

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

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
25 September 2008 (25.09.2008)

PCT

(10) International Publication Number  
WO 2008/116178 A2

(51) International Patent Classification:

G01N 33/574 (2006.01)

(21) International Application Number:

PCT/US2008/057884

(22) International Filing Date: 21 March 2008 (21.03.2008)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/896,020	21 March 2007 (21.03.2007)	US
60/938,307	16 May 2007 (16.05.2007)	US
60/968,988	30 August 2007 (30.08.2007)	US
60/971,839	12 September 2007 (12.09.2007)	US

(71) Applicants (for all designated States except US): **VANDERBILT UNIVERSITY** [US/US]; 305 Kirkland Hall, Nashville, Tennessee 37240 (US). **H. LEE MOFFITT CANCER AND RESEARCH INSTITUTE, INC** [US/US]; 12902 Magnolia Drive, Tampa, Florida 33612-9497 (US). **THE CINCINNATI CHILDREN'S HOSPITAL RESEARCH FOUNDATION** [US/US]; 3333 Burnet Avenue, Cincinnati, Ohio 45229 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **COFFEY, Robert J.**

[US/US]; 1809 Laurel Ridge Drive, Nashville, Tennessee 37215 (US). **ARONOW, Bruce J.** [US/US]; 30 E. Central Parkway, Cincinnati, Ohio 45202 (US). **FRANKLIN, Jeffrey L.** [US/US]; 1923 18th Avenue South, Nashville, Tennessee 37212 (US). **YEATMAN, Timothy** [US/US]; 17341 Morris Bridge Road, Thonotosassa, Florida 33592 (US). **OLESEN, Sanne K. H.** [DK/US]; Apt. 2125, 14535 Bruce B Downs Blvd, Tampa, Florida 33613 (US).

(74) Agents: **MYERS JR., Richard S** et al.; Suite 800, 401 Commerce Street, Nashville, Tennessee 37219 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, 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),

[Continued on next page]

(54) Title: SYSTEMS AND METHODS FOR DIAGNOSIS AND PROGNOSIS OF COLORECTAL CANCER

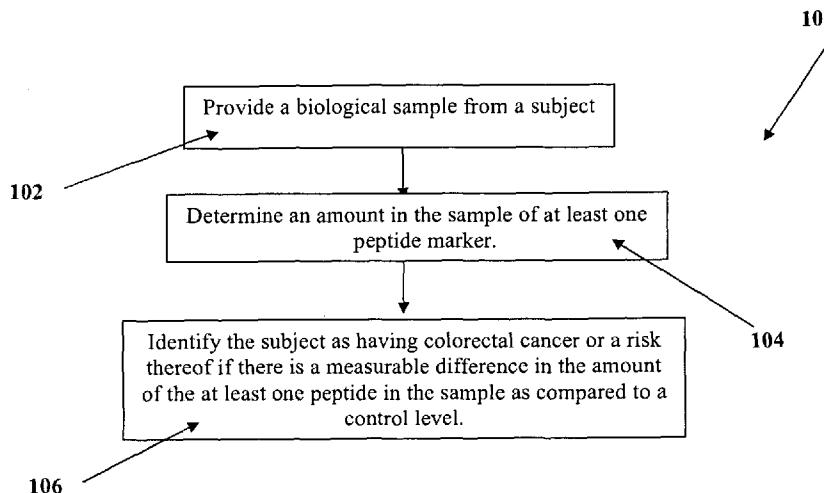


FIG. 1

WO 2008/116178 A2

(57) Abstract: Methods and systems are described for using in making a diagnosis of colorectal cancer in a subject. A method can include providing a biological sample from the subject; determining an amount in the sample of at least one peptide marker, selected from a group of markers as described herein; and comparing the amount of the at least one peptide marker in the sample, if present, to a control level of the at least one peptide marker, wherein the subject is diagnosed as having colorectal cancer or a risk thereof if there is a measurable difference in the amount of the at least one peptide marker in the sample as compared to the control level.



European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— *without international search report and to be republished upon receipt of that report*

SYSTEMS AND METHODS FOR DIAGNOSIS AND PROGNOSIS OF  
COLORECTAL CANCER

Attorney Docket No.: 11672N/7117WO(VC/VCM)

RELATED APPLICATIONS

**[0001]** This application claims priority from U.S. Provisional Application Serial Nos. 60/896,020 filed on March 21, 2007; 60/938,307 filed on May 16, 2007; 60/968,988 filed on August 30, 2007; and 60/971,839 filed on September 12, 2007; the entire disclosures of which are incorporated herein by this reference.

GOVERNMENT INTEREST

**[0002]** Subject matter described herein was made with U.S. Government support under Grant Numbers 5P50 CA095103-03, 2U01 CA084239-06, and UO1CA085052 awarded by the National Cancer Institute (NCI). The government has certain rights in the described subject matter.

TECHNICAL FIELD

**[0003]** The presently-disclosed subject matter relates to methods for diagnosis and prognosis of colorectal cancer. In particular, the presently-disclosed subject matter relates to diagnostic and prognostic methods based on determining an amount of protein markers in a biological sample from a subject.

INTRODUCTION AND GENERAL CONSIDERATIONS

**[0001]** Colorectal cancer is the second leading cause of cancer-related deaths in the western world, and accounts for about 655,000 deaths/year worldwide. Surveillance by

colonoscopy is currently considered to be the most effective screening method for colorectal cancer; however, the process is invasive, unpleasant, and time-consuming.

**[0002]** The identification of colorectal cancer biomarkers suitable for the early detection and diagnosis of colorectal cancer holds great promise for improving the clinical outcome of subjects, and would allow for more comprehensive population screening because it is less invasive, less expensive, and is less time consuming than methods currently in use. The identification of such biomarkers would be especially important for subjects presented with vague or no symptoms, as well as subjects who have tumors that are relatively inaccessible to physical examination. Despite considerable effort directed at early detection, few reliable and cost-effective screening tests have been developed.

**[0003]** In addition to improving early-stage cancer detection screening procedures, a set of markers that could, for an individual, quantitatively indicate the presence of residual or recurrent disease would be particularly useful for assessing effectiveness of treatment or a change in clinical status management.

**[0004]** Thus, there is an unmet need for biomarkers that individually, or in combination with other biomarkers or diagnostic modalities, could deliver the required sensitivity and specificity for early detection and prognosis of colorectal cancer, as well as quantitative detection following diagnosis and successive stages of therapeutic management. In particular, simple tests that can be performed on readily-accessible biological fluids are needed.

## SUMMARY

**[0005]** The presently-disclosed subject matter meets some or all of the above-identified needs, as will become evident to those of ordinary skill in the art after a study of information provided in this document.

**[0006]** This Summary describes several embodiments of the presently-disclosed subject matter, and in many cases lists variations and permutations of these embodiments. This Summary is merely exemplary of the numerous and varied embodiments. Mention of one or more representative features of a given embodiment is likewise exemplary. Such an embodiment can typically exist with or without the feature(s) mentioned; likewise, those features can be applied to other embodiments of the presently-disclosed subject matter, whether listed in this Summary or not. To avoid excessive repetition, this Summary does not list or suggest all possible combinations of such features.

**[0007]** The presently-disclosed subject matter includes methods and systems for making a diagnosis of colorectal cancer in a subject. Various peptides are disclosed herein for use as markers for colorectal cancer. Markers are identified in Tables A-O, set forth below.

**[0008]** In some embodiments, a method for making a diagnosis of a colorectal cancer in a subject includes: providing a biological sample from the subject; determining an amount in the sample of at least one peptide selected from the peptides set forth in Table A, Table B, and/or Table C; and comparing the amount of the at least one peptide in the sample, if present, to a control level of the at least one peptide, wherein the subject is diagnosed as having colorectal cancer or a risk thereof if there is a measurable difference in the amount of the at least one peptide in the sample as compared to the control level.

**[0009]** In some embodiments, the method includes determining an amount in the sample of at least two peptides, at least five peptides, or at least ten peptides selected from the peptides set forth in Table A, Table B, and/or Table C.

**[0010]** In some embodiments, the method includes determining an amount in the sample of at least one peptide selected from the peptides set forth in Table D, Table E, Table F, or Table G.

**[0011]** In some embodiments, the method includes determining an amount in the sample of at least one peptide selected from the peptides set forth in Table A, and at least one peptide selected from the peptides set forth in Table B. In some embodiments, the method includes determining an amount in the sample of at least one peptide selected from the peptides set forth in Table A, and at least one additional peptide selected from the peptides set forth in Table C. In some embodiments, the method includes determining an amount in the sample of at least one peptide selected from the peptides set forth in Table B, and at least one additional peptide selected from the peptides set forth in Table C.

**[0012]** In some embodiments, determining the amount in the sample of the at least one peptide is conducted using mass spectrometry (MS) analysis, immunoassay analysis, or both. In some embodiments where immunoassay analysis is conducted, an enzyme-linked immunosorbent assay (ELISA) can be used.

**[0013]** In some embodiments, the method also includes selecting a treatment or modifying a treatment for the colorectal cancer based on the determined amount of the at least one peptide.

**[0014]** In some embodiments, a method for making a diagnosis of a colorectal cancer in a subject includes: providing a biological sample from the subject; determining an amount in the sample of at least one peptide marker from at least one group of markers, selected from (i) the group of markers set forth in Tables H and/or I, which are associated with a first type of colorectal cancer; (ii) the group of markers set forth in Tables J and/or K, which are associated with a second

type of colorectal cancer; (iii) the group of markers set forth in Tables L and/or M, which are associated with a third type of colorectal cancer; and (iv) the group of markers set forth in Tables N and/or O, which are associated with a fourth type of colorectal cancer; and comparing the amount of the at least one peptide marker in the sample, if present, to a control level of the at least one peptide marker; wherein the subject is diagnosed as having the type of colorectal cancer associated with the group of markers of which the at least one peptide marker is a member if there is a measurable difference in the amount of the at least one peptide in the sample as compared to the control level.

**[0015]** In some embodiments, the method includes determining the amount in the sample of at least one peptide marker from at least two groups of markers. In some embodiments, the method includes determining the amount in the sample of at least one peptide marker from at least three groups of markers. In some embodiments, the method includes determining the amount in the sample of at least one peptide marker from four of the groups of markers.

**[0016]** In some embodiments, the biological sample includes blood, plasma, or serum. In some embodiments, the subject is human.

**[0017]** In some embodiments, a system for making a diagnosis of a colorectal cancer in a subject includes: probes for selectively binding each of one or more peptide markers, wherein the peptide markers are selected from the peptides set forth in Table A, Table B, and/or Table C; and means for detecting the binding of said probes to said one or more peptide markers. In some embodiments, the system includes at least one probe for selectively binding each of one or more peptide markers selected from the peptides set forth in Table A, and at least one probe for selectively binding each of one or more peptide markers selected from the peptides set forth in Table B. In some embodiments, the system includes at least one probe for selectively binding each of one or more peptide markers selected from the peptides set forth in Table B, and at least one probe for selectively binding each of one or more peptide markers selected from the peptides set forth in Table C. In some embodiments, the system includes at least one probe for selectively binding each of one or more peptide markers selected from the peptides set forth in Table A, and at least one probe for selectively binding each of one or more peptide markers selected from the peptides set forth in Table C.

**[0018]** In some embodiments, the system includes at least one probe for selectively binding each of one or more peptide markers selected from the peptides set forth in Table D, Table E, Table F, or Table G.

**[0019]** In some embodiments, a system for making a diagnosis of a colorectal cancer in a subject includes: probes for selectively binding each of one or more peptide markers, wherein the

peptide markers are selected from at least one group of markers, selected from (i) the group of markers set forth in Tables H and/or I, which are associated with a first type of colorectal cancer; (ii) the group of markers set forth in Tables J and/or K, which are associated with a second type of colorectal cancer; (iii) the group of markers set forth in Tables L and/or M, which are associated with a third type of colorectal cancer; and the group of markers set forth in Tables N and/or O, which are associated with a fourth type of colorectal cancer; and means for detecting the binding of said probes to said one or more peptide markers.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0004] **FIG. 1.** Flow chart illustrating the steps involved in an embodiment of a method of the presently-disclosed subject matter.

[0005] **FIG. 2.** depicts a stratification of murine colon tumor models by localization of  $\beta$ -catenin and a plan for analysis.

[0006] **FIG. 3.** is a graph depicting the results of a study to determine the differential expression of transcripts identified by microarray analyses using quantitative real-time PCR (qRT-PCR).

[0007] **FIG. 4.** includes an analysis of genes over-expressed and under-expressed in embryonic colon and in tumors.

[0008] **FIG. 5.** includes an analysis of gene lists with criteria of over-expression or under-expression in development, or over-expression or under expression in human CRCs.

[0009] **FIG. 6A** is a schematic diagram of the canonical WNT signaling pathway showing elements present in C6 (gene symbols with gray background).

[0010] **FIG. 6B** depicts relative gene expression for MYC and SOX4 for individual murine and human tumors.

[0011] **FIG. 7A** is a diagram showing that tumors exhibit large-scale activation of developmental patterns, where nuclear  $\beta$ -catenin-positive ( $Apc^{Min/+}$  and AOM) tumors map more strongly to early development stages during (more proliferative, less differentiated), whereas nuclear  $\beta$ -catenin-negative ( $Tgfb1^{-/-}$ ;  $Rag2^{-/-}$  and  $Smad3^{-/-}$ ) tumors map more strongly to later stages consistent with increased epithelial differentiation.

[0012] **FIG. 7B** illustrates an overall representation of the relationship of mouse colon tumor models and human CRC to development and non-developmental expression patterns.

## DESCRIPTION OF EXEMPLARY EMBODIMENTS

**[0020]** The details of one or more embodiments of the presently-disclosed subject matter are set forth in this document. Modifications to embodiments described in this document, and other embodiments, will be evident to those of ordinary skill in the art after a study of the information provided. The information provided in this document, and particularly the specific details of the described exemplary embodiments, is provided primarily for clearness of understanding and no unnecessary limitations are to be understood therefrom. In case of conflict, the specification of this document, including definitions, will control.

**[0021]** While the terms defined herein are believed to be well understood by one of ordinary skill in the art, the definitions are set forth to facilitate explanation of the presently-disclosed subject matter.

**[0022]** Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the presently-disclosed subject matter belongs. Although any methods, systems, and materials similar or equivalent to those described herein can be used in the practice or testing of the presently-disclosed subject matter, representative methods, devices, and materials are now described.

**[0023]** Following long-standing patent law convention, the terms “a”, “an”, and “the” refer to “one or more” when used in this application, including the claims. Thus, for example, reference to “a cell” includes a plurality of such cells, and so forth.

**[0024]** Unless otherwise indicated, all numbers expressing quantities of ingredients, properties such as reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term “about”. Accordingly, unless indicated to the contrary, the numerical parameters set forth in this specification and claims are approximations that can vary depending upon the desired properties sought to be obtained by the presently-disclosed subject matter.

**[0025]** As used herein, the term “about,” when referring to a value or to an amount of mass, weight, time, volume, concentration or percentage is meant to encompass variations of in some embodiments  $\pm 20\%$ , in some embodiments  $\pm 10\%$ , in some embodiments  $\pm 5\%$ , in some embodiments  $\pm 1\%$ , in some embodiments  $\pm 0.5\%$ , and in some embodiments  $\pm 0.1\%$  from the specified amount, as such variations are appropriate to perform the disclosed method.

**[0026]** The presently-disclosed subject matter includes methods and systems for diagnosis and prognosis of colorectal cancer. Colorectal cancer can include, for example, cancer of the colon, rectum, and/or appendix. The presently-disclosed subject matter includes methods and systems

for diagnosing colorectal cancer in a subject, and for determining whether to initiate or continue treatment of colorectal cancer in a subject. In some embodiments, the method includes identifying at least one marker in a biological sample from a subject. The at least one marker can be a secreted protein. In some embodiments, the at least one marker is a polypeptide selected from the following, or a polypeptide fragment thereof:

**Table A**

Abbreviation	Peptide Name
AZGP1	Alpha-2-glycoprotein 1, zinc
BGN	Biglycan
CHI3L1	Chitinase 3-like 1 (cartilage glycoprotein-39)
COL10A1	Collagen, type X, alpha 1(
COL11A1	Collagen, type XI, alpha 1
COL12A1	Collagen, type XII, alpha 1
COL1A1	Collagen, type I, alpha 1
COL1A2	Collagen, type I, alpha 2
COL5A2	Collagen, type V, alpha 2
COMP	Cartilage oligomeric matrix protein
CXCL1	Chemokine (C-X-C motif) ligand 1
CXCL2	Chemokine (C-X-C motif) ligand 2
CXCL3	Chemokine (C-X-C motif) ligand 3
ENC1	Ectodermal-neural cortex
FN1	Fibronectin 1
GDF15	Growth differentiation factor 15
IL8	Interleukin 8
INHBA	Inhibin beta A
MMP12	Matrix metalloproteinase 12
MMP7	Matrix metalloproteinase 7
PTGS2	COX-2
SFRP4	Secreted frizzled-related protein 4
SPP1	Osteopontin
TGFBI	Transforming growth factor beta 1
THBS2	Thrombospondin 2

**[0027]** In some embodiments, the at least one marker is a polypeptide selected from the following, or a polypeptide fragment thereof:

**Table B**

Abbreviation	Peptide Name
ANGPT2	Angiopoietin 2
BRCA2	Breast cancer 2, early onset
COL4A1	Collagen, type IV, alpha 1
COL8A1	Collagen, type VIII, alpha 1
CSPG6	Chondroitin sulfate proteoglycan 6
CTHRC1	Collagen triple helix repeat
CXCL5	Chemokine (C-X-C motif) ligand 5
CXCL6	Chemokine (C-X-C motif) ligand 6
HAMP	Hepcidin antimicrobial peptide
IF	I factor (complement)
INSIG1	Insulin induced gene 1
KIAA0963	Strawberry notch homolog 2
KRT16	Keratin 16
LYZ	Lysozyme (renal amyloidosis)
MFAP2	Microfibrillar-associated glycoprotein 2

**Table B**

Abbreviation	Peptide Name
MMP11	Matrix metalloproteinase 11
MMP9	Matrix metalloproteinase 9
NLF2	Nuclear localized factor 2
S100A9	S100 calcium binding protein A9 (calgranulin B)
SERPINE2	Serpin peptidase inhibitor
SPARC	Osteonectin
SULF1	Sulfatase 1
TDGF1	Cripto-1
TGIF2	TGF beta-induced factor 2
TIMP1	Tissue inhibitor of metalloproteinase 1
WISP1	WNT1 inducible signaling peptide

**[0028]** In some embodiments, the at least one marker is a polypeptide selected from the following, or a polypeptide fragment thereof:

**Table C**

Abbreviation	Peptide Name
ABCB6	ATP-binding cassette, sub-family B (MDR/TAP), member 6
ACSL6	Acyl-CoA synthetase long-chain family member 6
ADAM12	ADAM metallopeptidase domain 12 (meltrin alpha)
ADAM22	Adam metallopeptidase domain 22
ANLN	Anillin, actin binding protein (scraps homolog, <i>Drosophila</i> )
AREG / LOC653193	amphiregulin (schwannoma-derived growth factor) / similar to Amphiregulin precursor (AR) (Colorectum cell-derived growth factor) (CRDGF)
ARNTL2	Aryl hydrocarbon receptor nuclear translocator-like 2
BAMBI	BMP and activin membrane-bound inhibitor homolog ( <i>Xenopus laevis</i> )
BAX	BCL2-associated X protein
BCL2L1	BCL2-like 1
BGN	Biglycan
C17orf69	Chromosome 17 open reading frame 69
C1orf34	Chromosome 1 open reading frame 34
C2	Complement component 2
CCL18	Chemokine (C-C motif) ligand 18 (pulmonary and activation-regulated)
CCL20	Chemokine (C-C motif) ligand 20
CCL3	Chemokine (C-C motif) ligand 3
CCL3L1	Chemokine (C-C motif) ligand 3 – like 1
CCL3L3	Chemokine (C-C motif) ligand 3 – like 3
CDH11	Cadherin 11, type 2, OB-cadherin (osteoblast)
CDH3	Cadherin 3, type 1, P-cadherin (placental)
CEL	Carboxyl ester lipase (bile salt-stimulated lipase)
CFB	Complement factor B
CLDN1	Claudin 1
CLDN2	Claudin 2
COL5A2	Collagen, type V, alpha 2
COMP	Cartilage oligomeric matrix protein
CSPG2	Chondroitin sulfate proteoglycan 2 (versican)
CST1	Cystatin SN
CXCL11	Chemokine (C-X-C motif) ligand 11
DKK2	Dickkopf homolog 2 ( <i>Xenopus laevis</i> )
DNAJC3	DnaJ (Hsp40) homolog, subfamily C, member 3
DPEP1	dipeptidase 1 (renal)
DUOX2	Dual Oxidase 2
DUSP4	dual specificity phosphatase 4
EREG	Epiregulin

