1605	136
TAGLN	TGAAGGCAAAGACATGGCAG	429	TTCCCTCTTATGCTCCTGCG	561	133
UTS2	AAGCCGTCTATCTTGTGGCGAT	10	CGTCTTCATGAGGTGCTGAGAGTT	150	141
WDFY1	AAGGACATGAAGGTAGTGTGCGCT	632	ATGGCCCTGAAGTAACAGCGT	768	137
ZNF6	TGCAGGGCTGATCTGGGTCT		CTTCCACCGCCTGAATCCATACTT		

Table 17

Symbol	RefSeq	Affy ID	5' Primer Primer Sequence	Taqman Probe	3' Primer Primer Sequence
BANK1	NM_017935	219667_s_at	CTGAAAGGCATGGTCACAAA	TCCTTCCACACAGAAACCCAGCA	TCAGCTCCAATCTGCAGTCTG
BCNP1	NM_173544	230983_at	CGCGTGCTGAAGAAATTCAA	TTGGCGCAGAGGAGGTTCAT	ACTCAGGCAGCTCCTTTTTTG
CD163	NM_004244;	203645_s_at;			
	NM_203416	215049_x_at	GCAGCACATGGGAGATTGTCCTGTAA	AGCAAGTGGCCTCTGTAAATCTGCTCA	ATTGCACGAGGACAGTGTTTGGGA
CDA	NM_001785	205627_at	GCCGTCTCAGAAGGGTACAA	CAGGGCAATTGCTATCGCCA	CCAGTTGGTGCCAAACTCTC
MGC20553	NM_174938	229893_at	AGGTGCACAGAGCCAACATTAC	CTCATCATTAAACATGGAACCCCTGC	AATGGAACACCCCTCACCCAGA
MS4A1	NM_152866	217418_x_at	AAAGAACGTGCTCCAGACCC	CATAGTTCTCCTGTCAGCAGAAGA	TTCAGTTAGCCCAACCACTTCTTC

Table 1

AffySpotID	GeneSymbol	GeneID	pvalue	RNAAccess'nNo.	ProteinAcc'nNo.	FoldChange	GeneDescription
1552256_a_at	SCARB1	949	0.04	NM_005505	NP_005496	1.15	scavenger receptor class B, member 1
1552283_s_at	ZDHHC11	79844	9.69e-03	NM_024786	NP_079062	0.85	zinc finger, DHHC domain containing 11
1552287_s_at	AFG3L1	172	2.95e-03	NM_001132	NP_001123	0.89	AFG3 ATPase family gene 3-like 1 (yeast)
1552302_at	MGC20235	113277	0.04	NM_145041	NP_659478	1.17	hypothetical protein MGC20235
1552303_a_at	MGC20235	113277	0.01	NM_145041	NP_659478	1.13	hypothetical protein MGC20235
1552310_at	MGC29937	123207	0.04	NM_144597	NP_653198	0.91	hypothetical protein MGC29937
1552312_a_at	MFAP3	4238	0.03	NM_005927	NP_005918	0.89	microfibrillar-associated protein 3
1552315_at	GIMAP1	170575	1.77e-04	NM_130759	NP_570115	0.83	GTPase, IMAP family member 1
1552316_a_at	GIMAP1	170575	4.31e-03	NM_130759	NP_570115	0.92	GTPase, IMAP family member 1
1552318_at	GIMAP1	170575	6.71e-03	AK091818	BAC03754	0.88	GTPase, IMAP family member 1
1552320_a_at	NYD-SP28	85478	0.03	NM_033124	NP_149115	0.84	NYD-SP28 protein
1552364_s_at	MSI2	124540	0.03	NM_138962	NP_620412	1.14	musashi homolog 2 (Drosophila)
1552497_a_at	SLAMF6	114836	6.4e-03	NM_052931	NP_443163	0.89	SLAM family member 6
1552536_at	VTI1A	143187	5.01e-04	NM_145206	NP_660207	1.20	vesicle transport through interaction with t-SNAREs
homolog 1A (yeast)							
1552554_a_at	CARD12	58484	0.02	NM_021209	NP_067032	0.83	caspase recruitment domain family, member 12
1552584_at	IL12RB1	3594	0.02	NM_153701	NP_714912	0.85	interleukin 12 receptor, beta 1
1552607_at	CXorf52	286967	7.16e-03	AY168775	AAO11658	0.89	chromosome X open reading frame 52
1552610_a_at	JAK1	3716	0.03	NM_002227	NP_002218	0.92	Janus kinase 1 (a protein tyrosine kinase)
1552612_at	CDC42SE2	56990	0.04	NM_020240	NP_064625	0.86	CDC42 small effector 2
1552613_s_at	CDC42SE2	56990	0.02	NM_020240	NP_064625	0.86	CDC42 small effector 2
1552633_at	ZNF101	94039	0.03	NM_033204	NP_149981	0.89	zinc finger protein 101
1552634_a_at	ZNF101	94039	0.01	NM_033204	NP_149981	0.89	zinc finger protein 101
1552660_a_at	FLJ11193	55322	0.01	NM_018356	NP_060826	0.88	hypothetical protein FLJ11193
1552701_a_at	COP1	114769	0.01	NM_052889	NP_443121	0.84	CARD only protein
1552703_s_at	COP1	114769	0.03	NM_052889	NP_443121	0.86	CARD only protein
1552719_at	H63	113201	0.02	NM_138423	NP_612432	1.17	H63 breast cancer expressed gene
1552772_at	CLEC4D	338339	0.01	NM_080387	NP_525126	0.79	C-type lectin domain family 4, member D
1552773_at	CLEC4D	338339	0.04	NM_080387	NP_525126	0.85	C-type lectin domain family 4, member D
1552789_at	FLJ32803	151826	0.04	AK057365	BAB71453	0.87	hypothetical protein FLJ32803
1552889_a_at	XTP7	90332	0.04	NM_138568	NP_612635	1.04	protein 7 transactivated by hepatitis B virus X antigen
(HBxAg)							
1552928_s_at	TAB3	257397	0.01	NM_152787	NP_690000	0.90	TAK1-binding protein 3
1552942_at	FLJ30430	149373	0.03	AK054992	BAB70837	1.09	hypothetical protein FLJ30430
1553048_a_at	PIP5K2B	8396	5.95e-03	NM_138687	NP_619632	0.85	phosphatidylinositol-4-phosphate 5-kinase, type II, beta
1553112_s_at	CDK8	1024	2.96e-03	NM_001260	NP_001251	0.86	cyclin-dependent kinase 8
1553117_a_at	STK38	11329	9.2e-03	NM_007271	NP_009202	0.91	serine/threonine kinase 38
1553133_at	C9orf72	203228	0.02	NM_145005	NP_659442	0.85	chromosome 9 open reading frame 72
1553153_at	ATP6V0D2	245972	8.33e-03	NM_152565	NP_689778	1.29	ATPase, H ⁺ transporting, lysosomal 38kDa, V0 subunit d isoform 2
1553167_a_at	SLA/LP	51091	0.01	NM_153825	NP_722547	0.85	soluble liver antigen/liver pancreas antigen

1553167_a_at	SLA/LP	51091	0.01	NM_016955	NP_058651	0.85	soluble liver antigen/liver pancreas antigen
1553177_at	EAT2	117157	7.62e-03	NM_053282	NP_444512	0.73	SH2 domain-containing molecule EAT2
1553185_at	RASEF	158158	0.03	NM_152573	NP_689786	0.83	RAS and EF hand domain containing
1553193_at	ZNF441	126068	0.01	NM_152355	NP_689568	0.81	zinc finger protein 441
1553216_at	ZNF41	7592	0.01	NM_007130	NP_009061	0.87	zinc finger protein 41
1553222_at	OXER1	165140	0.03	NM_148962	NP_683765	1.16	oxoeicosanoid (OXE) receptor 1
1553225_s_at	ZNF75	7626	0.03	NM_007131	NP_009062	0.84	zinc finger protein 75 (D8C6)
1553227_s_at	WDR9	54014	0.02	NM_018963	NP_061836	0.88	WD repeat domain 9
1553252_a_at	BRWD3	254065	0.04	NM_153252	NP_694984	0.90	bromodomain and WD repeat domain containing 3
1553292_s_at	FLJ25006	124923	0.01	NM_144610	NP_653211	0.90	hypothetical protein FLJ25006
1553380_at	PARP15	165631	7.17e-03	NM_152615	NP_689828	0.83	poly (ADP-ribose) polymerase family, member 15
1553387_at	ATM	472	0.01	NM_138293	NP_612150	0.84	ataxia telangiectasia mutated (includes complementation groups A, C and D)
1553417_at	FLJ39058	283171	2.32e-03	NM_173580	NP_775851	1.17	hypothetical protein FLJ39058
1553530_a_at	ITGB1	3688	2.51e-04	NM_033669	NP_391989	0.91	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)
1553533_at	JPH1	56704	0.01	NM_020647	NP_065698	1.19	junctophilin 1
1553581_s_at	FLJ36754	285672	8.82e-03	NM_173829	NP_776190	0.88	hypothetical protein FLJ36754
1553587_a_at	POLE4	56655	3.48e-03	NM_019896	NP_063949	0.92	polymerase (DNA-directed), epsilon 4 (p12 subunit)
1553612_at	ZNF354B	117608	0.02	NM_058230	NP_478137	1.15	zinc finger protein 354B
1553647_at	CDYL2	124359	0.03	NM_152342	NP_689555	1.10	chromodomain protein, Y-like 2
1553678_a_at	ITGB1	3688	3.6e-03	NM_133376	NP_596867	0.92	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)
1553694_a_at	PIK3C2A	5286	0.04	NM_002645	NP_002636	0.91	phosphoinositide-3-kinase, class 2, alpha polypeptide
1553698_a_at	LOC126731	126731	0.04	NM_145257	NP_660300	0.83	LOC126731
1553703_at	FLJ90396	163049	0.04	NM_153358	NP_699189	1.05	hypothetical protein FLJ90396
1553709_a_at	FLJ14936	84950	6.11e-03	NM_032284	NP_115660	0.93	hypothetical protein FLJ14936
1553719_s_at	ZNF548	147694	0.01	NM_152909	NP_690873	0.90	zinc finger protein 548
1553750_a_at	MGC33371	143684	0.04	NM_144664	NP_653265	0.85	hypothetical protein MGC33371
1553861_at	MGC40368	255394	0.02	NM_152772	NP_689985	1.27	hypothetical protein MGC40368
1553881_at	MGC16142	84849	0.03	BC007365	NP_689985	0.93	hypothetical protein MGC16142
1553901_x_at	ZNF486	90649	0.03	BC008936	AAH08936	1.14	zinc finger protein 486
1553920_at	C9orf84	158401	0.01	NM_173521	NP_775792	1.21	chromosome 9 open reading frame 84
1553954_at	MGC19780	199857	0.02	NM_144988	NP_659425	0.87	hypothetical protein MGC19780
1553957_at	ZNF564	163050	1.31e-03	NM_144976	NP_659413	0.78	zinc finger protein 564
1553987_at	FLJ39616	51275	0.03	NM_016534	NP_057618	0.95	apoptosis-related protein PNAS-1
1554021_a_at	ZNF12	7559	0.01	NM_006956	NP_008887	0.89	zinc finger protein 12 (KOX 3)
1554039_at	C14orf159	80017	0.01	BC009182	AAH09182	1.19	chromosome 14 open reading frame 159
1554078_s_at	DNAJA3	9093	0.01	NM_005147	NP_005138	0.91	DnaJ (Hsp40) homolog, subfamily A, member 3
1554080_at	RQCD1	9125	0.04	BC007102	AAH07102	0.87	RCD1 required for cell differentiation1 homolog (S. pombe)
1554086_at	TUBGCP3	10426	0.01	BC007763	AAH07763	1.18	tubulin, gamma complex associated protein 3
1554089_s_at	SBDS	51119	1.26e-03	NM_016038	NP_057122	0.87	Shwachman-Bodian-Diamond syndrome
1554093_a_at	SNAPC5	10302	2.78e-03	NM_006049	NP_006040	0.93	small nuclear RNA activating complex, polypeptide 5, 19kDa