**Table C**

Abbreviation	Peptide Name
ESM1	Endothelial cell-specific molecule 1
ETV4	ets variant gene 4 (E1A enhancer binding protein, E1AF)
FAP	fibroblast activation protein, alpha
FN1	Fibronectin 1
GDF15	Growth differentiation factor 15
GTF2IRD1	GTF2I domain containing 1
HEAT R3	HEAT repeat containing 3
HIG2	Hypoxia-inducible protein 2
ITGBL1	Integrin, beta-like 1 (with EGF-like repeat domains)
JUB	jub, ajuba homolog (Xenopus laevis)
KDELC1	KDEL (Lys-Asp-Glu-Leu) containing 1
KLK10	kallikrein 10
LCN2	Lipocalin 2 (oncogene 24p3)
LOC643930	Similar to Small inducible cytokine A3-like 1 precursor (Tonsillar lymphocyte LD78 beta protein) (LD78-beta (1-70)) (G0/G1 switch regulatory protein 19-2)(G0S19-2 protein) PAT 464.2)
LPGAT1	lysophosphatidylglycerol acyltransferase 1
LRP8	low density lipoprotein receptor-related protein 8, apolipoprotein e receptor
LRRC14	Leucine rich repeat containing 14
LY6G6D	lymphocyte antigen 6 complex, locus G6D
MGC50722	Hypothetical MGC50722
MMP1	Matrix metalloproteinase 1
MMP3	Matrix metalloproteinase 3
MSLN	Mesothelin
MTHFD1L	methylenetetrahydrofolate dehydrogenase (NADP + dependent) 1-like
NEBL	Nebulette
NFAT5	nuclear factor of activated T-cells 5, tonicity-responsive
NFE2L3	Nuclear factor (erythroid-derived 2)-like 3
NUP1	Nucleoporin like 1
OLR1	Oxidised low density lipoprotein (lectin-like) receptor 1
PHF20	PHD finger protein 20
PPM1H	Protein phosphatase 1H (PP2C domain containing)
PTPLAD1	protein tyrosine phosphatase-like A domain containing 1
QPCT	glutaminyl-peptide cyclotransferase (glutaminyl cyclase)
REG3A	Regenerating islet-derived 3 alpha
RNF43	Ring finger protein 43
SCD	Stearoyl-CoA desaturase (Delta-9-desaturase)
SERPINB5	Serpin peptidase inhibitor, clade B (ovalbumin), member 5
SIX3	Sine oculis homeobox homolog 3 (Drosophila)
SLC6A14	Solute carrier family 6 (amino acid transporter), member 14
SLC7A11	Solute carrier family 7, (cationic amino acid transporter, y <sup>+</sup> system) member 11
SLCO1B3	solute carrier organic anion transporter family, member 1B3
SLCO4A1	solute carrier organic anion transporter family, memer 4A1
SMG1	PI-3-kinase-related kinase SMG-1
SPATA21	Spermatogenesis associated 21
SRPX2	Sushi-repeat-containing protein, X-linked 2
TAC4	Tachykinin 4 (hemokinin)
TACSTD2	Tumor-associated calcium signal transducer 2
TMEM44	Transmembrane protein 44
TNFSF11	Tumor necrosis factor (ligand) superfamily, member 11
TNFRSF12A	Tumor necrosis factor receptor superfamily, member 12A secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)
TRIB3	Tribbles homolog 3 (Drosophila)
TSPAN5	tetraspanin 5

[0029] In some embodiments, the at least one marker is a polypeptide selected from the following, or a polypeptide fragment thereof:

Abbreviation	Peptide Name
ACSL6	acyl-CoA synthetase long-chain family member 6
ADAM12	ADAM metallopeptidase domain 12 (meltrin alpha)
AREG / LOC653193	amphiregulin (schwannoma-derived growth factor) / similar to Amphiregulin precursor (AR) (Colorectum cell-derived growth factor) (CRDGF)
BAMBI	BMP and activin membrane-bound inhibitor homolog (Xenopus laevis)
BAMBI	BMP and activin membrane-bound inhibitor homolog (Xenopus laevis)
BAX	BCL2-associated X protein
BGN	biglycan
C17orf69	chromosome 17 open reading frame 69
C2	complement component 2
CCL20	chemokine (C-C motif) ligand 20
CCL3 / CCL3L1 / CCL3L3 / LOC643930	chemokine (C-C motif) ligand 3 / chemokine (C-C motif) ligand 3-like 1 / chemokine (C-C motif) ligand 3-like 3 / similar to Small inducible cytokine A3-like 1 precursor (Tonsillar lymphocyte LD78 beta protein) (LD78-beta(1-70)) (G0/G1 switch regulatory protein 19-2) (G0S19-2 protein) (PAT 464.2)
CDH11	cadherin 11, type 2, OB-cadherin (osteoblast)
CFB	complement factor B
COL10A1	collagen, type X, alpha 1 (Schmid metaphyseal chondrodysplasia)
CSPG2	chondroitin sulfate proteoglycan 2 (versican)
CXCL11	chemokine (C-X-C motif) ligand 11
DKK2	dickkopf homolog 2 (Xenopus laevis)
DUOX2	dual oxidase 2
DUOX2	dual oxidase 2
DUSP4	dual specificity phosphatase 4
ETV4	ets variant gene 4 (E1A enhancer binding protein, E1AF)
ETV4	ets variant gene 4 (E1A enhancer binding protein, E1AF)
GTF2IRD1	GTF2I repeat domain containing 1
HEATR3	HEAT repeat containing 3
HIG2	hypoxia-inducible protein 2
KDELC1	KDEL (Lys-Asp-Glu-Leu) containing 1
LCN2	lipocalin 2 (oncogene 24p3)
MMP12	matrix metallopeptidase 12 (macrophage elastase)
MTHFD1L	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like
NEBL	nebulette
PTPLAD1	protein tyrosine phosphatase-like A domain containing 1
QPCT	glutaminyl-peptide cyclotransferase (glutaminyl cyclase)
REG3A	regenerating islet-derived 3 alpha
REG3A	regenerating islet-derived 3 alpha
RNF43	ring finger protein 43
SERPINB5	serpin peptidase inhibitor, clade B (ovalbumin), member 5
SLC7A11	solute carrier family 7, (cationic amino acid transporter, y <sup>+</sup> system) member 11
SLC7A11	solute carrier family 7, (cationic amino acid transporter, y <sup>+</sup> system) member 11

**Table D**

Abbreviation	Peptide Name
SLCO1B3	solute carrier organic anion transporter family, member 1B3
SLCO4A1	solute carrier organic anion transporter family, member 4A1
SMG1	PI-3-kinase-related kinase SMG-1
SRPX2	sushi-repeat-containing protein, X-linked 2
TAC4	tachykinin 4 (hemokinin)
TRIB3	tribbles homolog 3 (Drosophila)
TSPAN5	tetraspanin 5 /// tetraspanin 5

[0030] The markers set forth in Table D are each associated with one of four Dukes stages.

[0031] In some embodiments, the at least one marker is a polypeptide selected from the following, or a polypeptide fragment thereof:

**Table E**

Abbreviation	Peptide Name
ARNTL2	aryl hydrocarbon receptor nuclear translocator-like 2
CCL18	chemokine (C-C motif) ligand 18 (pulmonary and activation-regulated)
CEL	carboxyl ester lipase (bile salt-stimulated lipase)
CEL	carboxyl ester lipase (bile salt-stimulated lipase)
COL12A1	collagen, type XII, alpha 1
COL5A2	collagen, type V, alpha 2
CXCL11	chemokine (C-X-C motif) ligand 11
CXCL11	chemokine (C-X-C motif) ligand 11
DNAJC3	DnaJ (Hsp40) homolog, subfamily C, member 3
FN1	fibronectin 1 /// fibronectin 1
HEATR3	HEAT repeat containing 3
LCN2	lipocalin 2 (oncogene 24p3)
MMP12	matrix metallopeptidase 12 (macrophage elastase)
MSLN	mesothelin
MTHFD1L	methylenetetrahydrofolate dehydrogenase (NADP <sup>+</sup> dependent) 1-like
NEBL	nebulette
NFAT5	nuclear factor of activated T-cells 5, tonicity-responsive
PHF20	PHD finger protein 20
PPM1H	protein phosphatase 1H (PP2C domain containing)
PPM1H	protein phosphatase 1H (PP2C domain containing)
PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)
SFRP4	secreted frizzled-related protein 4
SLCO1B3	solute carrier organic anion transporter family, member 1B3
SRPX2	sushi-repeat-containing protein, X-linked 2
TGFBI	transforming growth factor, beta-induced, 68kDa
TRIB3	tribbles homolog 3 (Drosophila)

[0032] The markers set forth in Table E are each associated with two of four Dukes stages.

[0033] In some embodiments, the at least one marker is a polypeptide selected from the following, or a polypeptide fragment thereof:

**Table F**

Abbreviation	Peptide Name
ABCB6	ATP-binding cassette, sub-family B (MDR/TAP), member 6
ANLN	anillin, actin binding protein (scraps homolog, <i>Drosophila</i> )
CEL	carboxyl ester lipase (bile salt-stimulated lipase)
CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)
COL11A1	collagen, type XI, alpha 1
COL12A1	collagen, type XII, alpha 1
COL1A1	collagen, type I, alpha 1
COL1A2	collagen, type I, alpha 2
COMP	cartilage oligomeric matrix protein
CST1	cystatin SN
EREG	epiregulin
ESM1	endothelial cell-specific molecule 1
FAP	fibroblast activation protein, alpha
ITGBL1	integrin, beta-like 1 (with EGF-like repeat domains)
KLK10	kallikrein 10
LPGAT1	lysophosphatidylglycerol acyltransferase 1
NUPL1	nucleoporin like 1
OLR1	oxidised low density lipoprotein (lectin-like) receptor 1
SCD	stearoyl-CoA desaturase (delta-9-desaturase)
SERPINB5	serpin peptidase inhibitor, clade B (ovalbumin), member 5
SLC6A14	solute carrier family 6 (amino acid transporter), member 14
SPP1	secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)
TACSTD2	tumor-associated calcium signal transducer 2
THBS2	thrombospondin 2
TMEM44	transmembrane protein 44
TNFRSF12A	tumor necrosis factor receptor superfamily, member 12A
TNFSF11	tumor necrosis factor (ligand) superfamily, member 11

[0034] The markers set forth in Table F are each associated with three of four Dukes stages.

[0035] In some embodiments, the at least one marker is a polypeptide selected from the following, or a polypeptide fragment thereof:

**Table G**

Abbreviation	Peptide Name
ADAM22	ADAM metallopeptidase domain 22
AZGP1	alpha-2-glycoprotein 1, zinc
AZGP1	alpha-2-glycoprotein 1, zinc
BCL2L1	BCL2-like 1
C1orf34	chromosome 1 open reading frame 34
CCL20	chemokine (C-C motif) ligand 20
CDH3	cadherin 3, type 1, P-cadherin (placental)
CDH3	cadherin 3, type 1, P-cadherin (placental)
CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)
CLDN1	claudin 1
CLDN1	claudin 1
CLDN2	claudin 2
CLDN2	claudin 2

**Table G**

<b>Abbreviation</b>	<b>Peptide Name</b>
COL10A1	collagen, type X, alpha 1(Schmid metaphyseal chondrodysplasia)
COL11A1	collagen, type XI, alpha 1
COL11A1	collagen, type XI, alpha 1
COL12A1	collagen, type XII, alpha 1
COL1A2	collagen, type I, alpha 2
CST1	cystatin SN
CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)
CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)
CXCL11	chemokine (C-X-C motif) ligand 11
CXCL2	chemokine (C-X-C motif) ligand 2
CXCL2	chemokine (C-X-C motif) ligand 2
CXCL3	chemokine (C-X-C motif) ligand 3
CXCL3	chemokine (C-X-C motif) ligand 3
DPEP1	dipeptidase 1 (renal)
DPEP1	dipeptidase 1 (renal)
DUSP4	dual specificity phosphatase 4
DUSP4	dual specificity phosphatase 4
ENC1	ectodermal-neural cortex (with BTB-like domain)
ENC1	ectodermal-neural cortex (with BTB-like domain)
EREG	epiregulin
GDF15	growth differentiation factor 15
GDF15	growth differentiation factor 15
HIG2	hypoxia-inducible protein 2
IL8	interleukin 8
INHBA	inhibin, beta A (activin A, activin AB alpha polypeptide)
INHBA	inhibin, beta A (activin A, activin AB alpha polypeptide)
JUB	jub, ajuba homolog (Xenopus laevis)
LRP8	low density lipoprotein receptor-related protein 8, apolipoprotein e receptor
LRRC14	leucine rich repeat containing 14
LY6G6D	lymphocyte antigen 6 complex, locus G6D
LY6G6D	lymphocyte antigen 6 complex, locus G6D
MGC50722	hypothetical MGC50722
MMP1	matrix metallopeptidase 1 (interstitial collagenase)
MMP1	matrix metallopeptidase 1 (interstitial collagenase)
MMP3	matrix metallopeptidase 3 (stromelysin 1, progelatinase)
MMP3	matrix metallopeptidase 3 (stromelysin 1, progelatinase)
MMP7	matrix metallopeptidase 7 (matrilysin, uterine)
MMP7	matrix metallopeptidase 7 (matrilysin, uterine)
NFE2L3	nuclear factor (erythroid-derived 2)-like 3
NFE2L3	nuclear factor (erythroid-derived 2)-like 3
NUPL1	nucleoporin like 1
SCD	stearoyl-CoA desaturase (delta-9-desaturase)
SIX3	sine oculis homeobox homolog 3 (Drosophila)
SLC6A14	solute carrier family 6 (amino acid transporter), member 14
SPATA21	spermatogenesis associated 21
SPP1	secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-

Table G	
Abbreviation	Peptide Name
	lymphocyte activation 1)
TACSTD2	tumor-associated calcium signal transducer 2
TGFBI	transforming growth factor, beta-induced, 68kDa

**[0036]** The markers set forth in Table G are each associated with four of four Dukes stages. Stages of colorectal cancer can be defined according to different systems, as will be understood by those of ordinary skill in the art. For example, stages can be defined using the Dukes system, which stages cancer as Dukes A, B, C, or D. Dukes A can refer to a cancer affecting the innermost lining of the colon or rectum. Dukes B can refer to a cancer that has spread into the muscle layer of the colon or rectum. Dukes C can refer to a cancer that has spread to a lymph node in the region of the colon or rectum. Dukes D can refer to a primary colorectal cancer that has metastasized to a location elsewhere in the body.

**[0037]** The terms “polypeptide”, “protein”, and “peptide”, which are used interchangeably herein, refer to a polymer of the 20 protein amino acids, including modified amino acids (e.g., phosphorylated, glycated, etc.) and amino acid analogs, regardless of size or function. Thus, exemplary polypeptides include gene products, naturally occurring proteins, homologs, orthologs, paralogs, fragments and other equivalents, variants, and analogs of the foregoing.

**[0038]** The term “fragment”, when used with reference to a polypeptide, refers to a polypeptide in which amino acid residues are absent as compared to the full-length polypeptide itself, but where the remaining amino acid sequence is usually identical to the corresponding positions in the reference polypeptide. Such deletions can occur at the amino-terminus or carboxy-terminus of the reference polypeptide, or alternatively both. A fragment can retain one or more of the biological features of the reference polypeptide. When the term “peptide” is used herein, it is intended to include the full-length peptide as well as fragments of the peptide. Thus, an identified fragment of a peptide (e.g., by mass spectrometry) is intended to encompass the fragment as well as the full-length peptide.

**[0039]** In some embodiments of the presently-disclosed subject matter, a method for diagnosing colorectal cancer in a subject is provided. The terms “diagnosing” and “diagnosis” as used herein refer to methods by which one of ordinary skill in the art can estimate and/or even determine whether or not a subject has a colorectal cancer or a risk thereof. One of ordinary skill in the art can make a diagnosis on the basis of one or more diagnostic indicators, such as for example a marker, the amount (including presence or absence) of which is indicative of the presence, severity, or absence of the condition.

[0040] With reference to **FIG. 1**, in some embodiments, a method for diagnosing colorectal cancer in a subject **100** comprises providing a biological sample from the subject **102**; determining an amount of at least one peptide marker **104**; and identifying the subject as having colorectal cancer or a risk thereof if there is a measurable difference in the amount of the at least one peptide in the sample as compared to a control level **106**.

[0041] With regard to providing a biological sample from the subject **102**, the term “biological sample” as used herein refers to any body fluid or tissue potentially comprising secreted peptide markers, such as the peptides set forth in Tables A, B, and C. In some embodiments, for example, the biological sample can be a blood sample, a serum sample, a plasma sample, or sub-fractions thereof. In some embodiments, the biological sample can be derived from a normal stool specimen or from a stool specimen obtained following a treatment with a GI tract prokinetic and or secretory stimulant.

[0042] Turning now to identifying one or more markers in the biological sample **104**, various methods known to those of ordinary skill in the art can be used to identify the one or more markers in the provided biological sample. For example, mass spectrometry and/or immunoassay devices and methods can be used, although other methods are well known to those of ordinary skill in the art (for example, the measurement of marker RNA levels). See, e.g., U.S. Pat. Nos. 6,143,576; 6,113,855; 6,019,944; 5,985,579; 5,947,124; 5,939,272; 5,922,615; 5,885,527; 5,851,776; 5,824,799; 5,679,526; 5,525,524; and 5,480,792, each of which is hereby incorporated by reference in its entirety. Immunoassay devices and methods can utilize labeled molecules in various sandwich, competitive, or non-competitive assay formats, to generate a signal that is related to the presence or amount of an analyte of interest. Additionally, certain methods and devices, such as biosensors and optical immunoassays, can be employed to determine the presence or amount of analytes without the need for a labeled molecule. See, e.g., U.S. Pat. Nos. 5,631,171; and 5,955,377, each of which is hereby incorporated by reference in its entirety.

[0043] Thus, in some embodiments of the presently-disclosed subject matter, the marker peptides are analyzed using an immunoassay. The presence or amount of a marker can be determined using antibodies or fragments thereof specific for each marker, and detecting specific binding. For example, in some embodiments, the antibody specifically binds SPARC, which is inclusive of antibodies that bind the full-length peptide or a fragment thereof. In some embodiments, antibodies are provided, wherein each antibody specifically binds a full-length peptide or a fragment thereof that is selected from: a peptide of Tables A, B, and C. In some embodiments, antibodies are provided, wherein each antibody specifically binds a particular

isotype of a peptide or a fragment thereof. In some embodiments, the antibody or antibodies can be monoclonal.

**[0044]** Any suitable immunoassay can be utilized, for example, enzyme-linked immunoassays (ELISA), radioimmunoassays (RIAs), competitive binding assays, and the like. Specific immunological binding of the antibody to the marker can be detected directly or indirectly. Direct labels include fluorescent or luminescent tags, metals, dyes, radionuclides, and the like, attached to the antibody. Indirect labels include various enzymes well known in the art, such as alkaline phosphatase, horseradish peroxidase and the like.

**[0045]** The use of immobilized antibodies or fragments thereof specific for the markers is also contemplated by the presently-disclosed subject matter. The antibodies can be immobilized onto a variety of solid supports, such as magnetic or chromatographic matrix particles, the surface of an assay plate (such as microtiter wells), pieces of a solid substrate material (such as plastic, nylon, paper), and the like. An assay strip can be prepared by coating the antibody or a plurality of antibodies in an array on a solid support. This strip can then be dipped into the test biological sample and then processed quickly through washes and detection steps to generate a measurable signal, such as, for example, a colored spot.

**[0046]** In some embodiments, mass spectrometry (MS) analysis can be used alone or in combination with other methods (e.g., immunoassays) to determine the presence and/or quantity of the one or more markers of interest in a biological sample. In some embodiments, the MS analysis comprises matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) MS analysis, such as for example direct-spot MALDI-TOF or liquid chromatography MALDI-TOF mass spectrometry analysis. In some embodiments, the MS analysis comprises electrospray ionization (ESI) MS, such as for example liquid chromatography (LC) ESI-MS. Mass analysis can be accomplished using commercially-available spectrometers, such as for example triple quadrupole mass spectrometers. Methods for utilizing MS analysis, including MALDI-TOF MS and ESI-MS, to detect the presence and quantity of biomarker peptides in biological samples are known in the art. See, e.g., U.S. Patents 6,925,389; 6,989,100; and 6,890,763 for further guidance, each of which is incorporated herein by this reference.

**[0047]** Although some embodiments of the method only call for a qualitative assessment of the presence or absence of the one or more markers in the biological sample, other embodiments of the method call for a quantitative assessment of the amount of each of the one or more markers in the biological sample. Such quantitative assessments can be made, for example, using one of the above-mentioned methods, as will be understood by those of ordinary skill in the art.

**[0048]** In some embodiments of the method, a subject is identified as having colorectal cancer upon identifying in a biological sample obtained from the subject one or more markers selected from: a peptide of Table A. In some embodiments of the method, a subject is identified having colorectal cancer upon identifying in a biological sample obtained from the subject one or more markers selected from: a peptide of Table B. In some embodiments of the method, a subject is identified having colorectal cancer upon identifying in a biological sample obtained from the subject one or more markers selected from: a peptide of Table C. In some embodiments of the method, the identification of one or more of such markers in a biological sample obtained from the subject results in the subject being identified as having a risk of colorectal cancer.

**[0049]** In some embodiments of the method, it can be desirable to include a control sample that is analyzed concurrently with the biological sample, such that the results obtained from the biological sample can be compared to the results obtained from the control sample. Additionally, it is contemplated that standard curves can be provided, with which assay results for the biological sample can be compared. Such standard curves present levels of protein marker as a function of assay units, i.e., fluorescent signal intensity, if a fluorescent signal is used. Using samples taken from multiple donors, standard curves can be provided for control levels of the one or more markers in normal tissue.

**[0050]** It is contemplated that the efficacy, accuracy, sensitivity, and/or specificity of the method can be enhanced by probing for multiple markers in the biological sample. For example, in some embodiments of the method, the biological sample can be probed for at least one marker selected from: the peptides of Table A, Table B, or Table C. For another example, in some embodiments, the biological sample can be probed for 2–5 markers selected from: the peptides of Table A, Table B, or Table C. For another example, in some embodiments, the biological sample can be probed for 6–10 markers selected from: the peptides of Table A, Table B, or Table C.

**[0051]** The analysis of markers can be carried out separately or simultaneously with additional markers within one test sample. For example, several markers can be combined into one test for efficient processing of a multiple of samples and for potentially providing greater diagnostic and/or prognostic accuracy. In addition, one of ordinary skill in the art would recognize the value of testing multiple samples (for example, at successive time points) from the same subject. Such testing of serial samples can allow the identification of changes in marker levels over time. Increases or decreases in marker levels, as well as the absence of change in marker levels, can provide useful information about the disease status that includes, but is not limited to: identifying the approximate time from onset of the event; identifying the presence and amount of salvageable tissue; identifying the appropriateness of drug therapies, including starting,

stopping, and/or changing drug therapies; identifying the relative effectiveness of various therapies; and identifying the prediction and optimization of the subject's outcome, including risk of future events.

**[0052]** The analysis of markers can be carried out in a variety of physical formats as well. For example, the use of microtiter plates or automation can be used to facilitate the processing of large numbers of test samples. Alternatively, single sample formats could be developed to facilitate immediate treatment and diagnosis in a timely fashion, for example, in ambulatory transport or emergency room settings.

**[0053]** Referring again to **FIG. 1**, in some embodiments the subject is identified as having colorectal cancer or a risk thereof if there is a measurable difference in the amount of the at least one peptide in the sample as compared to a control level **106**. Conversely, when no probed marker is identified in the biological sample, the subject can be identified as not having colorectal cancer or a risk thereof, or as having a low risk of colorectal cancer.

**[0054]** As mentioned above, depending on the embodiment of the method, identification of the one or more markers can be a qualitative determination of the presence or absence of the markers, or it can be a quantitative determination of the concentration of the markers. In this regard, in some embodiments, the step of identifying the subject as having colorectal cancer or a risk thereof requires that certain threshold measurements are made, i.e., the levels of the one or more markers in the biological sample exceed control level. In some embodiments of the method, the control level is any detectable level of the marker. In some embodiments of the method where a control sample is tested concurrently with the biological sample, the control level is the level of detection in the control sample. In some embodiments of the method, the control level is based upon and/or identified by a standard curve. In some embodiments of the method, the control level is a specifically identified concentration, or concentration range. As such, the control level can be chosen, within acceptable limits that will be apparent to those of ordinary skill in the art, based in part on the embodiment of the method being practiced and the desired specificity, etc.

**[0055]** In some cases it can be desirable not only to identify a subject as having a colorectal cancer, but to identify a subject as having a particular type of colorectal cancer. As will be understood by those of ordinary skill in the art, information about the type of cancer can be useful for making a prognosis, choosing appropriate treatment, and in some instances less severe treatment for the subject can be chosen.

**[0056]** It is contemplated that particular markers can be associated with particular types of colorectal cancer that differ with respect to their activation of cell cycle proliferation, tissue disruption, and angiogenesis gene programs. In this regard, the biomarkers set forth in Tables A,

B, and C, above, can be divided into groups or modules, as set forth in Tables H-G, below. The relative activity of genes in each of these modules indicates tumor state and predicts behavior and can be used to guide the selection of therapies and to better predict outcome than simple single measures. For example, Tables H and I include markers that are the products of genes associated with tissue disruption and angiogenesis.

Table H
COL12A1
COL1A2

Table I
TDGF1
WISP1

[0057] For another example, Tables J and K include markers that are the products of genes associated with cell cycle progression, replication, cancer, tumor morphology, and cellular movement, and genes highly correlated with transformation.

Table J
COL1A1
ENC1
FN1
PTGS2
TGFBI

Table K
MMP9
S100A9
SERpine2
SPARC
TGIF2

[0058] For another example, Tables L and M include markers that are the products of genes highly associated with the disruption of basement membranes, invasion and cell cycle progression, as well as altered transcriptional control.

Table L
BGN
MMP7

Table M
CXCL6
KIAA0963 (SBNO2)
LYZ

[0059] For another example, Tables N and O include markers that are associated with advanced stages of colorectal cancer.

Table N
AZGP1
CHI3L1
COL10A1
COL11A1
COL1A2, alternate isotype (identified using isotype-specific antibody)
COL5A2
COMP
CXCL2
CXCL3
FN1, alternate isotype (identified using isotype-specific antibody)
GDF15
IL8
INHBA
MMP12

Table O
COL8A1
CTHRC1
CXCL5
MMP11
SULF1

Table N
PTGS2, alternate isotype (identified using isotype-specific antibody)
SFRP4
SPP1
THBS2

Table O
---------

**[0013]** In some embodiments, a method for diagnosing a type of colorectal cancer in a subject comprises providing a biological sample from the subject; determining an amount in the sample of at least one peptide marker from at least one group of markers, wherein each group of markers is associated with a different type of colorectal cancer; comparing the amount of the at least one peptide marker in the sample, if present, to a control level of the at least one peptide marker, wherein the subject is diagnosed as having the type of colorectal cancer associated with a group of markers of which the at least one peptide marker is a member if there is a measurable difference in the amount of the at least one peptide in the sample as compared to the control level.

**[0014]** In some embodiments, the at least one group of markers is selected from the group of markers set forth in Tables H and I; the group of markers set forth in Tables J and I; the group of markers set forth in Tables L and M; or the group of markers set forth in Tables N and O. In some embodiments, the at least one group of markers is selected from the group of markers set forth in Table I; the group of markers set forth in Table K; the group of markers set forth in Table M; or the group of markers set forth in Table O. In some embodiments, the at least one group of markers is selected from the group of markers set forth in Table H; the group of markers set forth in Table J; the group of markers set forth in Table L; or the group of markers set forth in Table N.

**[0015]** In some embodiments, the method includes determining the amount in the sample of at least one peptide marker from at least two groups of markers. In some embodiments, the method includes determining the amount in the sample of at least one peptide marker from at least three groups of markers. In some embodiments, the method includes determining the amount in the sample of at least one peptide marker from four groups of markers. In some embodiments, the groups can be selected from: the group of Tables H and I; the group of Tables J and K; the group of Tables L and M; and the group of Tables N and O. In some embodiments, the groups can be selected from: the group of Table H; the group of Table J; the group of Table L; and the group of Table N. In some embodiments, the groups can be selected from: the group of Table I; the group of Table K; the group of Table M; and the group of Table O.

**[0016]** In some embodiments, an amount of at least one peptide marker selected from the a group of markers set forth in Table H, Table I, Table J, Table K, Table L, Table M, Table N, or

Table O is determined, and if there is a measurable difference in the amount of the at least one peptide marker from the group in the sample as compared to a control level, then the subject is identified as having a type of colorectal cancer associated the group. For example, in some embodiments, an amount of at least one peptide marker selected from a group of markers set forth in Tables F and G is determined, and if there is a measurable difference in the amount of the at least one peptide marker from the group in the sample as compared to a control level, then the subject is identified as having a particularly malignant form of colorectal cancer that will require particular modalities of treatment to be designed based on knowledge of the implicated activated pathways, as will be understood by one of ordinary skill in the art.

**[0017]** In some embodiments, an amount of at least one peptide marker selected from the a group of markers set forth in Tables H and I; the group of Tables J and K; the group of Tables L and M; and the group of Tables N and O is determined, and if there is a measurable difference in the amount of the at least one peptide marker from the group in the sample as compared to a control level, then the subject is identified as having a type of colorectal cancer associated the group.

**[0018]** In some embodiments, an amount of at least one peptide marker selected from the group of markers set forth in Table H, Table J, Table L, or Table N is determined, and if there is a measurable difference in the amount of the at least one peptide marker from the group in the sample as compared to a control level, then the subject is identified as having a type of colorectal cancer associated the group.

**[0019]** In some embodiments, an amount of at least one peptide marker selected from the a group of markers set forth in Table I, Table K, Table M, or Table O is determined, and if there is a measurable difference in the amount of the at least one peptide marker from the group in the sample as compared to a control level, then the subject is identified as having a type of colorectal cancer associated the group.

**[0020]** Clinical cancer prognosis is also an area of great concern and interest. It is useful to know the aggressiveness, stage, and/or type of the cancer cells and the likelihood of tumor recurrence in order to plan the most effective treatment. If a more accurate prognosis can be made, appropriate treatment, and in some instances less severe treatment for the subject can be chosen.

**[0021]** As such, “making a diagnosis”, as used herein, is further inclusive of making a prognosis, which can provide for selecting an appropriate treatment, or monitoring a current treatment and potentially changing the treatment, based on the measure or identity of marker

peptides. In some embodiments a method for determining whether to initiate or continue treatment of a colorectal cancer in a subject comprises: providing a series of biological samples over a time period from the subject; analyzing the series of biological samples to determine an amount in each of the biological samples of at least one peptide marker; and comparing any measurable change in the amounts of the at least one peptide marker in each of the biological samples to thereby determine whether to initiate or continue the treatment.

**[0022]** As used herein, the terms treatment or treating relate to any treatment of a condition of interest, including but not limited to prophylactic treatment, i.e., prophylaxis, and therapeutic treatment. As such, the terms treatment or treating include, but are not limited to: preventing or arresting the development of a cancer; inhibiting the progression of a cancer; reducing the severity of a cancer; ameliorating or relieving symptoms associated with a cancer; and causing a regression of cancer or one or more of the symptoms associated with the cancer.

**[0023]** In some embodiments of the presently-disclosed subject matter, multiple determination of the peptide markers over time can be made to facilitate diagnosis and/or prognosis. A temporal change in the marker can be used to monitor the progression of the cancer and/or efficacy of appropriate therapies directed against the cancer. In such an embodiment, for example, one might expect to see a decrease in the amount of one or more of the marker peptides in a biological sample over time during the course of effective treatment.

**[0024]** In some embodiments of the presently-disclosed subject matter, a system for diagnosing colorectal cancer in a subject is provided, or a system for determining whether to initiate or continue treatment of colorectal cancer in a subject is provided. Such systems can be provided, for example, as commercial kits that can be used to test a biological sample, or series of biological samples, from a subject. In some embodiments, the system includes probes for selectively binding each of one or more peptide markers; and means for detecting the binding of said probes to said one or more markers. The system can also include certain samples for use as controls. The system can further include one or more standard curves providing levels of markers as a function of assay units.

**[0025]** In some embodiments, a system for the analysis of biomarkers is provided that comprises antibodies having specificity for one or more markers associated with colorectal cancer, including a marker as set forth in Table A, B, and C. Such a system can comprise devices and reagents for the analysis of at least one test sample. The system can further comprise instructions for using the system and conducting the analysis. Optionally the systems can contain

one or more reagents or devices for converting a marker level to a diagnosis or prognosis of the subject.

**[0026]** In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers selected from the peptides of Table A. In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers selected from the peptides of Table B. In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers selected from the peptides of Table C.

**[0027]** In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers selected from a group of peptides set forth in Tables H and I, Tables J and K, Tables L and M, or Tables N and O. In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers, wherein at least one peptide marker is selected from at least two of the following groups of peptides: Tables H and I, Tables J and K, Tables L and M, or Tables N and O. In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers, wherein at least one peptide marker is selected from at least three of the following groups of peptides: Tables H and I, Tables J and K, Tables L and M, or Tables N and O. In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers, wherein at least one peptide marker is selected from four of the following groups of peptides: Tables H and I, Tables J and K, Tables L and M, or Tables N and O.

**[0028]** In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers selected from a group of peptides set forth in Table H, Table J, Table L, or Table N. In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers, wherein at least one peptide marker is selected from at least two of the following groups of peptides: Table H, Table J, Table L, or Table N. In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers, wherein at least one peptide marker is selected from at least three of the following groups of peptides: Table H, Table J, Table L, or Table N. In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers, wherein at least one peptide marker is selected from four of the following groups of peptides: Table H, Table J, Table L, or Table N.

**[0029]** In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers selected from a group of peptides set forth in Table

I, Table K, Table M, and Table O. In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers, wherein at least one peptide marker is selected from at least two of the following groups of peptides: Table I, Table K, Table M, and Table O. In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers, wherein at least one peptide marker is selected from at least three of the following groups of peptides: Table I, Table K, Table M, and Table O. In some embodiments, a system is provided including probes for selectively binding each of one or more peptide markers, wherein at least one peptide marker is selected from four of the following groups of peptides: Table I, Table K, Table M, and Table O.

**[0030]** As used herein, the term “subject” includes both human and animal subjects. Thus, veterinary therapeutic uses are provided in accordance with the presently-disclosed subject matter. As such, the presently-disclosed subject matter provides for the diagnosis of mammals such as humans, as well as those mammals of importance due to being endangered, such as Siberian tigers; of economic importance, such as animals raised on farms for consumption by humans; and/or animals of social importance to humans, such as animals kept as pets or in zoos. Examples of such animals include but are not limited to: carnivores such as cats and dogs; swine, including pigs, hogs, and wild boars; ruminants and/or ungulates such as cattle, oxen, sheep, giraffes, deer, goats, bison, and camels; and horses. Also provided is the treatment of birds, including the treatment of those kinds of birds that are endangered and/or kept in zoos, as well as fowl, and more particularly domesticated fowl, *i.e.*, poultry, such as turkeys, chickens, ducks, geese, guinea fowl, and the like, as they are also of economic importance to humans. Thus, also provided is the treatment of livestock, including, but not limited to, domesticated swine, ruminants, ungulates, horses (including race horses), poultry, and the like.

**[0031]** The presently-disclosed subject matter is further illustrated by the following specific but non-limiting examples. The following examples may include prophetic examples, and may also include compilations of data that are representative of data gathered at various times during the course of development and experimentation related to the presently-disclosed subject matter.

## EXAMPLES

### **[0060] Example 1**

#### **[0061] Identification of Markers for Colorectal Cancer.**

**[0062]** To identify new markers, the present inventors developed a gene expression database of genome-wide expression derived from about 350 human colon cancer and normal specimens. By evaluating about 54,000 genetic elements for each tumor or normal specimen, the present inventors were able to develop a list of genes that were substantially over-expressed in tumor samples versus normal controls. A genetic approach, described below, was then used to identify the genes that might produce secreted proteins. This approach ultimately identified potential biomarker proteins that could be used as serum or plasma makers. These markers could have utility for predicting the diagnosis, prognosis, of a cancer or potentially its response to a particular therapy. The markers could also be useful to develop diagnostic imaging agents to image a cancer's location, size, or extent of spread.

**[0063]** Secreted protein predictions.

**[0064]** Representative Public IDs (GenBank Nucleotide IDs) matching Affymetrix probe IDs from a U133 Plus 2.0 GeneChip (Affymetrix, Santa Clara, CA) were retrieved from the Affymetrix website. The GenBank Nucleotide IDs were then translated into GenBank Protein IDs using the Batch Entrez function at <http://www.ncbi.nlm.nih.gov/entrez/batchentrez.cgi>. The Protein IDs were translated into batches of individual FASTA sequences. These FASTA sequences were run through the publicly available SIGCLEASE database at <http://bioweb.pasteur.fr/seqanal/interfaces/sigcleave.html>. The output from SIGCLEASE was then sorted, giving two lists of proteins: 1) proteins with one or more predicted signal cleavage sites, and 2) proteins with no predicted signal cleavage sites. The sorting was carried out using Perl Script. The GenBank Protein IDs were translated back to GenBank Nucleotide IDs and then Affymetrix probe IDs. This provided a list of potentially secreted proteins with genes represented on the Affymetrix U133 Plus 2.0 GeneChip.

**[0065]** This list of Affymetrix probe IDs was then compared to two different lists.

**[0066]** The first list (Table P) was composed of genes 4 times or more increased in expression in colon tumors relative to normal colonic mucosa stratified by the four Dukes' stages individually. The second list (Table Q) was composed of genes 4 times or more increased between the mean expression of normal colon tissue and the mean expression of colon tumors of at least one of the four Dukes' stages individually. The first comparison produced a list of genes that are 4 times or more increased from normal in all four Dukes stages, which genes have products that are potentially secreted. The second comparison produced a list of genes that is 4 times or more increased from normal to at least one of the four Dukes stages and potentially secreted.