1554105_at	FAM11A	84548	0.01	NM_032508	NP_115897	0.88	family with sequence similarity 11, member A
1554116_s_at	PARP11	57097	0.01	BC031073	AAH31073	0.84	poly (ADP-ribose) polymerase family, member 11
1554149_at	C3orf4	56650	0.02	NM_019895	NP_063948	0.88	chromosome 3 open reading frame 4
1554153_a_at	PHF21A	51317	0.02	NM_016621	NP_057705	1.08	PHD finger protein 21A
1554167_a_at	GOLGA7	51125	0.01	NM_016099	NP_057183	0.95	golgi autoantigen, golgin subfamily a, 7
1554178_a_at	MGC39518	285172	1.41e-03	NM_173822	NP_776183	1.26	hypothetical protein MGC39518
1554182_at	TRIM50C	378108	5.81e-04	NM_198853	NP_942150	0.82	tripartite motif-containing 50C
1554205_s_at	ALS2CR15	130026	1.17e-03	NM_138468	NP_612477	1.23	amyotrophic lateral sclerosis 2 (juvenile)
chromosome region, candidate 15							
1554250_s_at	MGC45477	387277	0.03	BC033812	AAH33812	0.86	hypothetical protein MGC45477
1554306_at	ITPKB	3707	4.36e-03	BC015009	AAH15009	1.27	inositol 1,4,5-trisphosphate 3-kinase B
1554335_at	PSCD4	27128	0.02	BC017780	AAH17780	1.17	pleckstrin homology, Sec7 and coiled-coil domains 4
1554339_a_at	COG3	83548	0.02	NM_031431	NP_113619	0.89	component of oligomeric golgi complex 3
1554342_s_at	HEL308	113510	0.01	NM_133636	NP_598375	0.87	DNA helicase HEL308
1554406_a_at	CLEC7A	64581	0.03	NM_197947	NP_922938	0.83	C-type lectin domain family 7, member A
1554410_a_at	WBSCR16	81554	0.03	NM_030798	NP_110425	0.87	Williams-Beuren syndrome chromosome region 16
1554423_a_at	FBXO7	25793	0.01	NM_012179	NP_036311	0.92	F-box protein 7
1554424_at	FIP1L1	81608	0.03	NM_030917	NP_112179	0.88	FIP1 like 1 (S. cerevisiae)
1554451_s_at	DNAJC14	85406	2.27e-04	NM_032364	NP_115740	0.79	DnaJ (Hsp40) homolog, subfamily C, member 14
1554455_at	WINS1	55180	0.02	NM_181739	NP_859523	0.82	WINS1 protein with Drosophila Lines (Lin) homologous domain
1554462_a_at	DNAJB9	4189	0.03	NM_012328	NP_036460	0.92	DnaJ (Hsp40) homolog, subfamily B, member 9
1554464_a_at	CRTAP	10491	3.94e-03	NM_006371	NP_006362	0.78	cartilage associated protein
1554465_s_at	FLJ20344	55634	0.04	NM_017776	NP_060246	0.88	hypothetical protein FLJ20344
1554482_a_at	SARA2	51128	0.04	NM_016103	NP_057187	0.89	SAR1a gene homolog 2 (S. cerevisiae)
1554493_s_at	THADA	63892	8.65e-03	NM_198554	NP_940956	0.88	thyroid adenoma associated
1554493_s_at	THADA	63892	8.65e-03	NM_022065	NP_071348	0.88	thyroid adenoma associated
1554510_s_at	GHITM	27069	0.01	NM_014394	NP_055209	0.93	growth hormone inducible transmembrane protein
1554538_at	RHOF	54509	0.02	BC018208	AAH18208	0.89	ras homolog gene family, member F (in filopodia)
1554545_at	MGC26594	403314	0.04	NM_203454	NP_982279	1.09	hypothetical protein MGC26594
1554549_a_at	WDR20	91833	0.01	BC030654	AAH30654	1.16	WD repeat domain 20
1554557_at	ATP11B	23200	0.01	BC033880	AAH33880	0.80	ATPase, Class VI, type 11B
1554577_a_at	PSMD10	5716	5.57e-03	NM_002814	NP_002805	0.95	proteasome (prosome, macropain) 26S subunit, non-ATPase, 10
1554606_at	FLJ36090	153241	0.03	BC040527	AAH40527	0.90	hypothetical protein FLJ36090
1554616_at	SERPINB8	5271	3.88e-03	BC034528	AAH34528	1.11	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 8
1554627_a_at	ASCC1	51008	1.11e-04	NM_015947	NP_057031	0.91	activating signal cointegrator 1 complex subunit 1
1554637_a_at	CBFA2T2	9139	0.02	NM_175864	NP_787060	0.92	core-binding factor, runt domain, alpha subunit 2; translocated to, 2
1554661_s_at	FLJ32001	163882	0.02	NM_152609	NP_689822	0.93	hypothetical protein FLJ32001
1554670_at	GGA1	26088	0.02	NM_001001561	NP_001001561	0.90	golgi associated, gamma adaptin ear containing, ARF binding protein 1
1554678_s_at	HNRPDL	9987	0.01	NM_005463	NP_005454	0.88	heterogeneous nuclear ribonucleoprotein D-like
1554726_at	ZNF655	79027	0.02	NM_138494	NP_612503	0.87	zinc finger protein 655