Table P

Table P

AF003934	NM_000089
AF010314	NM_001191
AF043337	NM_001394
AF101051	NM_001511
D90427	NM_002090
J04152	NM_002421
J04177	NM_002423
M13436	NM_004289
M57731	NM_004413
M80927	NM_032876
M83248	U73778
X98568	

Table Q

AB007899	AL049569	NM_000463	NM_002644	NM_014057	NM_030926
AB020635	AL049798	NM_000531	NM_002842	NM_014312	NM_032787
AB023148	AL109824	NM_000558	NM_003012	NM_014479	NM_144696
AB036706	AL133386	NM_000587	NM_003182	NM_014859	NM_144707
AF011390	AL137517	NM_000633	NM_003278	NM_014879	NM_152315
AF016004	AL137750	NM_000672	NM_003381	NM_016341	NM_153274
AF016535	AP001660	NM_000717	NM_003619	NM_016459	U10485
AF017987	BC000278	NM_000772	NM_003759	NM_016619	U19495
AF031587	BC001012	NM_000846	NM_003890	NM_017631	U43328
AF060152	BC001872	NM_000909	NM_003944	NM_017682	U64094
AF073299	BC002665	NM_000956	NM_004160	NM_017705	U71207
AF078844	BC003179	NM_001041	NM_004235	NM_017716	U73945
AF098951	BC005912	NM_001048	NM_004453	NM_017821	U80139
AF105974	BC005931	NM_001072	NM_004532	NM_018018	V00489
AF110384	BC006111	NM_001076	NM_004615	NM_018286	X16354
AF127036	BC006141	NM_001077	NM_004624	NM_018638	X17115
AF133425	BC009738	NM_001091	NM_004633	NM_018713	X51887
AF146343	BC016647	NM_001150	NM_004654	NM_019076	X72475
AF147790	BC021222	NM_001169	NM_004660	NM_019079	Y11339
AF157492	D12502	NM_001192	NM_004751	NM_019093	Y15724
AF195116	D16350	NM_001200	NM_004944	NM_020406	
AF228422	D50579	NM_001218	NM_005123	NM_020973	
AF233336	L27624	NM_001275	NM_005173	NM_020987	
AF240698	M20812	NM_001461	NM_005408	NM_020997	
AF246718	M36532	NM_001548	NM_005771	NM_021027	
AF264014	M95548	NM_001712	NM_005814	NM_021153	
AF284095	M98399	NM_001778	NM_006307	NM_022129	
AF307080	NM_000036	NM_001819	NM_006614	NM_022469	
AF311912	NM_000055	NM_001874	NM_006615	NM_022901	
AF323084	NM_000072	NM_001877	NM_006633	NM_024302	
AF325503	NM_000096	NM_001928	NM_007072	NM_024308	
AF349571	NM_000104	NM_002054	NM_007102	NM_024709	
AF350881	NM_000111	NM_002084	NM_007168	NM_024727	
AF478446	NM_000114	NM_002122	NM_007351	NM_024743	
AJ242547	NM_000115	NM_002153	NM_012128	NM_024922	
AK022831	NM_000198	NM_002196	NM_012307	NM_025087	
AL021786	NM_000222	NM_002380	NM_013261	NM_025163	
AL022324	NM_000240	NM_002407	NM_013381	NM_025184	
AL022724	NM_000336	NM_002457	NM_014033	NM_030781	

[0067] Table R include a list of genes, the expression of which is 4 times or more increased from normal in all four Dukes stages, which genes have products that are potentially secreted:

Table R			
Gene Symbol	Affymetrix Probe ID	Gene name	Number of Dukes' stages where upregulation is more than 4x
ADAM22	206616_s_at	ADAM metallopeptidase domain 22	4
AZGP1	209309_at	alpha-2-glycoprotein 1, zinc	4
AZGP1	209309_at	alpha-2-glycoprotein 1, zinc	4
BCL2L1	206665_s_at	BCL2-like 1	4
C1orf34	217101_at	chromosome 1 open reading frame 34	4
CCL20	205476_at	chemokine (C-C motif) ligand 20	4
CDH3	203256_at	cadherin 3, type 1, P-cadherin (placental)	4
CDH3	203256_at	cadherin 3, type 1, P-cadherin (placental)	4
CHI3L1	209395_at	chitinase 3-like 1 (cartilage glycoprotein-39)	4
CLDN1	222549_at	claudin 1	4
CLDN1	222549_at	claudin 1	4
CLDN2	223509_at	claudin 2	4
CLDN2	223509_at	claudin 2	4
COL10A1	217428_s_at	collagen, type X, alpha 1(Schmid metaphyseal chondrodysplasia)	4
COL11A1	204320_at	collagen, type XI, alpha 1	4
COL11A1	37892_at	collagen, type XI, alpha 1	4
COL12A1	231766_s_at	collagen, type XII, alpha 1	4
COL1A2	202404_s_at	collagen, type I, alpha 2	4
CST1	206224_at	cystatin SN	4
CXCL1	204470_at	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	4
CXCL1	204470_at	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	4
CXCL11	211122_s_at	chemokine (C-X-C motif) ligand 11	4
CXCL2	209774_x_at	chemokine (C-X-C motif) ligand 2	4
CXCL2	209774_x_at	chemokine (C-X-C motif) ligand 2	4
CXCL3	207850_at	chemokine (C-X-C motif) ligand 3	4
CXCL3	207850_at	chemokine (C-X-C motif) ligand 3	4
DPEP1	205983_at	dipeptidase 1 (renal)	4
DPEP1	205983_at	dipeptidase 1 (renal)	4
DUSP4	204014_at	dual specificity phosphatase 4	4
DUSP4	204015_s_at	dual specificity phosphatase 4	4
ENC1	201340_s_at	ectodermal-neural cortex (with BTB-like domain)	4
ENC1	201340_s_at	ectodermal-neural cortex (with BTB-like domain)	4
EREG	205767_at	epiregulin	4
GDF15	221577_x_at	growth differentiation factor 15	4
GDF15	221577_x_at	growth differentiation factor 15	4
HIG2	1554452_a_at	hypoxia-inducible protein 2	4

Table R

Gene Symbol	Affymetrix Probe ID	Gene name	Number of Dukes' stages where upregulation is more than 4x
IL8	202859_x_at	interleukin 8	4
IL8	211506_s_at	interleukin 8	4
IL8	202859_x_at	interleukin 8	4
IL8	211506_s_at	interleukin 8	4
INHBA	210511_s_at	inhibin, beta A (activin A, activin AB alpha polypeptide)	4
INHBA	210511_s_at	inhibin, beta A (activin A, activin AB alpha polypeptide)	4
JUB	1553764_a_at	jub, ajuba homolog ( <i>Xenopus laevis</i> )	4
LRP8	205282_at	low density lipoprotein receptor-related protein 8, apolipoprotein e receptor	4
LRRC14	203495_at	leucine rich repeat containing 14	4
LY6G6D	207457_s_at	lymphocyte antigen 6 complex, locus G6D	4
LY6G6D	207457_s_at	lymphocyte antigen 6 complex, locus G6D	4
MGC50722	1554681_a_at	hypothetical MGC50722	4
MMP1	204475_at	matrix metallopeptidase 1 (interstitial collagenase)	4
MMP1	204475_at	matrix metallopeptidase 1 (interstitial collagenase)	4
MMP3	205828_at	matrix metallopeptidase 3 (stromelysin 1, progelatinase)	4
MMP3	205828_at	matrix metallopeptidase 3 (stromelysin 1, progelatinase)	4
MMP7	204259_at	matrix metallopeptidase 7 (matrilysin, uterine)	4
MMP7	204259_at	matrix metallopeptidase 7 (matrilysin, uterine)	4
NFE2L3	204702_s_at	nuclear factor (erythroid-derived 2)-like 3	4
NFE2L3	204702_s_at	nuclear factor (erythroid-derived 2)-like 3	4
NUPL1	223984_s_at	nucleoporin like 1	4
SCD	200832_s_at	stearoyl-CoA desaturase (delta-9-desaturase)	4
SIX3	206634_at	sine oculis homeobox homolog 3 (Drosophila)	4
SLC6A14	219795_at	solute carrier family 6 (amino acid transporter), member 14	4
SPATA21	1570241_at	spermatogenesis associated 21	4
SPP1	209875_s_at	secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)	4
TACSTD2	202286_s_at	tumor-associated calcium signal transducer 2	4
TGFBI	201506_at	transforming growth factor, beta-induced, 68kDa	4

[0068] Table S include a list of genes, the expression of which is 4 times or more increased from normal in three of four Dukes stages, which genes have products that are potentially secreted:

Table S

Gene Symbol	Affymetrix Probe ID	Gene name	Number of Dukes' stages where upregulation is more than 4x
ABCB6	203191_at	ATP-binding cassette, sub-family B (MDR/TAP), member 6	3
ANLN	222608_s_at	anillin, actin binding protein (scraps homolog, Drosophila)	3
CEL	1553970_s_at	carboxyl ester lipase (bile salt-stimulated lipase)	3
CHI3L1	209395_at	chitinase 3-like 1 (cartilage glycoprotein-39)	3
COL11A1	37892_at	collagen, type XI, alpha 1	3
COL12A1	231879_at	collagen, type XII, alpha 1	3
COL1A1	202310_s_at	collagen, type I, alpha 1	3
COL1A2	202404_s_at	collagen, type I, alpha 2	3
COMP	205713_s_at	cartilage oligomeric matrix protein	3
CST1	206224_at	cystatin SN	3
EREG	205767_at	epiregulin	3
ESM1	208394_x_at	endothelial cell-specific molecule 1	3
FAP	209955_s_at	fibroblast activation protein, alpha	3
ITGBL1	205422_s_at	integrin, beta-like 1 (with EGF-like repeat domains)	3
KLK10	209792_s_at	kallikrein 10	3
LPGAT1	1555058_a_at	lysophosphatidylglycerol acyltransferase 1	3
NUPL1	223984_s_at	nucleoporin like 1	3
OLR1	210004_at	oxidised low density lipoprotein (lectin-like) receptor 1	3
SCD	200832_s_at	stearoyl-CoA desaturase (delta-9-desaturase)	3
SERPINB5	204855_at	serpin peptidase inhibitor, clade B (ovalbumin), member 5	3
SLC6A14	219795_at	solute carrier family 6 (amino acid transporter), member 14	3
SPP1	209875_s_at	secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)	3
TACSTD2	202286_s_at	tumor-associated calcium signal transducer 2	3
THBS2	203083_at	thrombospondin 2	3
TMEM44	1560275_at	transmembrane protein 44	3
TNFRSF12A	218368_s_at	tumor necrosis factor receptor superfamily, member 12A	3
TNFSF11	210643_at	tumor necrosis factor (ligand) superfamily, member 11	3

**[0069]** Table T include a list of genes, the expression of which is 4 times or more increased from normal in two of four Dukes stages, which genes have products that are potentially secreted:

Table T			
Gene Symbol	Affymetrix Probe ID	Gene name	Number of Dukes' stages where upregulation is more than 4x
ARNTL2	220658_s_at	aryl hydrocarbon receptor nuclear	2

Table T

Gene Symbol	Affymetrix Probe ID	Gene name	Number of Dukes' stages where upregulation is more than 4x
		translocator-like 2	
CCL18	32128_at	chemokine (C-C motif) ligand 18 (pulmonary and activation-regulated)	2
CEL	205910_s_at	carboxyl ester lipase (bile salt-stimulated lipase)	2
CEL	1553970_s_at	carboxyl ester lipase (bile salt-stimulated lipase)	2
COL12A1	231766_s_at	collagen, type XII, alpha 1	2
COL5A2	221730_at	collagen, type V, alpha 2	2
CXCL11	210163_at	chemokine (C-X-C motif) ligand 11	2
CXCL11	210163_at	chemokine (C-X-C motif) ligand 11	2
DNAJC3	208499_s_at	DnaJ (Hsp40) homolog, subfamily C, member 3	2
FN1	211719_x_at	fibronectin 1 // fibronectin 1	2
HEATR3	1554478_a_at	HEAT repeat containing 3	2
LCN2	212531_at	lipocalin 2 (oncogene 24p3)	2
MMP12	204580_at	matrix metalloproteinase 12 (macrophage elastase)	2
MSLN	204885_s_at	mesothelin	2
MTHFD1L	225520_at	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like	2
NEBL	203961_at	nebulette	2
NFAT5	208003_s_at	nuclear factor of activated T-cells 5, tonicity-responsive	2
PHF20	209422_at	PHD finger protein 20	2
PPM1H	212686_at	protein phosphatase 1H (PP2C domain containing)	2
PPM1H	212686_at	protein phosphatase 1H (PP2C domain containing)	2
PTGS2	204748_at	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	2
SFRP4	204052_s_at	secreted frizzled-related protein 4	2
SLCO1B3	206354_at	solute carrier organic anion transporter family, member 1B3	2
SRPX2	205499_at	sushi-repeat-containing protein, X-linked 2	2
TGFBI	201506_at	transforming growth factor, beta-induced, 68kDa	2
TRIB3	218145_at	tribbles homolog 3 (Drosophila)	2

[0070] Table U include a list of genes, the expression of which is 4 times or more increased from normal in one of four Dukes stages, which genes have products that are potentially secreted:

Table U

Gene Symbol	Affymetrix Probe ID	Gene name	Number of Dukes' stages where upregulation is more than 4x
ACSL6	211207_s_at	acyl-CoA synthetase long-chain family member 6	1

Table U

Gene Symbol	Affymetrix Probe ID	Gene name	Number of Dukes' stages where upregulation is more than 4x
ADAM12	202952_s_at	ADAM metallopeptidase domain 12 (meltrin alpha)	1
AREG /// LOC653193	205239_at	amphiregulin (schwannoma-derived growth factor) /// similar to Amphiregulin precursor (AR) (Colorectum cell-derived growth factor) (CRDGF)	1
BAMBI	203304_at	BMP and activin membrane-bound inhibitor homolog (Xenopus laevis)	1
BAMBI	203304_at	BMP and activin membrane-bound inhibitor homolog (Xenopus laevis)	1
BAX	211833_s_at	BCL2-associated X protein	1
BGN	201261_x_at	biglycan	1
C17orf69	1553399_a_at	chromosome 17 open reading frame 69	1
C2	1554533_at	complement component 2	1
CCL20	205476_at	chemokine (C-C motif) ligand 20	1
CCL3 /// CCL3L1 /// CCL3L3 /// LOC643930	205114_s_at	chemokine (C-C motif) ligand 3 /// chemokine (C-C motif) ligand 3-like 1 /// chemokine (C-C motif) ligand 3-like 3 /// similar to Small inducible cytokine A3-like 1 precursor (Tonsillar lymphocyte LD78 beta protein) (LD78-beta(1-70)) (G0/G1 switch regulatory protein 19-2) (G0S19-2 protein) (PAT 464.2)	1
CDH11	207173_x_at	cadherin 11, type 2, OB-cadherin (osteoblast)	1
CFB	202357_s_at	complement factor B	1
COL10A1	217428_s_at	collagen, type X, alpha 1(Schmid metaphyseal chondrodysplasia)	1
CSPG2	204620_s_at	chondroitin sulfate proteoglycan 2 (versican)	1
CXCL11	211122_s_at	chemokine (C-X-C motif) ligand 11	1
DKK2	219908_at	dickkopf homolog 2 (Xenopus laevis)	1
DUOX2	219727_at	dual oxidase 2	1
DUOX2	219727_at	dual oxidase 2	1
DUSP4	204014_at	dual specificity phosphatase 4	1
ETV4	1554576_a_at	ets variant gene 4 (E1A enhancer binding protein, E1AF)	1
ETV4	1554576_a_at	ets variant gene 4 (E1A enhancer binding protein, E1AF)	1
GTF2IRD1	218412_s_at	GTF2I repeat domain containing 1	1
HEATR3	1554478_a_at	HEAT repeat containing 3	1
HIG2	1554452_a_at	hypoxia-inducible protein 2	1
KDELC1	219479_at	KDEL (Lys-Asp-Glu-Leu) containing 1	1
LCN2	212531_at	lipocalin 2 (oncogene 24p3)	1
MMP12	204580_at	matrix metallopeptidase 12 (macrophage elastase)	1
MTHFD1L	225520_at	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like	1
NEBL	203961_at	nebulette	1
PTPLAD1	234000_s_at	protein tyrosine phosphatase-like A domain containing 1	1
QPCT	205174_s_at	glutaminyl-peptide cyclotransferase (glutaminyl cyclase)	1

Table U

Gene Symbol	Affymetrix Probe ID	Gene name	Number of Dukes' stages where upregulation is more than 4x
REG3A	205815_at	regenerating islet-derived 3 alpha	1
REG3A	205815_at	regenerating islet-derived 3 alpha	1
RNF43	218704_at	ring finger protein 43	1
SERPINB5	204855_at	serpin peptidase inhibitor, clade B (ovalbumin), member 5	1
SLC7A11	209921_at	solute carrier family 7, (cationic amino acid transporter, y <sup>+</sup> system) member 11	1
SLC7A11	209921_at	solute carrier family 7, (cationic amino acid transporter, y <sup>+</sup> system) member 11	1
SLCO1B3	206354_at	solute carrier organic anion transporter family, member 1B3	1
SLCO4A1	219911_s_at	solute carrier organic anion transporter family, member 4A1	1
SMG1	210057_at	PI-3-kinase-related kinase SMG-1	1
SRPX2	205499_at	sushi-repeat-containing protein, X-linked 2	1
TAC4	1553094_at	tachykinin 4 (hemokinin)	1
TRIB3	218145_at	tribbles homolog 3 (Drosophila)	1
TSPAN5	209890_at	tetraspanin 5 /// tetraspanin 5	1

**[0071] Example 2**

**[0072]** The colon is composed of a dynamic and self-renewing epithelium that turns over every 3-5 days. Without wishing to be bound by theory, it is thought that at the base of the crypt, variable numbers (between one and 16) of slowly dividing, stationary, pluripotent stem cells give rise to more rapidly proliferating, transient amplifying cells. These cells differentiate chiefly into post-mitotic columnar colonocytes, mucin-secreting goblet cells and enteroendocrine cells as they migrate from the crypt base to the surface where they are sloughed into the lumen [1].

**[0073]** Several signaling pathways, including Wnt, Tgf $\beta$ , Bmp, Hedgehog and Notch, play roles in the control of proliferation and differentiation of the developing and adult colon [2]. Their perturbation via mutation or epigenetic modification occurs in human colorectal cancer (CRC) and doing so via genetic engineering confers risk for neoplasia in mouse models. Moreover, tumor cell de-differentiation correlates with key tumor features such as tumor progression rates, invasiveness, drug resistance and metastatic potential [3-5].

**[0074]** A variety of scientific and organizational obstacles tend to interfere with the ability to integrate and compare detailed characterizations of human cancer to corresponding molecular and clinical characterizations from genetically engineered mouse models of normal, cancer, and other disease processes. Because of this, there integrated views of the molecular bases of cancer risk, development, and progression are lacking. To approach these problems and learn more about the opportunities potentially afforded by large integration and cross-models comparisons, a consortium-based combined analysis was undertaken of a variety of mouse models of colorectal

cancer (CRC) that have been developed and used individually (reviewed in [6, 7]), and the results were compared with human CRC, in an effort to develop an increased understanding of fundamental aspects of the pathogenesis of CRC.

**[0075]** In the studies described herein, gene expression patterns of human colorectal cancer (CRC) and mouse models of colon cancer were compared to that of embryonic mouse colon development using colon samples from embryonic days 13.5-18.5.

**[0076]** Mouse Models of Colorectal Cancer (CRC)

**[0077]** There are four mouse models of CRC that were used in the studies described herein: (1) *Apc*<sup>Min/+</sup> mouse model (multiple intestinal neoplasia), (2) azoxymethane (AOM) carcinogen model, (3) *Tgfb1*<sup>-/-</sup> / *Rag2*<sup>-/-</sup> mouse model, and (4) *Smad3*<sup>-/-</sup> mouse model.

**[0078]** The *Apc*<sup>Min/+</sup> (multiple intestinal neoplasia) mouse model harbors a germline mutation in the *Apc* tumor suppressor gene and exhibits multiple tumors in the small intestine and colon [8]. A major function of APC is to regulate the canonical WNT signaling pathway as part of a  $\beta$ -catenin degradation complex. Loss of APC results in a failure to degrade  $\beta$ -catenin, which instead enters the nucleus to act as a transcriptional co-activator with the lymphoid enhancer factor/T-cell factor (LEF/TCF) family of transcription factors [9]. The localization of  $\beta$ -catenin within the nucleus indicates activated canonical WNT signaling. In addition to germline *APC* mutations that occur in persons with familial adenomatous polyposis coli (FAP) and *Apc*<sup>Min/+</sup> mice, loss of functional APC and activation of canonical WNT signaling occurs in more than 80% of human sporadic CRCs [10].

**[0079]** Similar to the *Apc*<sup>Min/+</sup> model, tumors in the azoxymethane (AOM) carcinogen model, which occur predominantly in the colon [11], have signaling alterations marked by activated canonical WNT signaling.

**[0080]** Two other mouse models that carry different genetic alterations leading to colon tumor formation are based on the observation that TGF $\beta$  type II receptor (*TGFBR2*) gene mutations are present in up to 30% of sporadic CRC and in more than 90% of tumors that occur in patients with the DNA mismatch repair deficiency associated with hereditary non-polyposis colon cancer (HNPCC) [12].

**[0081]** In the mouse, a deficiency of TGF $\beta$ 1 combined with an absence of T-cells (*Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup>) results in a high occurrence of colon cancer [13]. These mice develop adenomas by two months of age, and adenocarcinomas, often mucinous, by 3-6 months of age. Immunohistochemical analyses of these tumors are negative for nuclear  $\beta$ -catenin, suggesting that TGF $\beta$ 1 does not suppress tumors via a canonical WNT signaling-dependent pathway.

**[0082]** The SMAD-family proteins are critical downstream transcription regulators activated by TGF $\beta$  signaling, in part through the TGF $\beta$  type II receptor. *Smad3*<sup>-/-</sup> mice also develop intestinal lesions that include colon adenomas and adenocarcinomas by six months of age [14].