1554747_a_at	SEPT2	4735	0.03	NM_004404	NP_004395	0.90	septin 2
1554783_s_at	ARHGEF2	9181	9.54e-03	NM_004723	NP_004714	0.82	rho/rac guanine nucleotide exchange factor (GEF) 2
1554794_a_at	UBE3C	9690	0.01	BC014029	AAH14029	0.86	ubiquitin protein ligase E3C
1554806_a_at	FBXO8	26269	0.03	NM_012180	NP_036312	0.91	F-box protein 8
1554821_a_at	ZBED1	9189	7.29e-03	NM_004729	NP_004720	1.21	zinc finger, BED domain containing 1
1554866_at	FLJ22104	65084	8.99e-03	BC030952	AAH30952	1.19	hypothetical protein FLJ22104
1554878_a_at	ABCD3	5825	0.02	NM_002858	NP_002849	0.86	ATP-binding cassette, sub-family D (ALD), member 3
1554890_a_at	TIA1	7072	0.03	NM_022037	NP_071320	0.80	TIA1 cytotoxic granule-associated RNA binding protein
1554899_s_at	FCER1G	2207	0.03	NM_004106	NP_004097	0.87	Fc fragment of IgE, high affinity I, receptor for; gamma
polypeptide							
1554930_a_at	FUT8	2530	0.02	NM_178155	NP_835368	0.90	fucosyltransferase 8 (alpha (1,6) fucosyltransferase)
1554952_s_at	NALP12	91662	0.02	NM_144687	NP_653288	1.22	NACHT, leucine rich repeat and PYD containing 12
1554952_s_at	NALP12	91662	0.02	NM_033297	NP_150639	1.22	NACHT, leucine rich repeat and PYD containing 12
1554986_a_at	SNX19	399979	0.04	BC031620	AAH31620	0.93	sorting nexin 19
1554997_a_at	PTGS2	5743	0.03	NM_000963	NP_000954	0.70	prostaglandin-endoperoxide synthase 2 (prostaglandin
G/H synthase and cyclooxygenase)							
1555001_at	IQCF1	132141	0.01	BC034228	AAH34228	1.17	IQ motif containing F1
1555016_at	IL16	3603	4.58e-03	BC040272	AAH40272	0.84	interleukin 16 (lymphocyte chemoattractant factor)
1555037_a_at	IDH1	3417	0.01	NM_005896	NP_005887	0.88	isocitrate dehydrogenase 1 (NADP+), soluble
1555042_at	TRPV5	56302	0.01	BC034740	AAH34740	1.16	transient receptor potential cation channel, subfamily V,
member 5							
1555088_x_at	STAT5B	6777	0.01	BC020868	AAH20868	1.29	signal transducer and activator of transcription 5B
1555091_at	PPM1F	9647	0.03	NM_014634	NP_055449	1.19	protein phosphatase 1F (PP2C domain containing)
1555092_at	FLJ12505	79805	0.03	BC028194	AAH28194	1.12	hypothetical protein FLJ12505
1555106_a_at	HSPC129	51496	7.34e-03	NM_016396	NP_057480	0.88	hypothetical protein HSPC129
1555141_a_at	LOC150159	150159	0.04	NM_139173	NP_631912	0.80	CG10806-like
1555192_at	ZNF277	11179	0.01	BC020626	AAH20626	0.80	zinc finger protein (C2H2 type) 277
1555193_a_at	ZNF277	11179	3.27e-03	NM_021994	NP_068834	0.84	zinc finger protein (C2H2 type) 277
1555226_s_at	C1orf43	25912	0.01	NM_138740	NP_620077	0.96	chromosome 1 open reading frame 43
1555226_s_at	C1orf43	25912	0.01	NM_015449	NP_056264	0.96	chromosome 1 open reading frame 43
1555241_at	LOC401466	401466	0.04	BC032347		0.82	hypothetical gene supported by BC055092
1555243_x_at	LOC401466	401466	0.04	BC032347		0.81	hypothetical gene supported by BC055092
1555247_a_at	RAPGEF6	51735	0.04	NM_016340	NP_057424	0.88	Rap guanine nucleotide exchange factor (GEF) 6
1555250_a_at	CPEB3	22849	0.04	NM_014912	NP_055727	1.14	cytoplasmic polyadenylation element binding protein 3
1555289_at	RGPR	89866	3.49e-03	BC009106	AAH09106	1.24	regucalcin gene promotor region related protein
1555301_a_at	C21orf106	23181	0.04	NM_206890	NP_996773	0.83	chromosome 21 open reading frame 106
1555334_s_at	SLC30A5	64924	2.8e-03	NM_022902	NP_075053	0.93	solute carrier family 30 (zinc transporter), member 5
1555334_s_at	SLC30A5	64924	2.8e-03	NM_024055	NP_076960	0.93	solute carrier family 30 (zinc transporter), member 5
1555351_s_at	PPHLN1	51535	4.61e-03	NM_016488	NP_057572	1.30	periphrilin 1
1555355_a_at	ETS1	2113	2.8e-03	NM_005238	NP_005229	0.80	v-ets erythroblastosis virus E26 oncogene homolog 1
(avian)							
1555372_at	BCL2L11	10018	0.01	AF455755	AAL57868	1.17	BCL2-like 11 (apoptosis facilitator)
1555373_at	C21orf114	378826	0.01	AF304443		1.14	chromosome 21 open reading frame 114
1555374_at	TTL	150465	0.03	AB071393	BAC06832	0.91	tubulin tyrosine ligase
1555423_at	SSH2	85464	0.04	AB072358	BAB84117	0.89	slingshot homolog 2 (Drosophila)

1555427_s_at protein	SYNCRIP	10492	3.29e-03	NM_006372	NP_006363	0.81	synaptotagmin binding, cytoplasmic RNA interacting
1555446_s_at	TMEM1	7109	0.01	NM_001001723	NP_001001723	1.12	transmembrane protein 1
1555467_a_at	CUGBP1	10658	0.02	NM_006560	NP_006551	0.80	CUG triplet repeat, RNA binding protein 1
1555468_at	NRP2	8828	0.01	BC009222	AAH09222	1.19	neuropilin 2
1555554_at	BASE	317716	0.02	NM_173859	NP_776258	1.14	breast cancer and salivary gland expression gene
1555559_s_at	USP25	29761	0.02	NM_013396	NP_037528	0.82	ubiquitin specific protease 25
1555613_a_at	ZAP70	7535	0.01	NM_001079	NP_001070	0.83	zeta-chain (TCR) associated protein kinase 70kDa
1555630_a_at	RAB34	83871	0.02	NM_031934	NP_114140	0.87	RAB34, member RAS oncogene family
1555638_a_at signals, 1	SAMSNI	64092	0.02	NM_022136	NP_071419	0.91	SAM domain, SH3 domain and nuclear localisation
1555691_a_at	KLRK1	22914	0.01	NM_007360	NP_031386	0.81	killer cell lectin-like receptor subfamily K, member 1
1555705_a_at	CKLFSF3	123920	0.02	NM_144601	NP_653202	0.93	chemokine-like factor super family 3
1555745_a_at	LYZ	4069	0.04	NM_000239	NP_000230	0.78	lysozyme (renal amyloidosis)
1555760_a_at	RBM15	64783	0.02	NM_022768	NP_073605	0.90	RNA binding motif protein 15
1555762_s_at	RBM15	64783	0.01	NM_022768	NP_073605	0.84	RNA binding motif protein 15
1555764_s_at homolog (yeast)	TIMM10	26519	5.48e-03	NM_012456	NP_036588	0.77	translocase of inner mitochondrial membrane 10
1555780_a_at	RHEB	6009	6.83e-05	NM_005614	NP_005605	0.94	Ras homolog enriched in brain
1555785_a_at	XRN1	54464	0.04	NM_019001	NP_061874	0.82	5'-3' exoribonuclease 1
1555797_a_at	ARPC5	10092	0.02	NM_005717	NP_005708	0.91	actin related protein 2/3 complex, subunit 5, 16kDa
1555803_a_at	FLJ10726	55216	3.1e-03	NM_018195	NP_060665	0.83	hypothetical protein FLJ10726
1555815_a_at	L3MBTL2	83746	0.02	NM_031488	NP_113676	0.91	l(3)mbt-like 2 (Drosophila)
1555827_at	CCNL1	57018	0.01	AY034790	AAK61551	0.94	cyclin L1
1555830_s_at	CHR2SYT	57488	4.6e-03	NM_020728	NP_065779	0.80	chr2 synaptotagmin
1555837_s_at 140kDa	POLR2B	5431	0.01	NM_000938	NP_000929	0.95	polymerase (RNA) II (DNA directed) polypeptide B,
1555851_s_at	SEPW1	6415	0.01	NM_003009	NP_003000	0.94	selenoprotein W, 1
1555852_at multifunctional protease 7)	PSMB8	5696	0.02	AI375915		0.94	proteasome (prosome, macropain) subunit, beta type, 8 (large
1555878_at	RPS24	6229	7.19e-03	AK094613		0.74	ribosomal protein S24
1555881_s_at	LZTS2	84445	2.95e-03	NM_032429	NP_115805	0.90	leucine zipper, putative tumor suppressor 2
1555882_at	SPIN3	169981	0.01	AL832091		0.85	spindlin family, member 3
1555886_at	C6orf210	57107	1.75e-03	BC029491		0.85	chromosome 6 open reading frame 210
1555889_a_at	CRTAP	10491	0.02	NM_006371	NP_006362	0.93	cartilage associated protein
1555933_at	MGC33867	90231	0.04	AB095933	BAC23109	1.13	hypothetical protein MGC33867
1555960_at	HINT1	3094	3.33e-03	AK054976		0.89	histidine triad nucleotide binding protein 1
1555961_a_at	HINT1	3094	0.01	NM_005340	NP_005331	0.87	histidine triad nucleotide binding protein 1
1555981_at	DKFZp762C2414	339201	1.98e-03	NM_178542	NP_848637	0.85	hypothetical protein DKFZp762C2414
1555996_s_at	EIF4A2	1974	0.01	AL137681		0.83	eukaryotic translation initiation factor 4A, isoform 2
1555998_at subunit d	ATP5H	10476	0.01	AK095671		0.87	ATP synthase, H+ transporting, mitochondrial F0 complex,
1556009_at	PEX13	5194	0.02	NM_002618	NP_002609	0.88	peroxisome biogenesis factor 13
1556035_s_at	ZNF207	7756	0.03	AL834501	CAD39159	0.94	zinc finger protein 207
1556090_at	HSMPP8	54737	0.01	BM970306		0.86	M-phase phosphoprotein, mpp8

1556105_at	PARD3	56288	0.03	AW079553	1.12	par-3 partitioning defective 3 homolog (C. elegans)
1556285_s_at	PPA2	27068	5.93e-03	NM_176869	0.89	inorganic pyrophosphatase 2
1556389_at	TNRC5	10695	4.39e-03	AF161347	0.82	trinucleotide repeat containing 5
1556420_s_at	YPEL2	388403	0.01	AK097253	1.14	yippee-like 2 (Drosophila)
1556471_at	SCML4	256380	2.72e-03	AK093571	0.82	sex comb on midleg-like 4 (Drosophila)
1556472_s_at	SCML4	256380	6.73e-04	AK093571	0.82	sex comb on midleg-like 4 (Drosophila)
1556588_at	LOC283687	283687	0.01	NM_175898	1.21	hypothetical protein LOC283687
1556732_at	EML4	27436	1.95e-03	BC039354	1.16	echinoderm microtubule associated protein like 4
1556787_s_at	PDE5A	8654	0.03	AW975338	1.14	phosphodiesterase 5A, cGMP-specific
1557036_at	ZBTB1	22890	0.02	BU950380	1.22	zinc finger and BTB domain containing 1
1557059_at	MKI67P	84365	0.02	AK098264	1.12	MKI67 (FHA domain) interacting nucleolar phosphoprotein
1557066_at	LUC7L	55692	0.01	BX647161	0.86	LUC7-like (S. cerevisiae)
1557067_s_at	LUC7L	55692	6.71e-03	NM_018032	0.85	LUC7-like (S. cerevisiae)
1557145_at	STK38	11329	5.07e-03	BU617137	1.12	serine/threonine kinase 38
1557170_at	NEK8	284086	0.03	NM_178170	1.15	NIMA (never in mitosis gene a)-related kinase 8
1557227_s_at	TPR	7175	0.01	NM_003292	0.91	translocated promoter region (to activated MET oncogene)
1557246_at	KIDINS220	57498	0.04	AA992480	0.83	likely homolog of rat kinase D-interacting substance of 220
kDa						
1557300_s_at	WHSC2	7469	0.02	AI741292	1.11	Wolf-Hirschhorn syndrome candidate 2
1557309_at	FAM31B	163486	0.01	BC016588	1.24	family with sequence similarity 31, member B
1557360_at	LRPPRC	10128	7.07e-05	BC038181	0.83	leucine-rich PPR-motif containing
1557394_at	DLGAP4	22839	0.03	AK057672	1.07	dises, large (Drosophila) homolog-associated protein 4
1557477_at	STIM1	6786	0.01	BQ022900	1.15	stromal interaction molecule 1
1557527_at	RUNX1	861	0.01	BU789637	1.17	runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)
1557539_at	KIAA1545	57666	0.02	BC008052	1.09	KIAA1545 protein
1557551_at	PSCD1	9267	2.39e-03	CA442689	1.27	pleckstrin homology, Sec7 and coiled-coil domains 1 (cytohesin 1)
1557586_s_at	ATP6V1H	51606	0.01	AW468382	1.08	ATPase, H ⁺ transporting, lysosomal 50/57kDa, V1 subunit H
1557611_at	TTL1A	9654	9.89e-03	AW779022	1.21	tubulin tyrosine ligase-like family, member 4
1557666_s_at	C9orf98	158067	5.7e-03	AW511637	1.16	chromosome 9 open reading frame 98
1557684_at	ZNF286	57335	0.04	AK021478	0.93	zinc finger protein 286
1557706_at	ZHX2	22882	2.67e-03	AF075099	1.18	zinc fingers and homeoboxes 2
1557749_at	EHBP1L1	254102	0.04	AK092750	1.13	EH domain binding protein 1-like 1
1557799_at	C11orf31	280636	0.04	AF085883	0.91	chromosome 11 open reading frame 31
1557820_at	AFG3L2	10939	0.01	AA905788	0.88	AFG3 ATPase family gene 3-like 2 (yeast)
1557852_at	PHC2	1912	0.04	AW418842	0.85	polyhomeotic-like 2 (Drosophila)
1557915_s_at	GSTO1	9446	6.92e-03	NM_004832	0.95	glutathione S-transferase omega 1
1557945_at	TCTE3	6991	0.04	BC043594	0.91	t-complex-associated-testis-expressed 3
1557950_at	SDCCAG1	9147	2.28e-03	NM_004713	0.87	serologically defined colon cancer antigen 1
1557954_at	RBBP7	5931	0.02	BM757251	0.94	retinoblastoma binding protein 7
1557965_at	MGC61716	130916	0.02	NM_182501	0.86	hypothetical protein MGC61716
1557987_at	LOC23117	23117	0.02	BC042832	0.77	KIAA0220-like protein
1558014_s_at	MLSTD2	84188	0.02	NM_032228	0.90	male sterility domain containing 2
1558094_s_at	C3orf19	51244	0.04	NM_016474	0.90	chromosome 3 open reading frame 19

1558097_at	MGC50372	253143	0.02	BX647272	0.93	hypothetical protein MGC50372
1558120_at	DDX3X	1654	1.99e-04	BE379787	1.29	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked
1558136_s_at	TAF11	6882	0.01	NM_005643	0.87	TAF11 RNA polymerase II, TATA box binding protein
(TBP)-associated factor, 28kDa						
1558199_at	FNI	2335	0.04	BX641150	1.11	fibronectin I
1558208_at	TARDBP	23435	0.01	AL039795	0.79	TAR DNA binding protein
1558217_at	FLJ31952	146857	0.01	NM_144682	0.83	hypothetical protein FLJ31952
1558233_s_at	ATF1	466	0.04	NM_005171	0.90	activating transcription factor 1
1558289_at	RFT1	91869	9.2e-03	NM_052859	1.13	RFT1 homolog (S. cerevisiae)
1558292_s_at	PIGW	284098	3.28e-03	NM_178517	0.82	phosphatidylinositol glycan, class W
1558304_s_at	SEN54L	283989	0.01	NM_207346	0.86	likely homolog of yeast SEN54
1558327_at	ZDHHC18	84243	7.35e-03	NM_032283	1.21	zinc finger, DHHC domain containing 18
1558392_at	SYNE2	23224	0.04	NM_015180	0.89	spectrin repeat containing, nuclear envelope 2
1558409_at	LOC442612	442612	0.02	AW974642	0.82	LOC442612
1558448_a_at	JMJD2B	23030	8.05e-03	BC032415	1.18	jumonji domain containing 2B
1558450_at	A2M	2	0.01	NM_000014	1.18	alpha-2-macroglobulin
1558511_s_at	CHR2SYT	57488	8.42e-03	NM_020728	0.80	chr2 synaptotagmin
1558648_at	MGC20481	90416	0.01	AL832032	1.14	hypothetical protein MGC20481
1558675_s_at	SDCCAG1	9147	0.04	NM_004713	0.92	serologically defined colon cancer antigen 1
1558685_a_at	LOC158960	158960	0.03	BC009467	1.08	hypothetical protein BC009467
1558692_at	MGC31963	112770	0.02	AW090182	0.87	kidney predominant protein NCU-GI
1558693_s_at	MGC31963	112770	0.01	NM_144580	0.84	kidney predominant protein NCU-GI
1558742_at	DEXI	28955	7.02e-03	BE899474	1.10	dexamethasone-induced transcript
1558745_at	LOC92017	92017	0.02	BC029857	0.87	similar to RIKEN cDNA 4933437K13
1558759_s_at	CLASP2	23122	0.04	AB014527	1.09	cytoplasmic linker associated protein 2
1558761_a_at	C9orf10OS	158293	1.75e-03	AK093641	1.16	chromosome 9 open reading frame 10 opposite strand
1558802_at	BBP	83941	9.61e-03	BG401696	0.83	beta-amyloid binding protein precursor
1558953_s_at	Cep164	22897	0.04	NM_014956	0.87	KIAA1052 protein
1558996_at	FOXP1	27086	0.02	BX647682	0.89	forkhead box P1
1559025_at	SEPT9	10801	0.03	AL833770	0.86	septin 9
1559064_at	NUP153	9972	0.02	BC035502	0.88	nucleoporin 153kDa
1559094_at	FBXO9	26268	1.11e-03	AK095307	0.90	F-box protein 9
1559096_x_at	FBXO9	26268	5.53e-04	AK095307	0.92	F-box protein 9
1559101_at	FYN	2534	0.02	AK090692	0.76	FYN oncogene related to SRC, FGR, YES
1559128_at	HSDL2	84263	0.02	BC036620	1.25	hydroxysteroid dehydrogenase like 2
1559203_s_at	KRAS	3845	0.04	BC029545	1.27	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
1559347_at	CNOT4	4850	0.03	BI599409	1.11	CCR4-NOT transcription complex, subunit 4
1559391_s_at	B4GALT5	9334	0.01	AI084451	1.19	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase,
polypeptide 5						
1559399_s_at	ZCCHC10	54819	7.35e-03	BC015986	0.86	zinc finger, CCHC domain containing 10
1559413_at	MGC40368	255394	0.04	AF306858	1.13	hypothetical protein MGC40368
1559441_s_at	CYP4V2	285440	0.03	AK095895	1.21	cytochrome P450, family 4, subfamily V, polypeptide 2
1559502_s_at	LRRC25	126364	0.02	NM_145256	1.13	leucine rich repeat containing 25
1559582_at	RHOQ	23433	0.01	BC033251	1.15	ras homolog gene family, member Q
1559716_at	C18orf37	125476	0.02	AK095502	1.10	chromosome 18 open reading frame 37

1559739_at	CHPT1	56994	2.79e-03	AK025141	0.83	choline phosphotransferase 1
1559756_at	DKFZp667F0711	399716	0.02	AL713753	CAD28526	1.21 hypothetical protein DKFZp667F0711
1559881_s_at	ZNF12	7559	0.04	NM_006956	NP_008887	0.87 zinc finger protein 12 (KOX 3)
1559883_s_at	SAMHD1	25939	0.02	NM_015474	NP_056289	0.83 SAM domain and HD domain 1
1559942_at	HIC	29969	0.02	NM_199072	NP_951038	1.14 I-mfa domain-containing protein
1559946_s_at	RUVBL2	10856	0.03	NM_006666	NP_006657	1.12 RuvB-like 2 (E. coli)
1559954_s_at	DDX42	11325	7.69e-03	NM_007372	NP_031398	0.88 DEAD (Asp-Glu-Ala-Asp) box polypeptide 42
1560069_at	LOC389072	389072	0.02	BC020812	AAH20812	0.89 hypothetical protein LOC389072
1560474_at	NIPBL	25836	3.9e-03	BM728509	1.36	Nipped-B homolog (Drosophila)
1560486_at	STXBP3	6814	8.05e-03	BC028028	AAH28028	1.22 syntaxin binding protein 3
1560622_at	TPARL	55858	0.04	AK000203	0.90	TPA regulated locus
1560679_at	LOC151438	151438	0.02	AK055877	0.74	hypothetical protein LOC151438
1560741_at	UBE3A	7337	6.77e-03	AL832250	1.18	ubiquitin protein ligase E3A (human papilloma virus E6-associated protein, Angelman syndrome)
1560776_at	LOC201158	201158	0.03	AU121725	1.12	similar to CGI-148 protein
1560986_a_at	EHBPI	23301	9.33e-03	AF086019	1.15	EH domain binding protein 1
1561128_at	STK10	6793	0.03	AF088069	1.08	serine/threonine kinase 10
1561135_at	FLJ34870	401013	0.01	AF085980	1.15	FLJ34870 protein
1561303_at	LOC440384	440384	0.02	XM_498646	XP_498646	1.17 hypothetical gene supported by BC042734
1561442_at	LOC283585	283585	4.01e-04	BC034577	AAH34577	1.33 hypothetical protein LOC283585
1561453_at	KIAA1276	27146	0.03	BC037915	1.10	KIAA1276 protein
1561467_at	KIAA0804	23355	0.01	BC039111	1.17	KIAA0804 protein
1561502_x_at	LOC348180	348180	0.04	AW575915	0.88	hypothetical protein LOC348180
1561654_at	LIMK2	3985	0.03	AF085978	1.23	LIM domain kinase 2
1561690_at	SSBP2	23635	0.01	BC041991	1.19	single-stranded DNA binding protein 2
1561890_at	C10orf128	170371	0.02	BC037922	1.18	chromosome 10 open reading frame 128
1561908_a_at	HS3ST3B1	9953	0.02	AL832823	0.87	heparan sulfate (glucosamine) 3-O-sulfotransferase 3B1
1562051_at	LOC284757	284757	0.02	AK092805	0.89	hypothetical protein LOC284757
1562059_at	SVIL	6840	0.03	AW020871	1.21	supervillin
1562256_at	NALP1	22861	5.97e-03	AL832400	1.15	NACHT, leucine rich repeat and PYD (pyrin domain) containing 1
1562257_x_at	NALP1	22861	1.54e-03	AL832400	1.14	NACHT, leucine rich repeat and PYD (pyrin domain) containing 1
1562259_at	TEX9	374618	0.03	AL832050	0.80	testis expressed gene 9
1562305_x_at	ZNF306	80317	0.01	AK091225	1.18	zinc finger protein 306
1562314_at	NCOA1	8648	1.16e-03	AL713641	1.20	nuclear receptor coactivator 1
1562416_at	FLNB	2317	0.01	AI524619	0.89	filamin B, beta (actin binding protein 278)
1562505_at	STK35	140901	0.03	BC035700	1.18	serine/threonine kinase 35
1562612_at	ME2	4200	0.03	BC020933	0.88	malic enzyme 2, NAD(+)-dependent, mitochondrial
1562619_at	TXNDC6	347736	0.03	NM_178130	NP_835231	1.11 thioredoxin domain containing 6
1562650_at	LOC285527	285527	0.01	AL833444	0.82	hypothetical protein LOC285527
1562673_at	CHES1	1112	0.01	BC043161	1.17	checkpoint suppressor 1
1562733_at	LOC401544	401544	3.57e-03	BC043559	0.89	hypothetical gene supported by BC043559
1562894_at	ELMO1	9844	0.02	BC022406	1.16	engulfment and cell motility 1 (ced-12 homolog, C. elegans)
1562898_at	ATP11A	23250	0.04	AL713719	1.16	ATPase, Class VI, type 11A