**[0083]** To identify transcriptional programs that are activated or repressed in different colon tumor models, gene expression profiles of (1) 100 human CRCs and 39 colonic tumors from the four models of colon cancer to (2) mouse embryonic and mouse and human adult colon were compared. The results of these analyses demonstrate that tumors from the mouse models extensively adopt embryonic gene expression patterns, irrespective of the initiating mutation.

**[0084]** Although two of the mouse tumor subtypes were distinguishable by their relative shifts towards early or later stages of embryonic gene expression (driven principally by localization of  $\beta$ -catenin to the nucleus versus the plasma membrane), *Myc* was over-expressed in tumors from all four tumor-models. Further, by mapping mouse genes to their corresponding human orthologs, it was further shown that human CRCs share in the broad over-expression of genes characteristic of colon embryogenesis and the up-regulation of *MYC*, consistent with a fundamental relationship between embryogenesis and tumorigenesis. Large scale additional similarities could also be found at the level of developmental genes that were not activated in either mouse or human tumors. In addition, there were transcriptional modules consistently activated and repressed in human CRC that were not found in the mouse models. Taken together, this cross-species, cross-models analytical approach, filtered through the lens of embryonic colon development, provides an integrated view of gene expression that implicates the adoption of a broad program that encompasses embryonic activation, developmental arrest, and failed differentiation as fundamental characteristics of human CRC.

**[0085]** Example: Methods - Mouse models, human CRC patients, and tumor collection.

**[0086]** Mouse tumors.

**[0087]** All tumors were isolated as spontaneously-occurring lesions in *Apc*<sup>Min/+</sup> [56], *Smad3*<sup>-/-</sup> [57], and *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup> mice were collected at three-to-nine months of age depending on the model (for a review, see [6]). The only exceptions were two *Apc*<sup>Min/+</sup> tumors UW\_3\_2778 and UW\_6\_2748 that were 13 and 14 months and the three *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup> tumors, all five of which had histological features of locally invasive carcinoma [7].

**[0088]** Three-to-four month old mice from various AXB recombinant inbred lines were treated with AOM doses chosen for enhancement of inter-strain differences in susceptibility [11]. Mice were given four weekly i.p. injections of 10 mg AOM per kg body weight, and tumors were collected six months after the first injection. Animals were euthanized with CO<sub>2</sub>, colons removed,

flushed with 1X PBS, and laid out on Whatman 3MM paper. A summary of the mouse strains, mutant alleles and source laboratories is presented in Table V.

Table V: Mouse Models of Colon Cancer		
Model	Mm strain	n
Azoxymethane (AOM)	A x B	14
<i>Apc</i> <sup>Min/+</sup>	(SWR x B6) F1	2
<i>Apc</i> <sup>Min/+</sup>	(BR x B6) F1	2
<i>Apc</i> <sup>Min/+</sup>	C57BL/6	5
<i>Smad3</i> <sup>-/-</sup>	129	6
<i>Smad3</i> <sup>-/-</sup>	129	7
<i>Tgfb1</i> <sup>-/-</sup> ; <i>Rag2</i> <sup>-/-</sup>	C57BL/6	3

Tumors from four established mouse models of CRC (*Apc*<sup>Min/+</sup>, AOM, *Smad3*<sup>-/-</sup> and *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup>) were analyzed. The table provides details on the mouse strains used for the four models, as well as information on the number of samples generated per model and sample-originating laboratory.

**[0089]** All tumors were obtained from the colon only, the particular segment of which is indicated in the GEO database reposed sample information ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/), GSE5261). The majority of *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup> and *Smad3*<sup>-/-</sup> tumors occur in the cecum and proximal colon and all samples isolated for characterization were obtained from there. In contrast tumors isolated from *Apc*<sup>Min/+</sup> and AOM mice occurred predominantly in the mid- and distal colon. A small portion of the tumor was placed in formalin for histology, with the remainder finely dissected into RNAlater (Ambion) and stored at -20°C. Normal adult colon RNA for reference was obtained from whole colon samples harvested from ten 8-week old C57BL/6 male mice. The tissue was lysed in Trizol Reagent and homogenized. Total RNA was purified using a Qiagen kit.

**[0090]** Human samples – collection/biopsies, regulatory aspects, compliance and informed consents.

**[0091]** Sample collection protocol and analyses were performed as described in Kwong, et al., *Genomics* (2005) [36]. Information collected with the samples for this study includes TNM and Dukes staging/presentation criteria, pathological diagnosis and differentiation criteria.

**[0092]** RNA isolation.

**[0093]** All RNA samples were purified using Trizol Reagent (Invitrogen Systems, Inc.) from finely dissected tumors and were subjected to quality control screening using the Agilent BioAnalyzer 2100.

**[0094]** Example: Methods - Microarray procedures and data analysis

**[0095]** Mouse cDNA arrays.

**[0096]** Mouse tumors were analyzed on Vanderbilt University Microarray Core (VUMC)-printed 20K mouse cDNA arrays, composed principally of PCR products derived from three sources: the 15K National Institute of Aging mouse cDNA library; the Research Genetics mouse 5K set; and an additional set of cDNAs mapped to RefSeq transcripts. Labeling, hybridization, scanning, and quantitative evaluation of these two-color channel arrays were performed according to VUMC protocols (<http://array.mc.vanderbilt.edu/microarray/spotted.vmsr>) using a Universal Reference standard (embryonic day 17.5 (E17.5) whole fetal mouse RNA). Arrays were analyzed by GenePix version 3.0, flagged and filtered for unreliable measurements, with dye channel ratios corrected using Lowess and dye-specific correction normalization as described in Park, et al. Genesis (2005) [15].

**[0097]** Human Affymetrix oligonucleotide arrays.

**[0098]** Human RNA samples were labeled for hybridization to Affymetrix HG-U133plus2 microarrays using the Affymetrix-recommended standard labeling protocol (Small-scale labeling protocol version 2.0 with 0.5 µg of total RNA; Affymetrix Technical Bulletin). Microarrays were scanned with MicroarraySuite version 5.0 to generate “CEL” files that were processed using the RMA algorithm as implemented by Bioconductor [15].

**[0099]** Analysis strategy.

**[00100]** The four different mouse models of CRC were compared for model-specific differences, then compared to mouse colon development stages, and then to human CRC samples (**FIG. 2**). Colon tumors from four etiologically distinct mouse models of CRC were subjected to microarray gene expression profiling. The gene expression profiles from the different mouse model tumors were compared and contrasted to each other, as well as to those from embryonic mouse colon development and 100 human CRCs. The mouse tumor sample array data are comprised of Lowess-normalized Cy3:Cy5 labeling ratios of each individual tumor sample versus a Universal E17.5 whole fetal mouse reference RNA (See description found in the NCBI GEO database under series accession numbers [GSE5261], which is incorporated herein in its entirety by this reference; See also supplementary information, which is available at <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5204>, which is incorporated herein in its entirety by this reference, Query DataSets for GSE5204, Title-Mouse models of human colon cancer and mouse colon development, Organism(s) <http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=10090>, Type-

**[00101]** Tumors from four murine models of colorectal cancer and normal mouse colon samples at different developmental stages, Summary – experiments to identify transcriptional patterns across tumors from colorectal cancer murine models and normal mouse colon samples at

different developmental stages, Experiment description - Colorectal cancer (CRC) results from multiple genetic and epigenetic events that produce variable histologies and clinical outcomes. To identify gene regulatory programs that underlie colon tumorigenesis, gene expression was profiled in 39 mouse colon tumors from four independent mouse models and compared to mouse colon embryonic development, as well as with 100 human colon carcinomas. There was a striking recapitulation of embryonic patterns of gene expression in both mouse and human colon tumors. All four of the mouse colon tumor models exhibited large-scale activation of embryonic gene expression signatures. The two nuclear beta-catenin-positive mouse tumors (azoxymethane-treated [AOM] and ApcMin/+) exhibited strong activation of genes characteristic of those expressed in the earliest embryonic stages, while tumors from two other models (Smad3-/- and Tgfb1-/- x Rag2-/-) exhibited lower activation of early stage-specific genes but substantial expression of general embryonic colon genes. Human colon cancer cases over-expressed genes characteristic of both early and late embryonic stages. Examining tumor gene expression through the lens of development has revealed an extensive network of therapeutic targets for cancer control. Overall design - Four colorectal tumor models were compared using the approach of finding model-specific differences with gene expression levels referenced to the median gene expression value across all models. A second complementary strategy compared the gene expression levels of the murine tumors to those of mouse normal adult colon).

**[00102]** The first approach to referencing was to compare normalized ratios across the tumor series. To do this, for each gene, the Lowess-corrected ratio for each probe element (sample vs. E17.5 whole fetal mouse reference) was divided by the median ratio for that probe across the entire tumor sample series. This is termed the median-per-tumor expression ratio and was useful for identifying, clustering and visualizing differences that occur between the different tumor samples. Mouse expression data had been collected for normal E13.5-E18.5 colon samples from inbred C57BL/6J and outbred CD-1 mice [15] using the identical E17.5 whole fetal mouse reference, allowing the data to be combined directly.

**[00103]** Differential expression profiles in the tumors were combined with relative developmental gene expression levels by direct comparisons of ratios determined within each experimental series. Initial comparisons were made between median normalized tumor data to that of gene expression levels observed in the E13.5-E18.5, and adult (8 week post-natal) colon samples that were referenced to either E13.5 samples or to the adult colon. The latter approach subsequently allowed for the broadest comparison of mouse and human data using gene ortholog mapping. Correlated phenomena could be observed from any of the different referencing strategies.

**[00104]** Inter-organism gene ortholog and inter-platform comparison strategy.

**[00105]** Pairs of human and mouse ortholog genes (12,693) were curated using Mouse Genome Informatics (MGI – The Jackson Laboratory - <http://www.informatics.jax.org/>) and the National Center for Biotechnology Information - Homologene (NCBI - <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene>) databases. Individual microarray elements or features were mapped to these. The concatenated human and mouse RefSeq IDs was used as the composite ID for the orthologous gene pair in the ortholog genome definition. NIA/Research Genetics mouse cDNAs were mapped to human orthologs using a variety of resources, usually via the Stanford Online Universal Reference resource (<http://source.stanford.edu/>). Gene-transcript assignments are made unique by choosing the longest corresponding transcript.

**[00106]** To map the Affymetrix human and mouse array data into the ortholog genome, a sequence matching approach was used. First, human and mouse transcript sequences were obtained from RefSeq (<ftp://ftp.ncbi.nih.gov/refseq/>) and probe sequences from the manufacturer's website

(<http://www.affymetrix.com/support/technical/byproduct.affx?cat=arrays>). Next, all perfect probe-transcript pairs were computed. Probes that matched multiple gene symbols were excluded, but probes that matched multiple transcripts were accepted. Probe sets were assigned to represent a given transcript if at least 50% of the perfect match probes of the probe set matched to that transcript. The newly assigned transcript identifiers were then used to map probe sets to ortholog genes.

**[00107]** Since some transcripts have multiple probe-set representations on both the Affymetrix and cDNA microarrays to one ortholog identifier, an *ad hoc* strategy was employed to use the average of those probe sets or cDNAs that exhibited consistent regulation across a sample series. In such cases, the signals of the regulated probe sets that were interpreted as being in agreement were averaged and assigned to the corresponding ortholog. Probe sets or cDNAs were excluded if it was known that they corresponded to non-transcript genomic sequence as tested using BLAT at <http://genome.ucsc.edu/>.

**[00108]** Mouse-human RefSeq gene ortholog assignments can be found at <http://genometrafac.cchmc.org/>. All ortholog assignments and cross-species mapping annotations were incorporated into annotations associated with the Affymetrix HG-U133 plus2.0 genome. Gene expression ratios obtained for the mouse samples were then represented as expression values within the human platform for all of the probe sets that mapped to the corresponding mouse gene ortholog. Data for the primary human sample series, as well as the

combined mouse-human data sets, are available at <http://genet.cchmc.org/> in the HG-U133 genome under the KaiserEtAl\_2006 folders (“guest” login, all cross-platform ortholog gene identifiers are contained as annotation fields within the HG-U133 genome table).

**[00109]** Statistical and data visualization approaches.

**[00110]** Most normalization, expression-level referencing, statistical comparisons, and data visualization were performed using GeneSpring v7.0 (Silicon Genetics-Agilent). Fisher’s exact test was performed online at <http://www.matforsk.no/ola/fisher.htm>. To identify differentially expressed features between 2 or more classes, GeneSpring’s Wilcoxon-Mann-Whitney or the Kruskal-Wallis test were applied, respectively. For 3 or more classes, the initial non-parametric test was followed by the Student-Newman-Keuls *post-hoc* test. Results from the primary analyses were corrected for multiple testing effects by applying Benjamini and Hochberg false discovery rate correction (FDR, [58]). In general, due to the referencing strategies, good platform technical performances, and moderately low within-group biological variation of gene expression, stringent cutoffs could be used, i.e., the FDR level of significance was set between  $FDR < 5.10^{-5}$  and  $FDR < 5.10^{-4}$ . K-means clustering was performed using the GeneSpring K-means tool and the Pearson correlation similarity measure.

**[00111]** Example: Methods - Ontology-based analysis of gene cluster-associated functional correlates.

**[00112]** Gene expression clusters were analyzed for the occurrence of multiple genes involved in related gene function categories by comparing each list of coordinately regulated clustered genes to categories within GeneOntology, pathways, or literature-based gene associations using GATACA (<http://gataca.cchmc.org/>), Ontoexpress (<http://vortex.cs.wayne.edu/projects.htm/>), and Ingenuity Pathway Analysis, version 3 (IPA; <http://www.ingenuity.com/>, Ingenuity Systems, Redwood City, CA). To do this, each cluster indicated in **FIGS. 3, 4, 6, 7, and 9** was converted to a list of gene identifiers, uploaded to the application, and examined for over-representation of multiple genes from one or more molecular networks, or functional or disease associations as developed from literature mining. Networks of these focus genes were algorithmically generated based on the relationships of individual genes as derived from literature review and used to identify the biological functions and/or associated pathological processes most significant for each gene cluster. Fisher’s exact test was used to calculate a *p*-value estimating the probability that a particular functional classification or category of genes is associated with a particular pattern or cluster of gene expression more than would be expected by chance.

**[00113]** For each cluster, only the top significant functional classes and canonical pathways are shown. **FIG. 6A** shows a diagram of the canonical WNT signaling pathway and an associated-gene network that was a top-ranked association of the clusters that exhibited significant overexpression in AOM and *Apc*<sup>Min/+</sup> versus *Smad3*<sup>-/-</sup> and *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup> mouse models. Genes or gene products are represented as nodes, and biological relationships between nodes are represented as edges (lines). All edges are supported by at least one literature reference from a manuscript, or from canonical information stored in the Ingenuity Pathways Knowledge Base.

**[00114]** With reference to FIGS. 8A and 8B, the up-regulated signature in tumors from *Apc*<sup>Min/+</sup> (M) and AOM (A) models (cluster C6) is enriched with genes associated with the activation of the canonical WNT signaling pathway, as determined by nuclear  $\beta$ -catenin positivity. FIG. 6A depicts a schematic diagram of the canonical WNT signaling pathway showing elements present in C6 (gene symbols with gray background). Key elements of this pathway (*Ctnnb1*, *Lef1*, *Tcf* and *Myc*) are outlined in blue. FIG. 6B depicts relative gene expression for *MYC* and *SOX4* is plotted for individual murine and human tumors. The relative expression level of *MYC* and *SOX4* is normalized to adult colon. Note that whereas *Sox4*, a canonical WNT target gene, is expressed at high levels in all human CRCs, A/M tumors and during embryonic mouse colon development, it is not expressed in S and T tumors (black). In contrast, *MYC* is over-expressed in all human and murine tumors and during colonic embryonic development (red), irrespective of the activation of canonical WNT signaling, as determined by nuclear  $\beta$ -catenin positivity.

**[00115]** Example: Methods - Real-time quantitative PCR (RTQ-PCR)

**[00116]** To confirm the validity of data normalization and referencing procedures as well as the cDNA gene assignments of the printed arrays used in the microrarray analyses, quantitative real-time polymerase chain reaction (RTQ-PCR) was used to measure relative levels of nine genes found by microrarray data analysis to be differentially expressed (FDR<5.10<sup>-5</sup>) in tumors from *Apc*<sup>Min/+</sup> and *Smad3*<sup>-/-</sup> mice.

**[00117]** Total RNAs from C57BL6 *Apc*<sup>Min/+</sup> and 129 *Smad3*<sup>-/-</sup> tumor samples (20  $\mu$ g) were reverse-transcribed to cDNA using the High Capacity cDNA Archive Kit (oligo-dT primed; Applied Biosystems). RTQ-PCR reactions (20  $\mu$ l) were set up in 96-well MicroAmp Reaction Plates (Applied Biosystems) using 10 ng of cDNA template in Taqman Universal PCR Master Mix and 6-FAM-labeled Assays-on-Demand primer-probe sets (Applied Biosystems). Reactions were run on an MX3000P (Stratagene) with integrated analysis software. Threshold cycle numbers (C<sub>t</sub>) were determined for each target gene using an algorithm that assigns a fluorescence baseline based on measurements prior to exponential amplification. Relative gene expression

levels were calculated using the  $\Delta\Delta Ct$  method [59], with the *Gusb* gene as a control. Fold-change was determined relative to expression in normal adult colon from two C57BL/6J mice.

**[00118]** Example: Methods – Immunohistochemistry

**[00119]** Immunohistochemical procedures were performed as described in Park, et al., *Genesis* (2005) [15]. *Apc*<sup>Min/+</sup> and *Smad3*<sup>-/-</sup> colon tumors were rapidly dissected, fixed in 4% paraformaldehyde, and embedded in paraffin before cutting 10 $\mu$ m thick sections. Antigen-retrieval was performed by boiling for 20 min in citrate buffer, pH 6.0. Sections were treated with 0.3% hydrogen peroxide in PBS for 30 min, washed in PBS, blocked in PBS plus 3% goat serum and 0.1% Triton X-100, and then incubated with primary antibodies and HRP-conjugated goat anti-rabbit secondary antibody (Sigma, St. Louis, MO). Antigen-antibody complexes were detected with a DAB peroxidase substrate kit (Vector Laboratories, Burlingame, CA) according to the manufacturer's protocol.

**[00120]** Example: Results

**[00121]** Thirty-nine 39 colon tumors from four independent mouse models and 100 human CRCs all share recapitulation of embryonic colon gene expression. A large-scale expression module based on increased expression by embryonic colon, all mouse tumor models, and human CRCs was composed of genes responsible for control of cell cycle progression, proliferation, and migration such as MYC, AKT2, ANLN, BIRC5, CSE1L, ITGB5, MAD2L1, MIF, MSF, PLK1 and SPARC. Nuclear  $\beta$ -catenin status subdivided mouse tumor models into nuclear positive (azoxymethane-treated [AOM] and *Apc*<sup>Min/+</sup>) tumors exhibiting greater expression of early colon development genes versus nuclear negative (*Smad3*<sup>-/-</sup> and *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup>) tumors with a later developmental profile. Both human and mouse tumors differed from embryonic colon by loss of a shared tumor suppressor-containing module (EDNRB, HSPE, KIT and LSP1). Human CRC adenocarcinomas lost an additional suppressor module (BCL2, IGFBP4, MAP4K1, PDGFRA, STAB1 and WNT4) and in many of the tumor samples, gained a module associated with advanced malignancy (ABCC1, FOXO3A, LIF, PIK3R1, PRNP, TNC, TIMP3 and VEGF).

**[00122]** Example: Results - Strategy for cross-species analysis

**[00123]** With reference to **FIG. 2**, colon tumors from four etiologically distinct mouse models of CRC were subjected to microarray gene expression profiling. The gene expression profiles from the different mouse model tumors were compared and contrasted to each other, as well as to those from embryonic mouse colon development and 100 human CRCs. The strategy for the characterization of mouse models of human CRC relies on gene expression differences and relative patterning across a range of mouse CRC models, normal mouse colon developmental stages, and human CRCs. Achieving this comparison was facilitated by the use of reference

RNAs from whole-mouse and normal adult colon reference RNAs for both mouse and human measurements. Mouse tumor samples were profiled on cDNA microarrays using the embryonic day (E)17.5 whole mouse reference RNA identical to that described in Park, et al., *Genesis* (2005) [15] to examine embryonic mouse colon gene expression dynamics during from E13.5 to E18.5 during which time the primitive, undifferentiated, pseudo-stratified colonic endoderm becomes a differentiated, single-layered epithelium. This strategy allowed us to construct a gene expression database of mouse colon tumors in which gene expression levels of the tumors could be referenced, ranked, and statistically compared to an average value among the tumors or to embryonic or adult colon gene expression levels on a per-gene basis. First, the four models were compared with each other, then to mouse colon development, and finally to human CRCs using gene ortholog mapping (**FIG. 2**).