1563455_at	KIAA0999	23387	0.01	NM_025164	NP_079440	1.12	KIAA0999 protein
1563498_s_at	LOC283130	283130	0.02	AK090434	BAC03415	0.83	hypothetical protein LOC283130
1563621_at	KIAA1706	80820	0.04	AL713724		1.18	KIAA1706 protein
1563657_at	PLXND1	23129	7.09e-03	NM_015103	NP_055918	1.18	plexin D1
1563687_a_at	KIAA0826	23045	0.02	U80082	AAD00351	0.87	KIAA0826
1563796_s_at	KIAA1970	124454	0.03	AK095998		0.91	KIAA1970 protein
1564063_a_at	ATP11B	23200	0.04	BC033880	AAH33880	0.84	ATPase, Class VI, type 11B
1564093_at	NEK1	4750	0.01	AK024912		0.88	NIMA (never in mitosis gene a)-related kinase 1
1564139_at	LOC144571	144571	0.01	AK056852		0.70	hypothetical protein LOC144571
1564203_at	LOC147004	147004	0.02	AK057317		1.10	hypothetical protein LOC147004
1564301_a_at	RIP	84268	0.02	AK096001		1.14	RPA interacting protein
1564310_a_at	PARP15	165631	0.01	NM_152615	NP_689828	0.87	poly (ADP-ribose) polymerase family, member 15
1564443_at	RFP2OS	448961	0.02	AF529010		1.11	ret finger protein 2 opposite strand
1564494_s_at	P4HB	5034	0.01	NM_000918	NP_000909	0.89	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide (protein disulfide isomerase; thy
1564521_x_at	SKB1	10419	0.02	NM_006109	NP_006100	0.81	SKB1 homolog (S. pombe)
1564798_at	EEA1	8411	0.01	AK000811		1.09	early endosome antigen 1, 162kD
1565269_s_at	ATF1	466	0.01	NM_005171	NP_005162	0.89	activating transcription factor 1
1565358_at	RARA	5914	0.01	AJ417079		1.17	retinoic acid receptor, alpha
1565657_at	CKLFSF6	54918	0.04	BC042995		0.75	chemokine-like factor super family 6
1565672_at	LOC440703	440703	0.01	BC031957		1.14	hypothetical gene supported by AK026359
1565673_at	FCGR2A	2212	3.78e-03	AF416711	AAL78867	1.30	Fc fragment of IgG, low affinity IIa, receptor (CD32)
1565717_s_at	FUS	2521	2.44e-05	NM_004960	NP_004951	0.77	fusion (involved in t(12;16) in malignant liposarcoma)
1565804_at	LOC360030	360030	5.07e-03	AK025161		1.24	homeobox C14
1565830_at	KIAA1731	85459	0.01	AL833615		0.83	KIAA1731 protein
1565978_a_at	FCHO2	115548	0.01	N33616		1.18	FCH domain only 2
1566142_at	GBF1	8729	2.98e-03	AK095268		1.15	golgi-specific brefeldin A resistance factor 1
1566144_at	LOC401131	401131	9.73e-03	AK098337		1.15	hypothetical LOC401131
1566509_s_at	FBXO9	26268	8.32e-03	NM_012347	NP_036479	0.93	F-box protein 9
1566577_at	DPYSL5	56896	0.03	AL831879		1.20	dihydropyrimidinase-like 5
1566959_at	GAB2	9846	0.02	AL049273		1.34	GRB2-associated binding protein 2
1567032_s_at	ZNF160	90338	0.02	NM_033288	NP_150630	1.19	zinc finger protein 160
1567044_s_at	C19orf13	26065	0.02	AF085866		1.13	chromosome 19 open reading frame 13
1567107_s_at	TPM4	7171	0.01	NM_003290	NP_003281	1.08	tropomyosin 4
1567213_at	PNN	5411	0.01	U59479	AAB03391	0.92	pinin, desmosome associated protein
1567214_a_at	PNN	5411	0.02	NM_002687	NP_002678	0.90	pinin, desmosome associated protein
1568592_at	LOC400368	400368	0.02	XM_378511	XP_378511	0.85	hypothetical gene supported by BC031266
1568619_s_at	LOC162073	162073	0.01	BC015343		1.17	hypothetical protein LOC162073
1568620_at	CSAD	51380	0.04	BC018042		1.16	cysteine sulfinic acid decarboxylase
1568704_a_at	CHERP	10523	0.03	NM_006387	NP_006378	1.06	calcium homeostasis endoplasmic reticulum protein
1568720_at	ZNF505	81931	0.01	NM_031218	NP_112495	1.24	zinc finger protein 505
1568780_at	LOC389634	389634	5.93e-03	AK124896		1.25	hypothetical LOC389634
1568830_at	IRAK3	11213	0.03	BC029493		0.88	interleukin-1 receptor-associated kinase 3
1568836_at	CLK4	57396	9.75e-03	BC034633		0.84	CDC-like kinase 4
1568856_at	NBR1	4077	0.04	NM_005899	NP_005890	0.77	neighbor of BRCA1 gene 1

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 3
CONTENANT LES PAGES 1 À 230

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 3
CONTAINING PAGES 1 TO 230

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

AMENDED CLAIMS**received by the International Bureau on 08 May 2007 (08.05.2007)**

1. A method of testing for one or more colorectal pathologies in a test subject, the method comprising:

(a) providing data representing a level of one or more products of each of at least two biomarkers in a sample isolated and/or derived from said test subject, wherein said biomarkers are genes selected from the group consisting of BCNP1, CD163, CDA, MS4A1, BANK1, and MGC20553; and

(b) applying to said data a formula based on (i) a dataset representing levels of one or more products of each of said biomarkers in each subject of a reference population having said one or more pathologies, and (ii) a dataset representing levels of one or more products of each of said biomarkers in each subject of a reference population not having said one or more pathologies, thus providing an indication of a probability that said test subject has said one or more colorectal pathologies.

2. A computer-based method for testing for one or more colorectal pathologies in a test subject, the method comprising:

inputting, to a computer, data representing a level of products of each of at least two biomarkers in a sample isolated and/or derived from a test subject, wherein said biomarkers are genes selected from the group consisting of BCNP1, CD163, CDA, MS4A1, BANK1 and MGC20553; and

causing said computer to apply to said data, a formula based on (i) a dataset representing levels of one or more products of each of said biomarkers in each subject of a reference population having said one or more pathologies, and (ii) a dataset representing levels of one or more products of each of said biomarkers in each subject of a reference population not having said one or more pathologies, thus providing an indication of a probability that said test subject has said one or more colorectal pathologies.

3. The method of claims 1 or 2, wherein said formula has a form:

$$V = C + \sum \beta_i X_i,$$

wherein V is a value indicating a probability that said test subject has said one or more colorectal pathologies, X_i is a level of products of an i th biomarker of said biomarkers, β_i is a coefficient, and C is a constant.

4. The method of claims 1 or 2, wherein said formula has a form:

$$V = C + \sum \beta_{ij}(X_i / X_j),$$

wherein V is a value indicating a probability that said test subject has said one or more colorectal pathologies, X_i is a level of products of an i th biomarker of said biomarkers, X_j is a level of products of a j th biomarker of said biomarkers where the i th biomarker is not the j th biomarker, β_{ij} is a coefficient and C is a constant.

5. The method of claims 1 or 2, wherein said formula is derived by logistic regression.

6. The method of claims 1 or 2, wherein said sample is selected from the group consisting of blood, lymph and lymphoid tissue.

7. The method of claims 1 or 2 wherein said products in said sample are RNA.

8. The method of claims 1 or 2, wherein said sample is selected from the group consisting of a sample of serum-reduced blood, a sample of erythrocyte-reduced blood, a sample of serum-reduced and erythrocyte-reduced blood, a sample of unfractionated cells of lysed blood, and a sample of fractionated blood.