**[00124]** Example: Results - Mouse colon tumors partition into classes reflecting differential canonical WNT signaling activity.

**[00125]** To discover gene expression programs underlying differences between etiologically distinct mouse models of CRC, gene expression level values for each transcript in each tumor sample was set to its ratio relative to its median across the series of tumor models. Using non-parametric statistical analyses, 1798 cDNA transcripts were identified as differentially expressed among the four mouse models of CRC. Five major gene patterns were identified using K-means clustering (clusters C1-C5). Genes belonging to these clusters were strongly associated with annotated gene function categories (see Table W for detailed biological descriptions and associations). For example, cluster C1, composed of transcripts that exhibited lower expression in *Smad3*<sup>-/-</sup> tumors and higher expression in AOM, *Apc*<sup>Min/+</sup> and *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup> tumors, contains 391 transcripts including *Cdk4*, *Ctnnb1*, *Myc*, *Ezh2*, *Mcm2* and *Tcf3*. Gene list over-representation analysis using DAVID and Ingenuity Pathway Analysis applications demonstrated highly significant associations to cell cycle progression, replication, post-transcriptional control and cancer. Similarly, cluster C2, composed of 663 transcripts that exhibited high expression in AOM and *Apc*<sup>Min/+</sup> tumors, but low in *Smad3*<sup>-/-</sup> and *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup> tumors, included transcripts for contact growth inhibition (*Metap1*, *Pcyox1*), mitosis (*Mif*, *Pik1*), cell cycle progression and checkpoint control (*Id2*, *Ptp4A2*, *Tp53*).

**[00126]** Active canonical WNT signaling (as determined by nuclear  $\beta$ -catenin) stratifies the four murine colon tumor models into two groups. 1798 gene transcripts are identified as differentially expressed among any of the four mouse tumor models (Kruskal-Wallis test + Student-Newman-Keuls test + FDR<5.10<sup>-5</sup>) (Data not shown). Results demonstrate that AOM (A) and *Apc*<sup>Min/+</sup> (M) tumors are transcriptionally more similar to each other than to tumors from

*Smad3*<sup>-/-</sup> (S) and *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup> (T) mice. Five clusters have been identified (C1-C5) that correspond to the K-means functional clusters listed in Table W. Please refer to Table W for a description of the functional classification of the genes found in these clusters. The lower panel illustrates the extent of the similarity between A/M and S/T tumors by identifying the top-ranked 1265 transcripts of the 1798 that were higher or lower in the two tumor super-groups (rank based on Wilcoxon-Mann-Whitney test for between-group differences with a FDR<5.10<sup>-5</sup> cutoff). Up-regulated transcripts in A/M tumors are highly enriched for genes associated with canonical WNT signaling activity, cell proliferation, chromatin remodeling, cell cycle progression and mitosis; transcripts over-expressed in S/T tumors are highly enriched for genes related to immune and defense responses, endocytosis, transport, oxidoreductase activity, signal transduction and metabolism. Representative histologies for each of the four tumor models were obtained (data not shown). Model-dependent localization of  $\beta$ -catenin was shown. Tumors from M and A (not shown) mice exhibited prominent nuclear  $\beta$ -catenin accumulation and reduced cell surface staining. Conversely, tumors from S and T (not shown) mice exhibited retention of plasma membrane  $\beta$ -catenin immunoreactivity.

Table W: Detailed cluster analysis: Differential and statistically significant biological functions in clusters C1-C7

Cluster #	PS #	Reference	Pattern	Biology	Example genes
1	391	global	Up (A/M/T); down (S)	RNA post-transcriptional modification, cell cycle, DNA replication/recombination/repair, molecular transport, post-translational modification, cellular assembly and organization, cellular movement, cardiovascular system develop. and function, connective tissue develop. and function, cancer	cell cycle progression (Cdk4, Ctnnb1, Id1, Id3, Myc, PcnA, Tcf3), replication of DNA (Id1, Mcm2, Myc, Oro4l, PcnA, Polb, Set), checkpoint control (Bub3, Myc, Rae1, Smc111), invasion of mammary epithelial cells (Ezh2), recovery of ATP (Hspd1, Hspe1), hyperplasia of secretory structure (Cdk4, Ctnnb1, Ptpre, Sdc1), cell proliferation (Id1, Id3, Myc, PcnA)
2	663	global	up(A/M); down (S/T)	cell cycle, cellular response to therapeutics, cellular assembly and organization, molecular transport, connective tissue develop. and function, genetic disorder, gastrointestinal disease, cancer, Wnt-signaling pathway	contact growth inhibition of connective tissue cells (Metap2, Pcyox1), mitosis of tumor cells Mif, Plk1), cell cycle progression (Id2, Tp53). Checkpoint control (Mad2l1, Tp53), DNA modification (Apex1, Dnmt3a, Dnmt3b), infiltrating duct carcinoma (Esr1, Ing4), mitosis of tumor cells (Mif, Plk1), myotonic dystrophy (Dmpk, Znf9),

Table W: Detailed cluster analysis: Differential and statistically significant biological functions in clusters C1-C7

Cluster #	PS #	Reference	Pattern	Biology	Example genes
					Wnt-signaling (Csnk1d, Csnk1e, Lef1, Nlk, Tcf3, Tcf4, Wif1)
3	170	global	up (A/S); down (M/T)	cancer, cell death, cellular develop., cellular growth and proliferation, cell cycle	apoptosis of colon carcinoma cells (Tnfsf10), sarcoma (Ewsr1, Mdm2, Tnfsf10), hyperproliferation (Map2k7), survival (Mdm2, Nras, Tnfsf10), tumorigenesis (Ewsr1, Mdm2, Nras, Tnfsf10), fibroblast proliferation (Arid5b, E4f1, Map2k7, Mdm2, Nras), mitosis of embryonic cells (E4f1)
4	142	global	up (M/S); down (A/T)	cellular movement, hematological system develop. and function, immune response, hematological disease, immune and lymphatic system develop. and function, organ morphology, cell-to-cell signaling and interaction, cell death, molecular transport	Cell movement/chemotaxis (Alox5AP, C3, Ctsb, Cxcl12, Dcn, Fcgr3a, Fgfr1, Hif1a, Igf2, Itgb2, Lsp1, S100A9, Slp1), invasion of tumor cell lines (Cbx5, Ctsb, Cxcl12, Fstl1, Hif1a, Igf1, Igf2, Itgb2), chemotaxis/migration of leukocytes (C3, Cxcl12, Icam2, Itgb2, Lgals1, Lsp1, S100a9, Slp1), growth of tumor (Fgfr1, Hif1a, Igf2, Igfbp5, Igf1), invasion of tumor cell lines (Cbx5, Ctsb, Cxcl12, Fstl1, Hif1a, Igf2, Igf1, Itgb2)
5	432	global	up (S/T); down (A/M)	cell death, neurological disease, drug metabolism, endocrine system develop. and function, cancer, drug metabolism, lipid metabolism, gastrointestinal disease, organismal functions, organismal injury and abnormalities	gut epithelium differentiation (Chgb, Klf4, Klf4d, Sst), cell death/apoptosis of microglia (Btg1, Casp3, Casp9, Cx3cl1, Grin1, Myd88), uptake of prostaglandin E2 (Slco2a1), tumorigenesis of brain tumor (Nf1, Stat2), tumorigenesis of polyp (Asph, Smad4), aggregability of colon cancer cell lines (Cd82), cell spreading of colon cancer cell lines (Smad4), contact inhibition of colon cancer cell lines (Prkg1)
6	904	global	up (A/M); down (S/T)	cell proliferation, cell cycle progression and mitosis, DNA replication/recombination/repair, molecular transport, RNA post-transcriptional modification, post-translational modification, cellular growth and proliferation, connective tissue develop. and function, cancer, gastrointestinal disease, digestive system develop., and function	cell cycle progression/proliferation (Cdk4, Clu, Id2, Mki67, Magoh, Myc, Pcn, Tcf3, Tp53), tumor cell mitosis (Mif, Plk1). DNA excision repair (Apex1, Ddb1, Hmgb1, Polb), DNA methylation (Dnmt3a, Dnmt3b), accumulation of colonocytes (Clu, Myc),

Table W: Detailed cluster analysis: Differential and statistically significant biological functions in clusters C1-C7					
Cluster #	PS #	Reference	Pattern	Biology	Example genes
					tumorigenesis (Cd44, Cdk4, Ctnnb1, Esr1, Myc, Prkar1a, Tp53), Wnt-signalling pathway (Csnk1a1, Csnk1d, Csnk1e, Ctnnb1, Lef1, Myc, Nlk, Ppp2cb, Tcf3, Tcf4, Wif1)
7	361	global	up (S/T); down (A/M)	cell death, neurological disease, cancer, drug metabolism, embryonic development, endocrine system develop. and function, lipid metabolism, organismal injury and abnormalities, infectious disease, immune response, immunological disease, hematological disease, gastrointestinal disease, antigen presentation pathway	Antigen presentation (B2m, Cd74, H2-D1, HLA-DMA, HLA-DRB, Psmb8, Tap2), embryonic develop. (C3, Celsr1, Erbb3, Impk, Mcl1), infectious disease (B2m, Ifngr1, Irf1, Myd88, Nr3c1), mast cell chemotaxis (C3, Cx3c1), apoptosis of microglia (Btg1, Casp3, Cx3c1, Myd88), tumorigenesis of polyp (Asph, Smad4), transport of prostaglandin E2 (Slco2a1), quantity of colonocytes (Guca2a), gastrointestinal disease (Asph, Cd84, Smad4)

**[00127]** From the 1798 transcripts differentially expressed among the four mouse models of CRC, more than 70% (n=1265) distinguished  $Apc^{Min/+}$  and AOM tumors versus  $Smad3^{-/-}$  and  $Tgfb1^{-/-}$ ;  $Rag2^{-/-}$  tumors. If a random or equivalent degree of variance occurred among all classes, there would be far less overlap. The majority of this signature (~75%, n=904 features) derived from genes over-expressed in  $Apc^{Min/+}$  and AOM tumors relative to the  $Smad3^{-/-}$  and  $Tgfb1^{-/-}$ ;  $Rag2^{-/-}$  tumors (cluster C6). Cluster C6 was functionally enriched for genes linked to canonical WNT signaling (Table W). These included genes previously identified to be part of this pathway (Cd44, Myc, Stra6, Tcf1, Tcf4 [16], Id2, Lef1, Nkd1, Nlk, Twist [http://www.stanford.edu/~rnusse/], Catnb, Csnk1a1, Csnk1d, Csnk1e, Plat, Wif1) as well as genes that appear to be novel canonical WNT signaling targets (e.g. Cryl1, Expi, Ifitm3l, Pacsin2, Sox4 [16], Ets2, Hnrnpg, Hnrpa1, Id3, Kpnbp3, Pais, Pcn, Ranbp11, Rbbp4, Yes [17], Hdac2 [18]). Moreover, consistent with the over-expression of Myc in tumors from the  $Apc^{Min/+}$  and AOM models, enrichment of Myc targets such as Apex, Eef1d, Eif2a, Eif4e, Hsp90, Mif, Mitf, Npm1 [19] and the repression of Nibam [19] were detected.

**[00128]** Example: Results - Nuclear  $\beta$ -catenin expression distinguishes murine models

**[00129]** To establish a molecular basis for over-expression of canonical WNT target genes in  $Apc^{Min/+}$  and AOM tumors, immunohistochemistry was used to characterize the relative cellular

distribution of  $\beta$ -catenin. Tumors from  $Apc^{Min/+}$  and AOM mice exhibited strong nuclear  $\beta$ -catenin immunoreactivity and reduced membrane staining, whereas tumors from  $Smad3^{-/-}$  and  $Tgfb1^{-/-}; Rag2^{-/-}$  mice showed strong plasma membrane  $\beta$ -catenin staining with no nuclear accumulation (see inset). Additional tests to confirm the microarray results were also carried out using an independent set of C57BL/6  $Apc^{Min/+}$  colon tumor samples analyzed by qRT-PCR (FIG. 3) and immunohistochemistry. Colon tumors from five  $Apc^{Min/+}$  (M; nuclear  $\beta$ -catenin-positive) mice and four  $Smad3^{-/-}$  (S; nuclear  $\beta$ -catenin-negative) mice were harvested, and qRT-PCR was performed on nine genes that exhibited representative strong or subtle patterns in the microarray analyses. All nine patterns detected in the microarray set were validated by the qRT-PCR results. Alox12=Arachidonate 12-lipoxygenase; Casp6=Caspase 6; Matn2=Matrilin 2; Ptplb=Protein tyrosine phosphatase-like B; Sox21=SRY (sex determining region Y)-box 21; Spock2=Spard/osteonectin 2; Tesc=Tescalcin; Tpm2=Tropomysin 2; Wif1=WNT inhibitory factor; Stmn1=stathmin 1; Ptp4a2=phosphatase 4a2. In (a), \* p<0.05 and \*\* p<0.01. All expression patterns identified via microarray analysis were consistent with the qRT-PCR results (n=9 transcripts, chosen for their demonstration of a range of differential expression characteristics). *In situ* hybridization analyses using C57BL/6  $Apc^{Min/+}$  colon tumor samples also validated that Wif, Tesc, Spock2 and Casp6 were strongly expressed in dysplastic cells of the tumors (data not shown). At the protein level, immunohistochemical analyses confirmed relatively greater expression of the oncoprotein stathmin 1 in  $Apc^{Min/+}$  mice and tyrosine phosphatase 4a2 in  $Smad3^{-/-}$  mice.

**[00130]** Overall, cluster C6 genes (i.e. genes with greater up-regulation in tumors from  $Apc^{Min/+}$  and AOM models than in  $Smad3^{-/-}$  and  $Tgfb1^{-/-}; Rag2^{-/-}$ ) were consistent with increased tumor cell proliferation (e.g. Myc, Pnca), cytokinesis (e.g. Amot, Cxcl5), chromatin remodeling (e.g. Ets2, Hdac2, Set) as well as cell cycle progression and mitosis (e.g. Cdk1, Cdk4, Cul1, Plk1). Myc is up-regulated in all four mouse tumors models relative to normal colon tissue (see FIG. 7B). Biological processes showing increased transcription in tumors from the  $Smad3^{-/-}$  and  $Tgfb1^{-/-}; Rag2^{-/-}$  models (cluster C7) included immune and defense responses (e.g. Il18, Irf1, Myd88), endocytosis (e.g. Lrp1, Ldlr, Rac1), transport (e.g. Abca3, Slc22a5, Slc30a4), and oxidoreductase activity (e.g. Gcdh, Prdx6, Xdh) (Table W). Taken together, these transcriptional observations are both consistent with and extend the understanding of the histological features of the CRC models [7]. For example, while  $Apc^{Min/+}$  and AOM tumors are characterized by cytologic atypia (i.e. nuclear crowding, hyperchromasia, increased nucleus-to-cytoplasm ratios and minimal inflammation), tumors from  $Smad3^{-/-}$  and  $Tgfb1^{-/-}; Rag2^{-/-}$  mice show less overt dysplastic changes but exhibit a significant inflammatory component.

**[00131]** Multiperspective views of cancer sample transcriptional programs provide integrative insight into large-scale transcriptional patterns adopted by colon cancer and adenomatous tumor models. Murine colon tumor adenomas and human CRCs both show adoption and dysregulation of signatures tightly controlled during embryonic mouse colon development. The use of etiologically distinct mouse models of colon cancer allows for the identification of models that resemble different stages of embryonic mouse colon development and that are recapitulated by specific tumors types. With reference to FIG. 7A, all tumors exhibit large-scale activation of developmental patterns. Nuclear  $\beta$ -catenin-positive ( $Apc^{Min/+}$  and AOM) tumors map more strongly to early development stages during (more proliferative, less differentiated), whereas nuclear  $\beta$ -catenin-negative ( $Tgfb1^{-/-}$ ;  $Rag2^{-/-}$  and  $Smad3^{-/-}$ ) tumors map more strongly to later stages consistent with increased epithelial differentiation. With reference to FIG. 7B, overall representation of the relationship of mouse colon tumor models and human CRC to development and non-developmental expression patterns is shown. Gene expression clusters mapped to the progression of adenomatous and carcinomatous transformation are shown. Both mouse colon tumor models and human CRC share in the activation of embryonic colon expression, the repression of adult differentiation, and the loss of shared tumor suppressor genes. Many human CRCs also lack the expression of additional tumor suppressor programs and gain the expression of oncogenes that are not overexpressed during normal developmental morphogenesis.

**[00132]** Example: Results - Large-scale activation of the embryonic colon transcriptome in mouse tumor models.

**[00133]** The studies described herein were used to compare genes over-expressed in both colon tumors and embryonic mouse colon in order to provide insights into tumor programs important for fundamental aspects of tumor growth and regulation of differentiation. To identify genes and observe regulatory patterns that were shared or differed between colon tumors and embryonic development, a global quantitative referencing strategy was applied to both tumor and embryonic samples by calculating the relative expression of each gene as the ratio of its expression in any sample as that relative to its mean level in adult colon. From this adult baseline reference, genes over-expressed in the four mouse tumor models appeared strikingly similar. Moreover, the vast majority of genes over-expressed in tumors were also over-expressed in embryonic colon. If the fraction of fetal over-expressed genes from the entire microarray (5796 of 20393 features; 28.4%) was maintained at a similar occurrence frequency in the tumor over-expressed fraction (8804 of 20393), one would expect an overlap of 2502 transcripts ((8804/20393)\*28.4%). Rather, 4693 out of the 5796 fetal over-expressed transcripts were observed to be over-expressed in the 8804 tumor over-expressed genes (FIG. 4). The probability

calculated by Fisher's exact test is  $p < 1^{-300}$ , and thus represents highly significant over-representation of fetal genes among the tumor over-expressed genes. Similarly, genes under-expressed in developing colon were disproportionately underexpressed in tumors relative to normal adult colon (3282 of 3541;  $p < 1^{-300}$ ). Combining these results, approximately 85% of the developmentally regulated transcripts (7975 out of 9337 features) were recapitulated in tumor expression patterns relative to adult colon.

**[00134]** To explore the potential biological significance of genes over-expressed in both embryonic colon development and mouse tumors, K-means clustering was used to generate C8-C10 cluster patterns as shown in a hierarchical tree heatmap (Table X). Several sub-patterns were evident, some of which clearly separated  $Apc^{Min/+}$  and AOM from  $Smad3^{-/-}$  and  $Tgfb1^{-/-}$ ;  $Rag2^{-/-}$  tumors. One strong cluster, cluster C8, consisted of genes more strongly expressed in  $Apc^{Min/+}$  and AOM than  $Smad3^{-/-}$  and  $Tgfb1^{-/-}$ ;  $Rag2^{-/-}$  tumors. This group of genes represented a large fraction of all differences found between nuclear  $\beta$ -catenin-positive ( $Apc^{Min/+}$  and AOM) and negative ( $Smad3^{-/-}$  and  $Tgfb1^{-/-}$ ;  $Rag2^{-/-}$ ) tumors (~45%; 1636 out of 3592 features), as well as differences detected between early (i.e. E13.5-E15.5, ED) and late (E.16.5-E18.5, LD) embryonic colon developmental stages. Thus, the fraction of developmentally regulated genes that are more characteristic of the earlier stages of normal colon development (E13.5-E15.5), are clearly expressed at higher levels in nuclear  $\beta$ -catenin-positive tumors. This observation is illustrated by 750 transcripts selected solely for stronger expression in ED versus LD. The majority of these transcripts overlap with cluster C6 containing 230 features and illustrate the tendency of the earlier-expressed developmental genes to be more strongly expressed in  $Apc^{Min/+}$  and AOM mice. In addition, transcripts associated with increased differentiation and maturation, observed at later stages of colon development E16.5-E18.5 (e.g. *Klf4* [20], Crohn's disease-related *Slc22a5/Octn2* [21], *Slc30a4/Znt4* [22], *Sst* [23]), were expressed at higher levels by tumors from  $Smad3^{-/-}$  and  $Tgfb1^{-/-}$ ;  $Rag2^{-/-}$  mice.