9. The method of claims 1 or 2, wherein said products in said sample are RNA, and whereas said data represents a level of cDNA, EST and/or PCR product derived from said RNA.

10. A kit comprising packaging and containing at least two primer sets, wherein each set of which is able to generate an amplification product by selective amplification of at least a portion of a polynucleotide complementary to one or more RNA products of a biomarker, wherein said biomarker is a gene selected from the group consisting of: BCNP1, CD163, CDA, MS4A1, BANK1, and MGC20553; and wherein each set of said primer sets is selective for a different biomarker of said group.

11. The kit of claim 10, wherein said polynucleotide is selected from the group consisting of total RNA, mRNA, DNA, cDNA and EST.

12. The kit of claim 10, further comprising two or more oligonucleotide probes, wherein each of said probes is capable of selectively hybridizing to either a sense or an antisense strand of said amplification product.

13. The kit of claims 10, 11 or 12, further comprising two or more components selected from the group consisting of: a thermostable polymerase, a reverse transcriptase, deoxynucleotide triphosphates, nucleotide triphosphates and enzyme buffer.

14. The kit of claims 10, 11, 12 or 13, further comprising a computer-readable medium encoded with instructions for applying a formula to data representing levels of said amplification product in a sample isolated and/or derived from a test subject, wherein said formula is said based on (i) a dataset representing levels of one or more products of each of said biomarkers in each subject of a reference population having one or more colorectal pathologies, and (ii) a dataset representing levels of one or more products of each of said biomarkers in each subject of a reference population not having said one or more pathologies, thus providing an indication of a probability that said test subject has said one or more colorectal pathologies.

15. The kit of claim 14, wherein said formula has a form:

$$V = C + \sum \beta_i X_i,$$

wherein V is a value indicating a probability that said test subject has said one or more colorectal pathologies, X_i is a level of products of an i th biomarker of said biomarkers, β_i is a coefficient, and C is a constant.

16. The kit of claim 14, wherein said formula has a form:

$$V = C + \sum \beta_{ij} (X_i / X_j),$$

wherein V is a value indicating a probability that said test subject has said one or more colorectal pathologies, X_i is a level of products of an i th biomarker of said biomarkers, X_j is a level of products of a j th biomarker of said biomarkers where the i th biomarker is not the j th biomarker, β_{ij} is a coefficient, and C is a constant.

17. The kit of claims 14, 15 or 16, wherein said sample is selected from the group consisting of blood, lymph and lymphoid tissue.

18. The kit of claims 14, 15 or 16, wherein said sample is selected from the group consisting of a sample of serum-reduced blood, a sample of erythrocyte-reduced blood, a sample of serum-reduced and erythrocyte-reduced blood, a sample of unfractionated cells of lysed blood and a sample of fractionated blood.

19. A computer-readable medium having instructions stored thereon that are operable when executed by a computer to apply a formula to data representing a level of one or more products of each of at least two biomarkers in a sample isolated and/or derived from a test subject, wherein said biomarkers are genes selected from the group consisting of BCNP1, CD163, CDA, MS4A1, BANK1, and MGC20553, wherein said formula is based on (i) a dataset representing levels of one or more products of each of said biomarkers in each subject of a reference population having one or more colorectal pathologies, and (ii) a dataset representing levels of one or more products of each of said biomarkers in each subject of a reference population not having said one or more pathologies, thus providing an indication of a probability that said test subject has said one or more colorectal pathologies.

20. The computer-readable medium of claim 19, wherein said formula has a form:

$$V = C + \sum \beta_i X_i,$$

wherein V is a value indicating a probability that said test subject has said one or more colorectal pathologies, X_i is a level of products of an i th biomarker of said biomarkers, β_i is a coefficient, and C is a constant.

21. The computer-readable medium of claim 19, wherein said formula has a form:

$$V = C + \sum \beta_{ij} (X_i / X_j),$$

wherein V is a value indicating a probability that said test subject has said one or more colorectal pathologies, X_i is a level of products of an i th biomarker of said biomarkers, and X_j is a level of products of a j th biomarker of said biomarkers, where the i th biomarker is not the j th biomarker, β_{ij} is a coefficient and C is a constant.

22. The computer-readable medium of claim 19 wherein said formula is derived by logistic regression.

23. The computer-readable medium of claim 19 wherein said sample is selected from the group consisting of blood, lymph and lymphoid tissue.

24. The computer-readable medium of claims 19 wherein said products in said sample are RNA.

25. The computer-readable medium of claim 19, wherein said sample is selected from the group consisting of a sample of serum-reduced blood, a sample of erythrocyte-reduced blood, a sample of serum-reduced and erythrocyte-reduced blood, a sample of unfractionated cells of lysed blood, and a sample of fractionated blood.

26. The computer-readable medium of claim 19 wherein said products in said sample are RNA, and whereas said data represents a level of cDNA, EST and/or PCR product derived from said RNA.

27. A computer system for providing an indication of a probability that a test subject has one or more colorectal pathologies, the computer system comprising
a processor; and

a memory configured with instructions that, when executed, cause said processor to provide a user with said indication, wherein said instructions comprise applying a formula to data representing a level of one or more products of each of at least two biomarkers in a sample isolated and/or derived from said test subject, wherein said biomarkers are genes selected from the group consisting of BCNP1, CD163, CDA, MS4A1, BANK1, and MGC20553; and wherein said formula is based on (i) a dataset representing levels of one or more products of each of said biomarkers in each subject of a reference population having said one or more pathologies, and (ii) a dataset representing levels of one or more products of each of said biomarkers in each subject of a reference population not having said one or more pathologies, thus providing said indication of said probability that said test subject has said one or more colorectal pathologies.

28. The system of claim 27, wherein said formula has a form:

$$V = C + \sum \beta_i X_i,$$

wherein V is a value indicating a probability that said test subject has said one or more colorectal pathologies, X_i is a level of products of an i th biomarker of said biomarkers, β_i is a coefficient, and C is a constant.

29. The system of claim 27, wherein said formula has a form:

$$V = C + \sum \beta_{ij}(X_i / X_j),$$

wherein V is a value indicating a probability that said test subject has said one or more colorectal pathologies, X_i is a level of products of an i th biomarker of said biomarkers, and X_j is a level of products of a j th biomarker of said biomarkers, where the i th biomarker is not the j th biomarker, β_{ij} is a coefficient, and C is a constant.

30. The system of claim 27, wherein said formula is derived by logistic regression.

31. The system of claim 27, wherein said sample is selected from the group consisting of blood, lymph and lymphoid tissue.

32. The system of claim 27, wherein said products in said sample are RNA.

33. The system of claim 27 wherein said sample is selected from the group consisting of a sample of serum-reduced blood, a sample of erythrocyte-reduced blood, a sample of serum-reduced and erythrocyte-reduced blood, a sample of unfractionated cells of lysed blood, and a sample of fractionated blood.

34. The system of claim 27 wherein said products in said sample are RNA, and whereas said data represents a level of cDNA, EST and/or PCR product derived from said RNA.

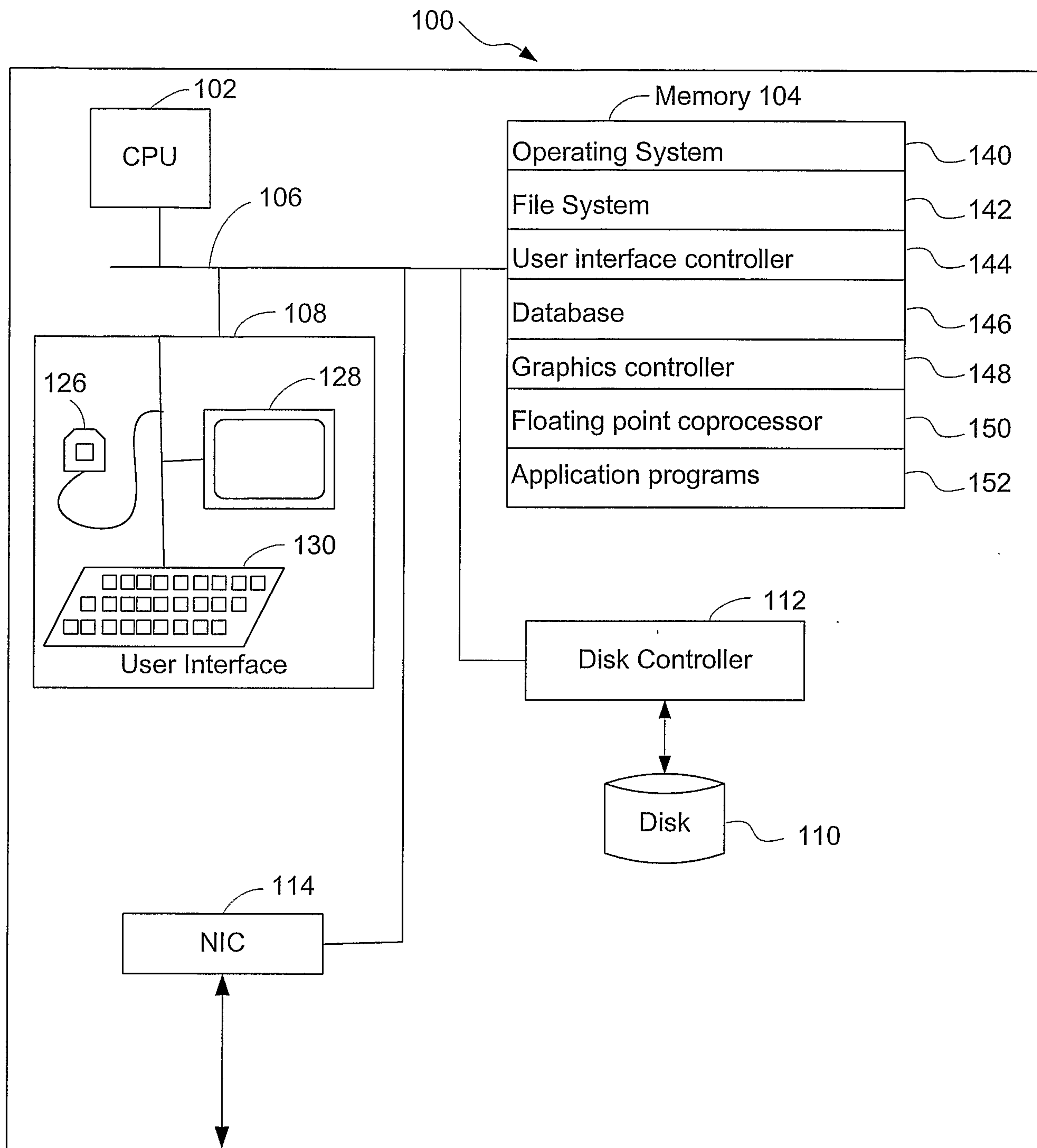
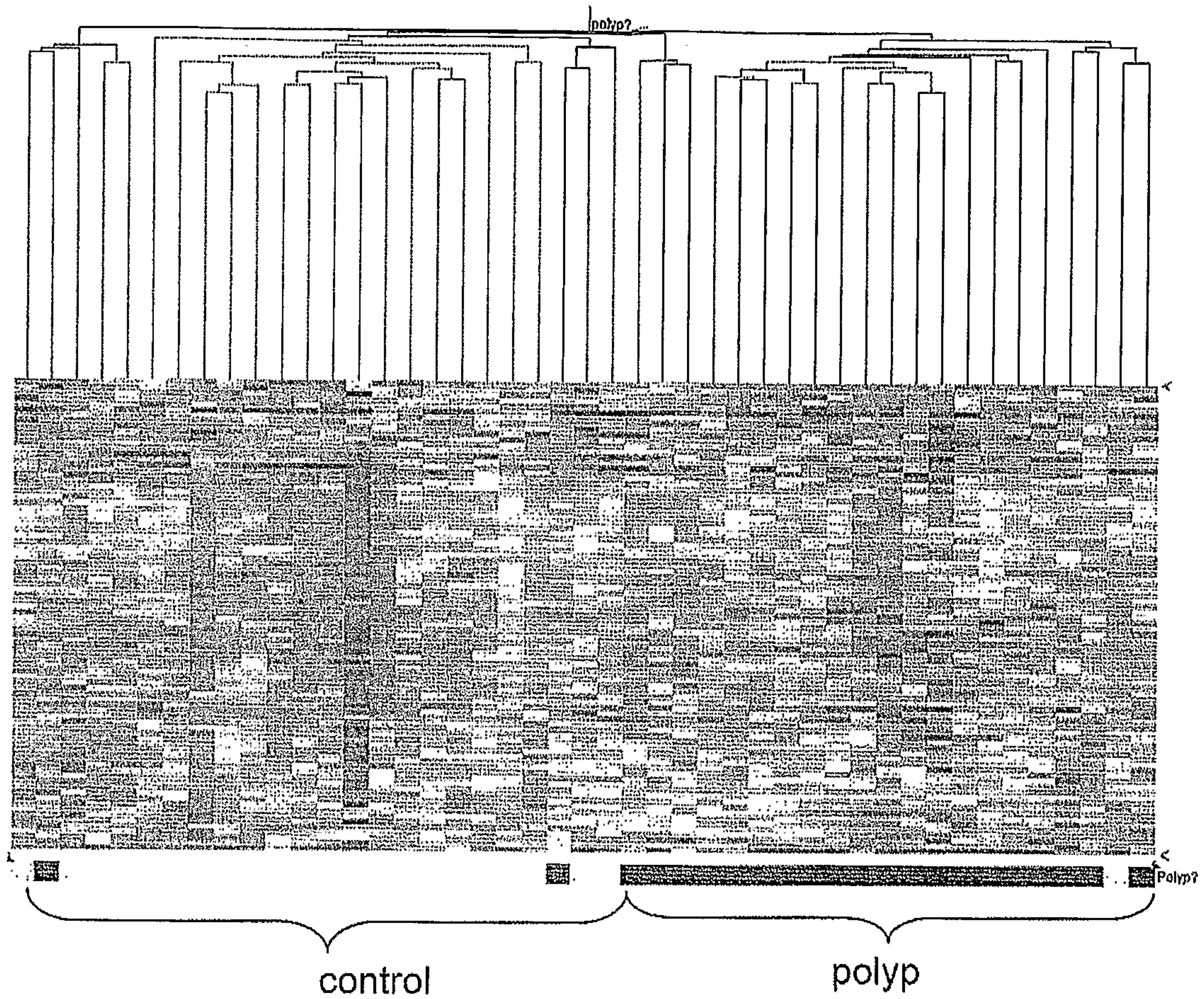