**[00135]** All four murine tumor models exhibit reactivation of embryonic gene expression. The expression level of each gene in each sample was calculated relative to that in adult colon. Genes and samples were subjected to unsupervised hierarchical tree clustering for similarities among genes and tumors. Heatmap showed the relative behaviors of 20393 transcripts that passed basic signal quality filters with gene transcripts shown as separate rows and samples as separate columns. The majority of genes over-expressed in tumors are also over-expressed in embryonic colon; similarly, the genes under-expressed in tumors are under-expressed in embryonic colon. In addition, there are genes over-expressed in embryonic colon that are under-expressed in tumors and vice versa. The genes represented in FIG. 3 were divided into those over-expressed and

under-expressed in embryonic colon and in the tumors, respectively. Fisher's exact test was used to calculate expected overlaps between lists and confirmed significant over-representation of development-regulated signatures among the tumors ( $*p < 1^{-300}$ ,  $**p < 1.3^{-19}$ ,  $***p < 4^{-296}$ ,  $****p < 1^{-300}$ ). A heatmap was generated showing the behavior of a subset of the transcripts in 4a (n=4693 features) that were over-expressed in both embryonic colon and tumor samples (not shown). Table W includes a description of the genes associated with these clusters. Embryonic gene expression can be further refined into genes expressed differentially during early (ED; E13.5-15.5) and late (LD; E16.5-18.5) embryonic development. A heatmap was generated showing the relative behaviors of 750 transcripts that are highest-ranked for early versus late embryonic regulation (not shown). Overall, transcripts with the highest early embryonic expression were expressed at higher levels in nuclear  $\beta$ -catenin-positive tumors (A and M), whereas nuclear  $\beta$ -catenin negative tumors (S and T) were representative of later stages of embryonic development. Sample groups: ED, early development (E13.5-E15.5); LD, late development (E16.5-E18.5); A, AOM-induced; M, *Apc*<sup>Min/+</sup>; T, *Tgfb1*<sup>-/-</sup>; Rag2<sup>-/-</sup>; S, *Smad3*<sup>-/-</sup>.

Table X. Detailed cluster analysis: Differential and statistically significant biological functions in clusters C8-C10.

Cluster #	PS #	Reference	Biology	Example genes
8	1240	adult	RNA post-transcriptional modification, cell cycle, cellular assembly and organization, DNA replication/recombination/repair, cancer, molecular transport, protein trafficking and synthesis, cellular development, gastrointestinal disease, IGF-1 signaling, Wnt-signaling	mitosis (Ask, <i>Birc5</i> , <i>Bcra1</i> , <i>Cdc2</i> , <i>Cdk4</i> , <i>Chek1</i> , <i>Mad2l1</i> , <i>Mif</i> , <i>Plk1</i> ), DNA mismatch repair ( <i>Hgmb1</i> , <i>Msh2</i> , <i>Pcna</i> , <i>Rev1l</i> , <i>Xrcc5</i> ), cell transformation ( <i>Cdc37</i> , <i>Id2</i> , <i>Myc</i> ), cell proliferation ( <i>Ctnnb1</i> , <i>Pcna</i> , <i>Plat</i> , <i>Plk1</i> , <i>Rala</i> , <i>Top2a</i> ), colorectal cancer ( <i>Birc5</i> , <i>Bcra1</i> , <i>Cdc37</i> , <i>Myc</i> , <i>Top53</i> ), IGF-1 signaling ( <i>Igf1</i> , <i>Igfb4</i> , <i>Mapk1</i> , <i>Prkc</i> , <i>Ptpn11</i> ), Wnt signaling ( <i>Csnk1a1</i> , <i>Csnk2a1</i> , <i>Ctnnb1</i> , <i>Gs3kb</i> , <i>Myc</i> , <i>Nlk</i> , <i>Tcf3</i> , <i>Tcf4</i> )
9	1676	adult	protein synthesis, RNA-post transcriptional modification, cancer, connective tissue development and function, embryonic development, organ morphology, tissue morphology, cell-to-cell signaling and interaction, tissue development	Protein synthesis ( <i>Csf1</i> , <i>Eif5</i> , <i>Gadd45g</i> , <i>Itgb1</i> , <i>Sars</i> , <i>Tnf</i> , <i>Traf6</i> ), transformation ( <i>Ccnd1</i> ), formation of hepatoma cell line ( <i>Hras</i> , <i>Pin1</i> , <i>Shfm1</i> ), cell growth ( <i>Nrp1</i> , <i>Tnf</i> ), invasion of lymphoma cell line ( <i>Itgb1</i> , <i>Itgb2</i> ), proliferation of ovarian cancer cell lines ( <i>Fst</i> , <i>Hras</i> , <i>Itgb5</i> , <i>Sod2</i> , <i>Sparc</i> ), fibroblast cell cycle progression ( <i>Ccnf</i> , <i>E2f5</i> , <i>Hras</i> , <i>Map4</i> , <i>Rhoa</i> , <i>Skil</i> ), survival of epiblast ( <i>Dag1</i> , <i>Itgb1</i> ), cell adhesion ( <i>Icam1</i> , <i>Itgb1</i> , <i>Itgb2</i> , <i>Lu</i> , <i>Rhoa</i> , <i>Tnf</i> )
10	1051	Adult	cell cycle, cellular assembly and organization, DNA replication, Recombination/repair, cellular function and maintenance, cancer, cardiovascular system development	Cell cycle ( <i>Cdk2</i> , <i>Ccnd3</i> , <i>Siah</i> ), exocytosis ( <i>Nos3</i> , <i>Snap23</i> , <i>Stx6</i> , <i>Vamp2</i> ), Burkitt's lymphoma ( <i>Dmtf1</i> ), cell transformation ( <i>Mmp2</i> , <i>Pecam1</i> ), angiogenesis ( <i>Mdk</i> , <i>Nos3</i> ),

Table X: Detailed cluster analysis: Differential and statistically significant biological functions in clusters C8-C10.

Cluster #	PS #	Reference	Biology	Example genes
			and function, gene expression, immunological disease, digestive system development and function, activin/inhibin signaling	activation of RNA (Hrsp12, Rps6kb1), development of gastrointestinal tract (Pdgfra, Sptbn1), activin/inhibin signaling (Acvr2b, Bmpr1b, Inha, Map3k7, Mapk8, Tgfbr1)
8	1240	adult	RNA post-transcriptional modification, cell cycle, cellular assembly and organization, DNA replication/recombination/repair, cancer, molecular transport, protein trafficking and synthesis, cellular development, gastrointestinal disease, IGF-1 signaling, Wnt-signaling	mitosis (Ask, Birc5, Bcra1, Cdc2, Cdk4, Chek1, Mad2l1, Mif, Plk1), DNA mismatch repair (Hgmb1, Msh2, Pcna, Revl1, Xrcc5), cell transformation (Cdc37, Id2, Myc), cell proliferation (Ctnnb1, Pcna, Plat, Plk1, Rala, Top2a), colorectal cancer (Birc5, Bcra1, Cdc37, Myc, Top53), IGF-1 signaling (Igf1, Igfb4, Mapk1, Prkc, Ptpn11), Wnt signaling (Csnk1a1, Csnk2a1, Ctnnb1, Gs3kb, Myc, Nlk, Tcf3, Tcf4)
9	1676	adult	protein synthesis, RNA-post transcriptional modification, cancer, connective tissue development and function, embryonic development, organ morphology, tissue morphology, cell-to-cell signaling and interaction, tissue development	Protein synthesis (Csf1, Eif5, Gadd45g, Itgb1, Sars, Tnf, Traf6), transformation (Cnd1), formation of hepatoma cell line (Hras, Pin1, Shfm1), cell growth (Nrp1, Tnf), invasion of lymphoma cell line (Itgb1, Itgb2), proliferation of ovarian cancer cell lines (Fst, Hras, Igfb5, Sod2, Sparc), fibroblast cell cycle progression (Ccnf, E2f5, Hras, Map4, Rhoa, Skil), survival of epiblast (Dag1, Itgb1), cell adhesion (Icam1, Itgb1, Itgb2, Lu, Rhoa, Tnf)
10	1051	Adult	cell cycle, cellular assembly and organization, DNA replication, Recombination/repair, cellular function and maintenance, cancer, cardiovascular system development and function, gene expression, immunological disease, digestive system development and function, activin/inhibin signaling	Cell cycle (Cdk2, Cnd3, Siah), exocytosis (Nos3, Snap23, Stx6, Vamp2), Burkitt's lymphoma (Dmtf1), cell transformation (Mmp2, Pecam1), angiogenesis (Mdk, Nos3), activation of RNA (Hrsp12, Rps6kb1), development of gastrointestinal tract (Pdgfra, Sptbn1), activin/inhibin signaling (Acvr2b, Bmpr1b, Inha, Map3k7, Mapk8, Tgfbr1)

[00136] Example: Results - Human CRCs reactivate an embryonic gene signature

[00137] Since mouse tumors recapitulated developmental signatures irrespective of their etiology, it was asked whether a similar commitment to embryonic gene programming was shared by sporadic human CRCs. Tumor classification by microarray profiling is usually accomplished by referencing relative gene expression levels to the median value for each gene across a series of tumor samples. Using this “between-tumors median normalization” approach, as well as a gene filtering strategy that detects significantly regulated genes in at least 10% of the cases, led to the

identification of a set of 3285 probe sets corresponding to transcripts whose expression was highly varied between independent human tumor cases. There was striking heterogeneity of gene expression among 100 human CRCs. For example, Cluster 15 contained a set of genes (principally metallothionein genes) recently identified to be predictive of microsatellite instability [24, 25]. This analysis indicates that human CRCs have a greater level of complexity than the mouse colon tumors studied here. There was no correlation between these distinguishing clusters and the stage of the tumor (note the broad overlapping distributions of Dukes stages A-D across these different clusters). However, as shown in Table Y, gene ontology and network analysis of the individual gene clusters (C11-C17) that were differentially active in subgroups of the tumors, map to genes highly associated with a diverse set of biological functions, including lipid metabolism, digestive tract development and function, immune response and cancer.

**[00138]** Human CRCs exhibit gene expression profile complexity consistent with significant tumor subclasses. Genes potentially able to distinguish cancer subtypes were identified from Affymetrix HG-U133 plus2 Genechip expression profiles by filtering for 3285 probe sets that were top-ranked by raw expression and their differential regulation in at least 10 out of 100 human colorectal cancer tumors. Coordinately regulated transcripts and similarly behaving samples were identified via hierarchical tree clustering. Seven different gene clusters (C11-17) were identified that distinguished 10 or more tumors from the other tumors. Gene clusters were found to be highly enriched for gene functions listed in Table Y. Data were processed using robust microarray analysis (RMA) with expression value ratios depicted as the relative expression per probe set in each sample relative to the median of its expression across the 100 CRCs. A striking heterogeneity of gene expression was observed, including metallothionein genes in cluster C15 previously shown to be predictive of microsatellite instability, and C17 represented by 734 probe sets rich in genes associated with extracellular matrix and connective tissue, tumor invasion and malignancy.

Table Y: Detailed cluster analysis: Differential and statistically significant biological functions in clusters C11-C17.

Cluster #	PS #	Reference	Biology	Example genes
11	167	global	molecular transport, protein trafficking lipid metabolism, small molecule biochemistry, cardiovascular system development, dermatological diseases and conditions, organismal development, organismal injury and abnormalities, cancer, digestive system develop, and function	protein excretion (BF, EDNRA, KL), corticosteroid/daunorubicin transport (ABCB1), modification of cholesterol (ABCB1, SULT2B1), neovascularization of animal (TNFRSF6B, TNFSF11), angiogenesis of granulation tissue (PTGES), blister formation (COL17A1, FRAS1), develop. of enteroendocrine calls (NEUROD1), crypt size (FOLR1), connective tissue formation (EDNRA, IL7, MSX2, PTGES, WT1), division of

Table Y: Detailed cluster analysis: Differential and statistically significant biological functions in clusters C11-C17.

Cluster #	PS #	Reference	Biology	Example genes
				mesenchymal cells (BMP7)
12	762	global	RNA post-translational modification, gene expression, cancer, renal and urological disease, RNA trafficking, embryonic development, cell-to-cell signaling and interaction, estrogen receptor signaling, EFG signaling, PI3K/AKT signaling	processing of RNA (HNRPA2B1, HNRPD, HNRPH1, PRPF4B, RBM6, RBPMS, SFPQ, SFRS3, SFRS4, SNRPA1, U2AF1, ZNF638), transactivation of glucocorticoid/thyroid hormone response element (FOXO1A NCOR1, NR3C1, RORA), tumorigenesis (CD44, CTNNB1, EGFR, NF1, PRKAR1A, PTEN, THBS1), adhesion of tumor cells (CD44, CD47, EGFR, PTK2, THBS1), juvenile/colonic polyposis (CTNBB1, PTEN, SMAD4), IGF1-signaling (CTNBB1, FOXO1A, PTEN, SOS2)
13	213	global	cell morphology, cellular development, hematological disease, generic disorder, embryonic development, cellular assembly and organization, hair and skin development and function, cardiovascular system development and function, cancer, digestive system develop and function	conversion of epithelial cells (ATOH1, DMBT1, FOS), depolarization of cells (CACNA1C, FOS, NTS), development of Goblet/Paneth/enteroendocrine cells (ATOH1), hematological disease (HBA1, HBA2, HBB, GTF), partington syndrome (ARX), muchopolysaccharidosis (HYAL1), Pfeiffer's syndrome (FGFR2), retinoic acid synthesis (ALDH1A1, ALDH1A2), adenoma inflammation (TFF1), density of connective tissue (MIA, TNFRSF11B)
14	161	global	cancer, cellular movement, skeletal and muscular disorders, immune response, gastrointestinal disease lipid metabolism, reproductive system disease, small molecule biochemistry, digestive system development and function, tissue development	migration/invasion of tumor cell lines (CDKN2A, CST6, DPP4, KITLG, LAMA3, LCK, MDK, SERPINB5, TFF2, TGFA), tumorigenesis of intestinal polyp (ASPH), proliferation of tumor cell lines (APRIN, CDKN2A, CST6, IMP3, LITLG, PIWIL1, SLP1, TGFA), cytotoxic reaction (CDKN2A, LCK), invasion of tumor cell lines (CDKN2A, CST6, DPP4, SERPINB5, TFF2, TGFA), tumorigenesis of small intestine (PLA2G4A), size/tumorigenesis of polyp (ASPH, CDKN2A, TGFA)
15	366	global	drug metabolism, endocrine system development and function, small molecule biochemistry, lipid metabolism, molecular transport, gene expression, cell death, cell morphology, cancer, gastrointestinal disease, digestive system development and function, tissue development	Steroid metabolism (AKR1C2, CYP3A5, UGT2B15, UGT2B17), conversion of progesterone (AKR1C3, HSD3B2), modification of dopamine (SULT1A3, XDH), oxidation of norepinephrine (MAOA), drug transport (ANCB1, ABCG2), transport of fludarabine (SLC28A2), hydrocortisone uptake (ABCB1) formation of aberrant crypt foci (NR5A2, PTGER4), cell death of enteroendocrine cells (GCG, PYY), growth of crypt cells (NKX2, NKX3)
16	221	global	cardiovascular system development and function, cellular movement,	Cell movement/proliferation of endothelial cells (ADIPOQ, CXCL12,

Table Y: Detailed cluster analysis: Differential and statistically significant biological functions in clusters C11-C17.

Cluster #	PS #	Reference	Biology	Example genes
			hematological system development and function, immune response, cancer, neurological disease, carbohydrate metabolism, organismal development, digestive system development and function, tissue development	ENPP2, FGF13, HGF, HHEX, MYH11, PTN), formation of endothelial tube and blood vessel (ADAMTS1, ANGPTL1, CCL11, CXCL12, ENPP2, F13A1, HGF, MEF2C, MYH11, PTEN), cell movement of cancer cells (CXCL12, CD36, HGF, IGF1, L1CAM, SFRP1, PTN), tumorigenesis (AGTR1, CNN1, ENPP2, FGF7, HGF, IGF1, KIT, L1CAM). Hirschprung disease (EDNRB, L1CAM)

[00139] To evaluate if similar sets of genes are systematically activated or repressed in human CRC, as in the mouse colon tumors, two procedures were undertaken to align the data. First, gene expression values for the mouse and human tumors were separately normalized and referenced relative to their respective normal adult colon controls; second, mouse and human gene identifiers were reduced to a single ortholog gene identifier. The latter is a somewhat complex procedure that requires identifying microarray probes from each platform that can be mapped to a single gene ortholog and undertaking a procedure to aggregate redundant probes within a platform (see Methods). This approach allowed the identification of 8621 gene transcripts on the HG-U133 plus2 and Vanderbilt NIA 20K cDNA arrays for which relative expression values could be mapped for nearly all mouse and human samples. A clustering-based assessment of expression across the whole mouse-human ortholog gene set identified a large number of transcripts behaving similarly across colon tumors, many irrespective, but some respective of species. Notably, the great majority of genes over-expressed in all tumors were also over-expressed during colon development. To evaluate the statistical significance of this pattern, a Venn overlap filtering strategy and Fisher's exact test analysis were used. Approximately 50% of the 2212 ortholog genes over-expressed in at least 10% of the human cancers relative to adult colon were also over-expressed in developing colon. If there was not a selection for developmental genes among the tumor-overexpressed, the expected overlap would be  $(2718/8621)*2212=697$  transcripts. Using Fisher's exact test for the significance of the increased overlap of 1080 versus 697 transcripts is  $p<1e-300$ . Similarly, genes under-expressed in mouse colon development and human CRCs also strongly overlapped (FIG. 5; 431 of 737,  $p<1e-76$ ). This result is significantly greater than the 8-19% of genes that were estimated to be over-expressed in human colon tumors and fetal gut morphogenesis based upon a computational extrapolation of SAGE data [26]. Thus, the findings not only confirm but also significantly expand and experimentally validate the previously suggested recapitulation of embryonic signatures by human CRCs.

**[00140]** All overlaps between tumor expression and development were pooled to form a set of 2116 ortholog gene transcripts. This was subjected to hierarchical tree and K-means clustering to define six expression clusters, C18-C23 (Table Z).

**[00141]** Both human CRCs and mouse colon tumors reactivate an embryonic gene signature. When human and murine tumors are compared, they both broadly re-express an embryonic gene expression pattern. Gene expression profiles from the mouse tumor models and human CRC samples were combined into a single non-redundant gene ortholog genome table structure and subjected to comparative profile analysis. Informative probe-sets from human and mouse platforms were selected, mapped to corresponding ortholog genes, and used to populate a table in which normalized expression for each gene is relative to normal adult colon. A heatmap plot for all cross-species gene orthologs both present and successfully measured on both the Affymetrix Hg-U133 and Vanderbilt Mouse NIA 20K microarrays (n=8621 features) suggested that a large number of human CRC signatures exhibit similar behaviors in the mouse tumors and during embryonic mouse colon development. Based on these results, four separate gene lists were generated with criteria of over- or under-expression in development or over-expression or under-expression in human CRCs (2718, 2365, 2212, and 737, respectively). Genes over-expressed and under-expressed in embryonic mouse colon and human CRCs were found to be over-represented as determined by Fisher's exact test analysis (\* $p < 7 \times 10^{-88}$ , \*\* $p < 1 \times 10^{-76}$ , \*\*\* $p < 5 \times 10^{-4}$ , \*\*\*\* $p < 1 \times 10^{-76}$ ). A heatmap plot of all genes co-regulated in human CRCs and during early (ED) and late (LD) mouse embryonic colon development (n=2216 features). Six predominant clusters (C18-C23) characterize the transcriptional relationship between human CRC and mouse colon tumor models and embryonic development. Two clusters (C20 and 21) primarily distinguish human CRCs from murine tumors (A, M, S and T). For example, CRC up-regulated transcripts that are either developmentally up- or down-regulated are represented by clusters C22 (n=860 features) and clusters C21/C23 (n=142 features), respectively. Conversely, CRC down-regulated transcripts that are either down- or up-regulated during development are shown in clusters C18/C19 (n=258 features) and cluster C20 (n=42 features), respectively. Interestingly, while ~80% and ~60% of genes up- and down-regulated in both human CRCs and mouse development were also up- and down-regulated in tumors from the various mouse models, several clusters provide very interesting exceptions: cluster C20 are genes down-regulated in human CRCs that are routinely over-expressed in mouse tumors and development; cluster C21 are genes robustly expressed in human CRC that are rarely expressed in embryonic colon or murine tumors.