FIG. 1

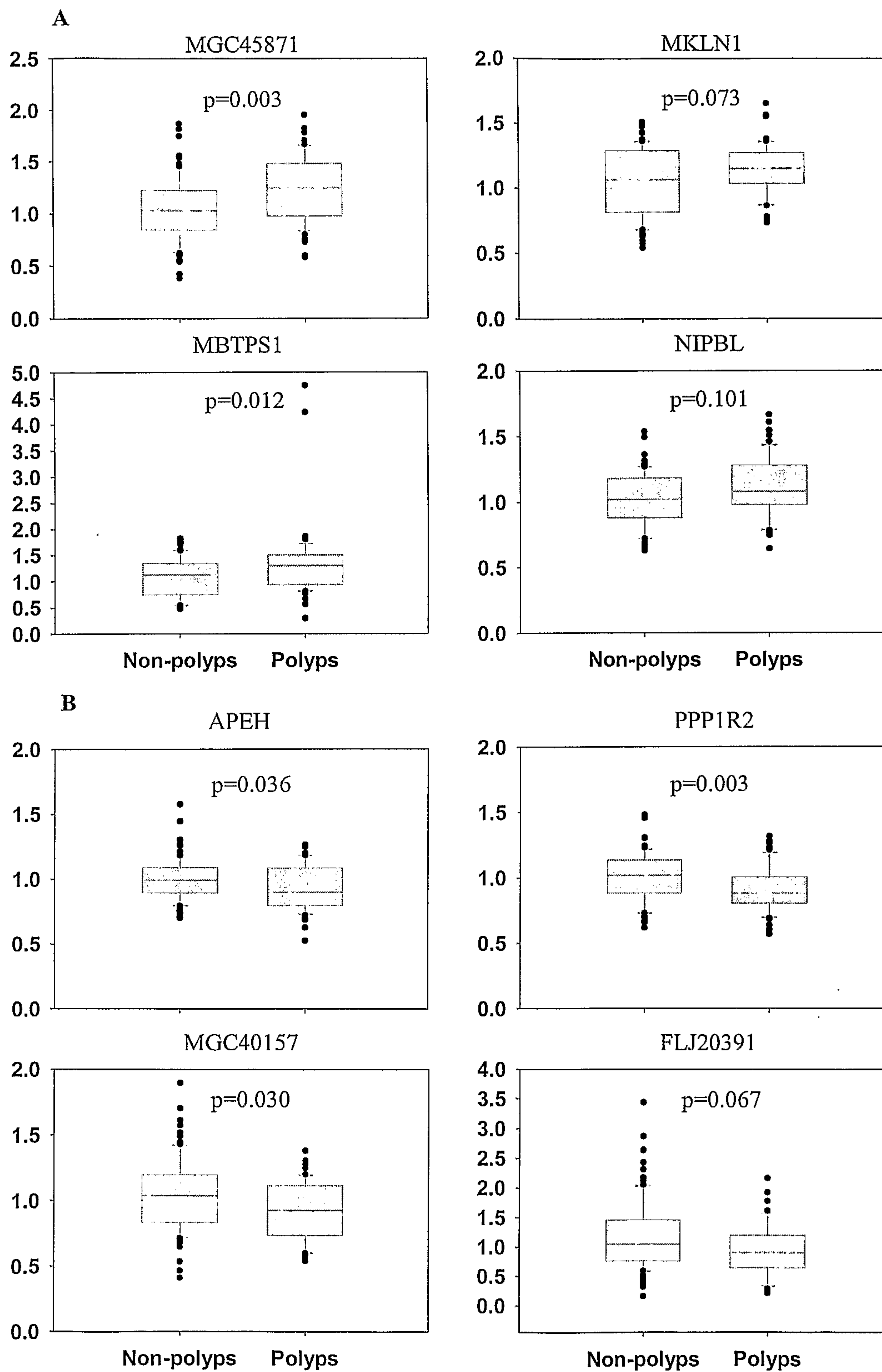
2/6

FIGURE 2



3/6

FIGURE 3



4/6

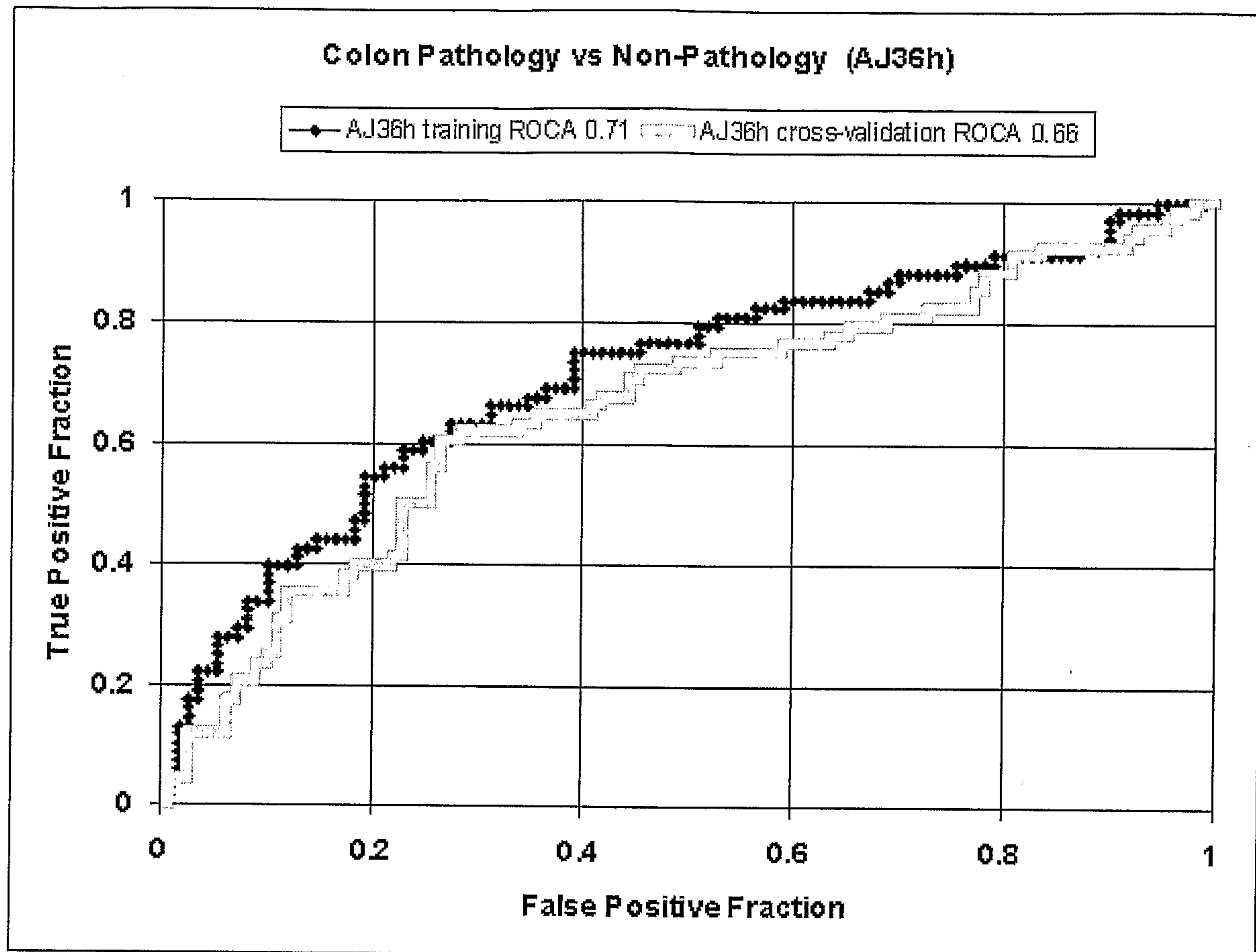
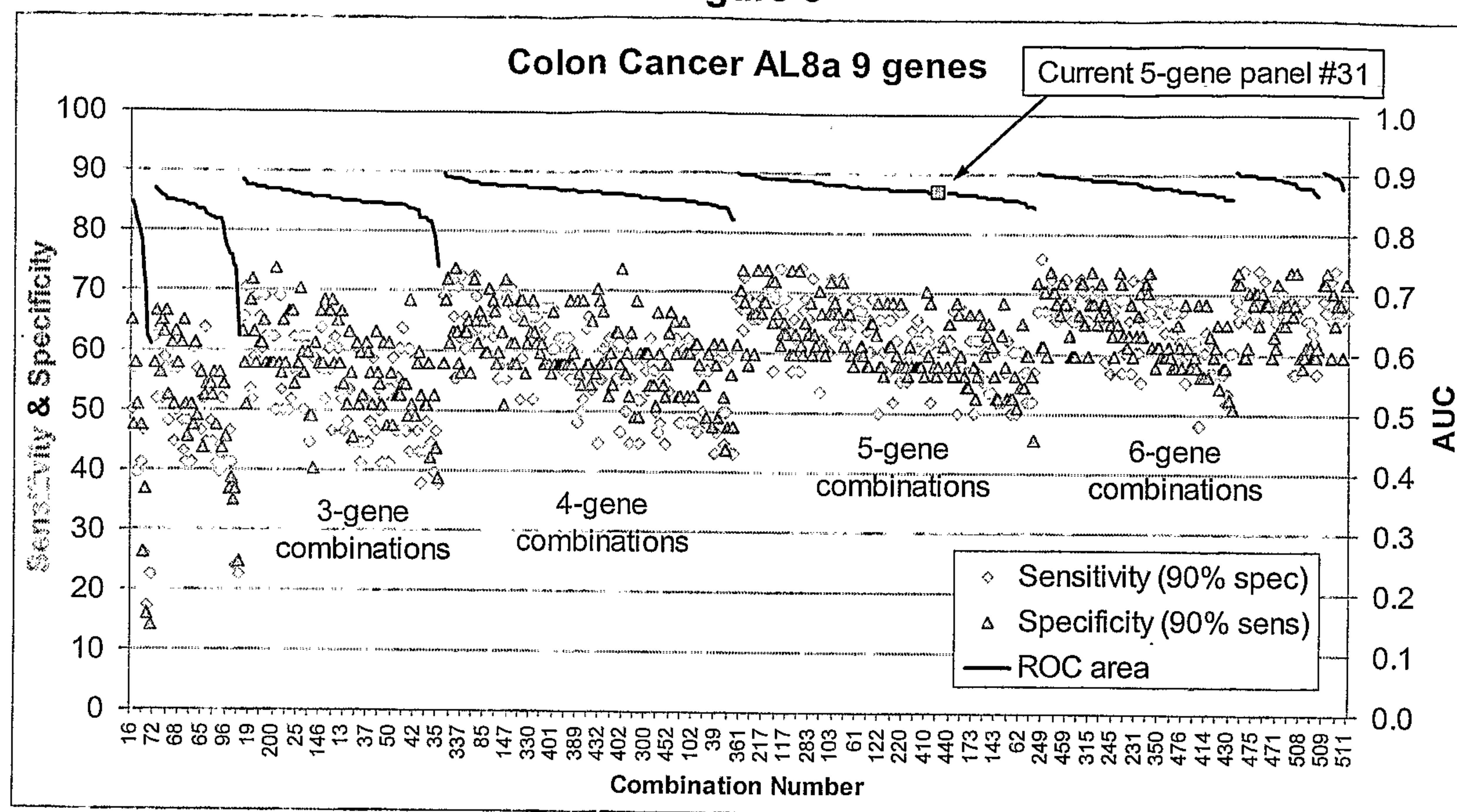
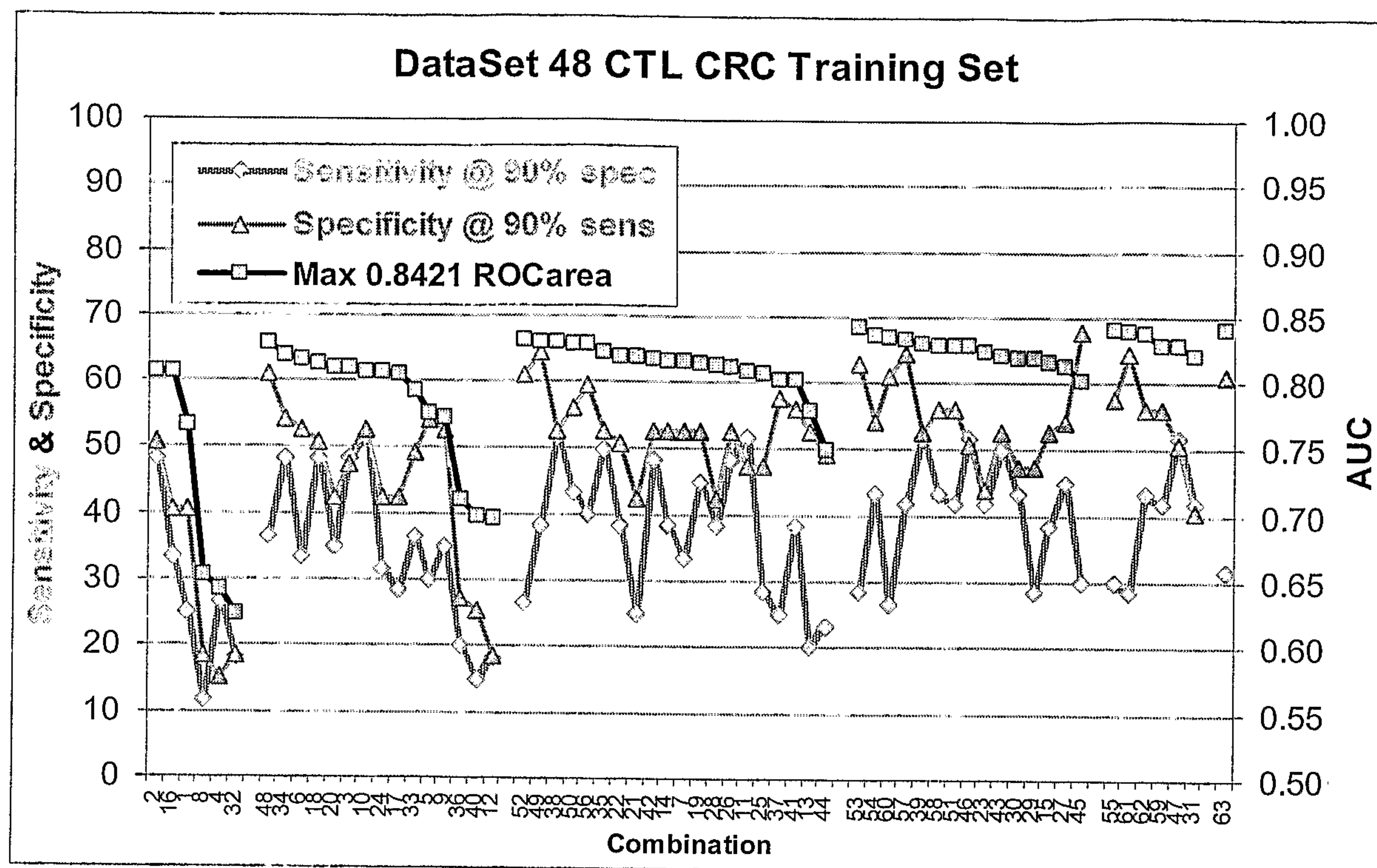
FIGURE 4

Figure 5



6/6

Figure 6



polyp?