Table Z: Detailed cluster analysis: Differential and statistically significant biological functions in clusters C18-C23.

Cluster #	PS #	Reference	Pattern	Biology	Example genes
18	258	adult colon	down (D); down (CRC); down (A/M/S/T)	lipid metabolism, molecular transport, cell death, cancer, cellular movement, drug metabolism, lipid metabolism, digestive system develop. and function, small molecule biochemistry, endocrine system develop. and function, neurological disease	gut epithelium differentiation (CA4, CA12, CBR1, CHGB, KLF4, KLF9, MCOLN2, SST, TFF3), apoptosis/cell death (CYCS, G3N, KITLG, SST, TFF3, TGFA), cytolysis/crypt damage (ABCB1, KLKR1, PTGER4), formation of aberrant crypt foci (NR5A2, PTGER4), drug tra
19	42	adult colon	up(D); down (CRC); down (A/M/S/T)	digestive system develop. and function, cancer, small molecule biochemistry, reproductive system develop. and function, organ morphology	colon and midgut develop. (EDNRB), gastrointestinal stromal tumor (KIT), apoptosis of mesothelioma cells (KIT), melanocyte differentiation (EDNRB, KIT), inhibition and morphology of melanoma cells (HSPE, LSP1), adhesion of lymphoma cells (HSPE)
20	91	adult colon	up (D); down (CRC); up (A/M/S/T)	cell death, hematological disease, immunological disease, cell-to-cell signaling and interaction, hematological system development and function, immune response, cancer cell morphology, tissue development, gastrointestinal disease	apoptosis of colon carcinoma, cells (BCL2), apoptosis of lymphoma cell lines (BCL2, IGFBP4, MAP4K1, PDGFRA), cell-cell contacts of endothelial cells (STAB1). Lymphocyte quantity (BCL2, CCR7, CD28, ITGB7, ITK, MUC1, WNT4), proliferation of lymphocytes (CD2)
21	313	adult colon	down (D); up (CRC); down (A/M/S/T)	cell death, nervous system development and function, drug metabolism, small molecule biochemistry, cancer, cell cycle, cellular growth and proliferation, tissue development	melanocyte survival (RB1), proliferation of neuronal progenitor cells (ATM, VEGF), heparin binding (PRNP, TNC, VEGF), dopamine formation (TH), drug resistance (ABCC1), quantity of tumor cell lines (LIF, PIK3R1, RB1, TIMP3, VEGF) transformation (FOXO3A)
22	860	adult colon	up (D); up (CRC); up (A/M/S/T)	cell proliferation, cancer, DNA replication/recombination/repair, cell cycle progression and mitosis, cellular movement, connective tissue develop. and function, tumor morphology; purine and pyrimidine metabolism, folate metabolism	cell transformation (Myc). Mismatch repair (HMGB1, MSH2, MSH6, PCNA), arrest in mitosis (BIRC5, BUB1B, CDC2, CHEK1, CSE1L, MAD2L1, MIF, PLK1), migration/cytokinesis (ANLN, CDC42, FN1, ITGB5, MSF, SPARC, TOP2A), survival (AKT2, APEX 1, BIRC5), gastric carcinoma progression (COL1A1, FUS), folate metabolism (MTHFD1, MTHFD2)
23	142	adult colon	down (D); up (CRC);	connective tissue develop. and function, cell-to-cell signaling	cell transformation (ESR1, SRC), basal membrane

Table Z: Detailed cluster analysis: Differential and statistically significant biological functions in clusters C18-C23.

Cluster #	PS #	Reference	Pattern	Biology	Example genes
			up (A/M/S/T)	development disorder, organismal injury and abnormalities, tumor morphology, hematological system develop. And function, immune and lymphatic system develop, and function, cancer	disruption (MMP7) cell extension (ATF3, CD82, IL6, SRC) contact growth inhibition (JUN, IL6), osteocyte differentiation (IL6, JUN, SMAD6, SRC), cell cycle progression (ESR1, IER3, IL6, PSEN2), ERK/MAPK sig

**[00142]** These clusters provide an impressive partitioning of groups of genes associated with different biological functions critical for colon development, maturation and oncogenesis. Cluster C22 (860 transcripts of genes strongly expressed both developmentally and across all tumors) is highly enriched with genes associated with cell cycle progression, replication, cancer, tumor morphology and cellular movement. Cluster C18 (258 transcripts down-regulated in mouse and human tumors, as well as in development) is highly enriched in genes associated with digestive tract function, biochemical and lipid metabolism. This cluster is clearly composed of genes associated with the mature GI tract. Thus, as opposed to recapitulating developmental gene activation, the cluster C18 pattern indicates a corresponding arrest of differentiation in both mouse and human tumors. Cluster C23 (142 transcripts over-expressed in all mouse models and human CRC, but with low expression in development) maps to genes highly associated with the disruption of basement membranes, invasion and cell cycle progression, as well as altered transcriptional control. Cluster C21 (313 transcripts in which human tumors somewhat variably express a set of genes that are rarely expressed by the mouse tumors) is remarkable for its composition of genes associated with cell cycle proliferation, tissue disruption and angiogenesis. Thus, while categorically quite similar to cluster C23, the genes in cluster C21 represent a separately regulated module that is enriched for genes associated with invasion. Clusters C21 and C23 reveal sets of genes likely involved in tumor progression. Cluster C22 (with genes over-expressed in all mouse and human tumors and strongly expressed in embryonic colon) represents a group of genes highly correlated with transformation. The top-ranked transcription factor present in this cluster, with regulation independent of  $\beta$ -catenin localization, is Myc/MYC (**FIG. 6B**). Although Myc was lower in expression in the  $\text{Smad3}^{-/-}$  tumors compared to tumors from the other three models, it was elevated in all four models relative to normal adult colon. Myc/MYC was over-expressed in all mouse and human tumors as well as in development. This contrasts with Sox4, which is unaltered in expression in the  $\text{Smad3}^{-/-}$  and  $\text{Tgfb1}^{-/-}$ ;  $\text{Rag2}^{-/-}$  tumors but is up-regulated in AOM and  $\text{Apc}^{\text{Min}+/+}$  tumors relative to normal adult colon (**FIG. 6B**). Myc/MYC

over-expression may be independent of nuclear  $\beta$ -catenin status. Increased Myc/MYC expression may reflect both activation of canonical Wnt signaling, as it is a target of nuclear  $\beta$ -catenin/TCF [27], and deregulation of TGF $\beta$  signaling, as TGF $\beta$ 1 is known to repress Myc/MYC [28-30].

These observations suggest a fundamental role for Myc/MYC in colonic neoplasia.

**[00143]** Example: Discussion

**[00144]** Numerous mouse models of intestinal neoplasia have been developed, each with unique characteristics. The models constructed to date, however, do not fully represent the complexity of human CRCs principally because most are unigenic in origin and produce primarily adenomas and early stage cancers. Although models like *Apc*<sup>Min/+</sup> show molecular similarities to human CRCs, such as initiation of adenoma formation by inactivation of *Apc*, little is known about the molecular similarities of tumors from the different mouse models. It is also unknown how such common and perhaps large-scale molecular changes in mouse models relate to the molecular programming of human CRC. To shed light on the underlying molecular changes in tumors from mouse models and human CRC, the relationship was assessed at the molecular level of four widely used, but genetically distinct, mouse models that develop colon tumors. A subsequent analysis of the models in the context of embryonic mouse colon development was also undertaken. Finally, to identify consensus species-independent cancer signatures that may define gene expression changes common to all CRCs, relevant mouse model signatures were projected onto a large set of human primary CRCs of varied histopathology and stage.

**[00145]** Example: Discussion - Differential canonical WNT signaling activity discriminates two major classes of mouse models of CRC with distinct molecular characteristics

**[00146]** Tumors from mouse models of CRC exhibit significant phenotypic diversity [6], and therefore were expected to exhibit differential gene expression patterns. Using a combination of inter-model and normal adult gene expression level referencing, the analysis of tumors from mouse models of CRC has revealed a low complexity between models and strains, and has identified common and unique transcriptional patterns associated with a variety of biological processes and pathway-associated activities. The results demonstrate an imbalance between proliferation and differentiation with nuclear  $\beta$ -catenin-positive tumors being more proliferative, less differentiated and with lower immunogenic characteristics than tumors from nuclear  $\beta$ -catenin-negative tumors. Mouse tumors characterized by signatures of relative up-regulation of genes associated with cell cycle progression also showed increased canonical WNT signaling activity (*Apc*<sup>Min/+</sup> and AOM). Tumors from mouse models not showing canonical WNT signaling pathway activation (*Smad3*<sup>-/-</sup> and *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup>) were characterized by up-regulation of genes associated with inflammatory and innate immunological responses, and intestinal epithelial cell

differentiation. Recent studies have indicated that chronic inflammation caused either by infection with Helicobacter pylori [31] or Helicobacter hepaticus [13] is a prerequisite for intestinal tumor development in *Smad3*<sup>-/-</sup> and *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup> mice, respectively.

**[00147]** The activation of canonical WNT signaling in AOM tumors was identified using a between-tumor global median normalization to gene expression data. However, when tumor sample expression was referenced to that of normal adult intestinal tissue, many more genes are up-regulated, including developmental genes that are not dependent on nuclear  $\beta$ -catenin. That canonical WNT signaling-related genes are altered similarly in both AOM and *Apc*<sup>Min/+</sup> tumors suggest biological similarities between the two models. In addition, the relatively consistent programming within the AOM model also emphasizes its value for examining the more complicated genetics that result in strain-specific sensitivity to environmental agents that induce cancer.

**[00148]** Activation of canonical WNT signaling leads to nuclear translocation of  $\beta$ -catenin and, through its interaction with LEF/TCF, the regulation of genes relevant to embryonic development and proliferation [16], as well as stem cell self-renewal [32]. Consequently, the activated canonical WNT signaling observed in *Apc*<sup>Min/+</sup> and AOM models suggests that tumors may arise as a consequence of proliferation of the stem cell or “transient amplifying” compartment. In the colonic crypt, loss of TCF4 [33] or DKK1 over-expression [34] promotes loss of stem cells, suggesting that canonical WNT signaling is required for the maintenance of the intestinal stem cell compartment [33-35]. Conversely, increased nuclear  $\beta$ -catenin/TCF4 activity imposes a crypt progenitor phenotype on tumor cells [17]. In this study, transcriptional activation of the canonical WNT signaling pathway was identified in tumors from *Apc*<sup>Min/+</sup> and AOM mice. This was confirmed by immunohistochemistry.

**[00149]** In colon tumors and perhaps intestinal stem cells, activation of canonical WNT signaling promotes a hyperproliferative state. Proliferation-related characteristics of nuclear  $\beta$ -catenin-positive tumors include increased expression of CCND1, *MYC*, *PCNA* [17], and *Sox4* [16]. These genes were also identified as a component of the nuclear- $\beta$ -catenin-positive signatures. In turn, increased *MYC* decreases intestinal cell differentiation by binding to and repressing the *Cdkn1a* (coding for p21<sup>CIP1/WAF1</sup>) promoter [36], the gene encoding Wnt-inhibitory factor *Wif1*, the gene encoding the negative regulator of WNT Naked1 [37], and the gene encoding the Tak1/Nemo-like kinase, *Nlk* [38]. *Wif1* displays a graded expression in colonic tissue, with higher expression in the stem cell compartments and lower expression in the more differentiated cells at the luminal surface, suggesting that *Wif1* may contribute to stem cell pool maintenance independent of WNT signaling inhibition. [39].

**[00150]** Canonical WNT signaling not only governs intestinal cell proliferation, but also cell differentiation and cell positioning along the crypt-lumen axis of epithelial differentiation. Increased canonical WNT signaling activity enhances MATH1-mediated amplification of the gut secretory lineages [40]. Canonical WNT signaling also influences cell positioning by regulating the gradient of EPHB2/EPHB3 and EPHB1 ligand expression [41, 42]. Together, the data suggest a complex imbalance of crypt homeostasis due to enhanced canonical WNT activity.

**[00151]** The results indicate that tumors arising in response to abnormal TGF $\beta$ 1/SMAD signaling [14, 43] are similar to one another in their specific gene signatures and broadly distinct from those with activated canonical WNT signaling by their absence of nuclear  $\beta$ -catenin. Unique to the dysregulated TGF $\beta$ 1/SMAD4 signaling models is the strong signature of an immunologically altered state, with up-regulation of genes determining immune and defense responses such as *Il18*, *Irf1* and mucin pathway-associated genes. Again, these tumors are usually characterized by a strong inflammatory component when evaluated histopathologically, even in the absence of T- and B-cells such as in the *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup> background.

**[00152]** The microarray patterns of gene expression for AOM and *Apc*<sup>Min/+</sup> tumors are mirror images of those for *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup> tumors. It is perhaps not surprising that combining these two transcriptional programs results in increased number and invasiveness of colonic tumors as recently reported for *Apc*<sup>Min/+</sup> mice crossed to *Smad3*<sup>-/-</sup> mice [44]. Moreover, combined activation of canonical WNT signaling and inhibition of TGF $\beta$  signaling also results in more advanced intestinal tumors (*Apc*<sup>delta716/+</sup>; *Smad4*<sup>+/+</sup> mice [45] and intestine-specific deletion of the type II TGF $\beta$  receptor in *Apc*<sup>1638N/wt</sup> mice [46].

**[00153]** The findings that shared over-expressed signatures are identifiable in all four mouse models of CRC, that are also representative of the majority of embryonic colonic over-expressed signatures, and that these signatures are also present in all human CRCs, suggest that colon tumors may arise independently of canonical WNT signaling status. A likely candidate to impart this oncogenic signaling is Myc, which is an embryonic up-regulated transcript that is also upregulated in all human CRCs and mouse tumor models independently of nuclear  $\beta$ -catenin status.

**[00154]** Example – Discussion - Embryology provides insight into the biology of mouse and human colon tumors

**[00155]** It has long been suggested that cancer represents a reversion to an embryonic state, partly based upon the observation that several oncofetal antigens are diagnostic for some tumors [47, 48]. To assess the embryology-related aspects of tumorigenesis and tumor progression in CRC, the transcriptomes of normal mouse colon development and models of CRC were analyzed

and compared. The data show that developmentally regulated genes represent ~56% of mouse tumor signatures, and that the tumor signatures from the four mouse models recapitulate ~85% of developmentally regulated genes.

**[00156]** There are at least two regulatory programs that determine the expression of developmental genes by mouse tumors (FIG. 7). The simpler program is evident by the over-expression of the earliest genes of colon development by the nuclear  $\beta$ -catenin-positive models. The more subtle program is that which is also shared by nuclear  $\beta$ -catenin-negative models that are also highly enriched in developmentally expressed genes despite the fact that they lack activation of canonical WNT signaling. Genes found within this signature are consistent with those present in the colon at later developmental stages (E16.5-E18.5).

**[00157]** How do genes tightly regulated during mouse colon development become activated in colon tumors? While activated canonical WNT signaling imparts a strong influence, its absence in *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup> and *Smad3* <sup>-/-</sup> tumors, as determined by the absence of nuclear  $\beta$ -catenin, did not prevent the large scale activation of developmental/embryonic gene expression. One mechanism may be through epigenetic alterations. In human CRCs, these types of alterations in gene expression programs [49], suggest a link between cellular homeostasis and tumorigenesis. The recruitment of histone acetyltransferases and histone deacetylases (HDACs) are key steps in the regulation of cell proliferation and differentiation during normal development and carcinogenesis [50]. Induction of *Hdac2* expression occurs in 82% of human CRCs as well as in tumors from *Apc*<sup>Min/+</sup> mice [18]. Alternatively, common regulatory controls may operate in parallel growth and differentiation/anti-differentiation pathways such that a single or small subset of regulators, such as MYC or one or more miRNAs, may be responsible for the control of multiple pathways. Indeed, consistent with the observation of nuclear beta catenin independent activation of Myc in all mouse models and across the board for human CRC, deletion of Myc has recently been demonstrated to completely abrogate nuclear beta catenin-driven small bowel oncogenesis in mouse models [51].

**[00158]** Example: Discussion - Comparative analysis reveals underlying development-related signatures in human CRCs

**[00159]** As shown in FIG. 6, considerable and intriguing heterogeneity of human CRC is observed among genes highly relevant for differential malignant behavior. However, employing the between-tumors normalization and referencing strategies prevent the detection of gene expression patterns that are shared between tumors. Using the adult normal colon as a reference, as shown in FIG. 5, a large fraction of differential gene expression relative to adult colon could be demonstrated that recapitulated developmental gene expression by virtue of both activating

embryonic colon gene expression and failing to express genes associated with normal colon maturation. Within these developmentally regulated gene sets, the analyses revealed little evidence of CRC subsets, including those suggestive of nuclear beta catenin negative tumors that might approximate the *Smad3*<sup>-/-</sup> and *Tgfb1*<sup>-/-</sup>; *Rag2*<sup>-/-</sup> signature. The inability to identify distinct subclasses with respect to developmental genes in the human CRCs is perhaps not surprising in that over 80% of MSI+ CRCs from HNPCC families exhibit nuclear  $\beta$ -catenin [52]. In addition, within the developmental genes, little evidence was apparent for signatures related to microsatellite-unstable tumors (MSI+), often associated with HNPCC although perhaps some of this type of signature was apparent in the median normalized depiction of the tumors as highlighted in **FIG. 6**.

**[00160]** This report constitutes a comprehensive molecular evaluation and comparison of mouse and human colon tumor gene expression profiles. The ability to compare tumor gene expression profiles between mouse and human tumors has been improved by using a referencing strategy in which gene expression levels in the tumor samples are analyzed in relation to gene expression in corresponding normal colon epithelium. This approach has revealed that gene expression patterns are both shared and distinct between mouse models and human CRCs. The present study actually demonstrates the magnitude of the similarity between tumors and embryonic gene expression.

**[00161]** Finally, the results suggest that comparisons made between mouse tumor models, developing embryonic tissues, and human CRCs provides a powerful biological framework from which to observe shared and unique genetic programs associated with human cancer. While ortholog-gene based analyses have been used previously to obtain direct comparison of the molecular features of mouse and human hepatocellular carcinomas [55], the results provide striking support for the hypothesis that cancer represents a subversion of normal embryonic development. By inclusion of detailed mouse embryonic and developmental profile information, the results have revealed critical similarities and differences between the mouse and human tumors that are particularly revealing of oncogenic and tumor suppressor programs, some genes from which should be useful for development of diagnostic biomarkers and identification of therapeutic targets and pathways.

**[00162]** Example: Conclusions

**[00163]** Cross-species and multi-model comparisons of gene expression in colon cancer versus that of normal embryonic development provide an integrated and versatile framework from which to recognize transcriptional patterns associated with oncogenesis and a general method for pattern-specific identification of biomarkers and therapeutic targets. A pronounced feature of all

CRCs is a simultaneous exploitation and subversion of normal organogenesis. Irrespective of nuclear  $\beta$ -catenin status, all mouse and human colonic tumors exhibited an embryonic expression pattern and Myc/ MYC overexpression suggesting its fundamental oncogenic role in colonic neoplasia. Compared to the mouse model tumors, human CRC adenocarcinomas both gained and lost additional non-developmental gene expression modules whose specific targeting may lead to both improved mouse models and novel combinatorial therapeutics.

**[00164]** Throughout this document, various references are mentioned. All such references are incorporated herein by reference, including the references set forth in the following list:

#### REFERENCES

1. Heath JP: Epithelial cell migration in the intestine. *Cell Biol Int* 1996, 20(2):139-146.
2. Sancho E, Batlle E, Clevers H: Signaling pathways in intestinal development and cancer. *Annu Rev Cell Dev Biol* 2004, 20:695-723.
3. Brabletz T, Hlubek F, Spaderna S, Schmalhofer O, Hiendlmeyer E, Jung A, Kirchner T: Invasion and metastasis in colorectal cancer: epithelial-mesenchymal transition, mesenchymal-epithelial transition, stem cells and beta-catenin. *Cells Tissues Organs* 2005, 179(1-2):56-65.
4. Helm J, Enkemann SA, Coppola D, Barthel JS, Kelley ST, Yeatman TJ: Dedifferentiation precedes invasion in the progression from Barrett's metaplasia to esophageal adenocarcinoma. *Clin Cancer Res* 2005, 11(7):2478-2485.
5. Kaihara T, Kusaka T, Nishi M, Kawamata H, Imura J, Kitajima K, Itoh-Minami R, Aoyama N, Kasuga M, Oda Y *et al*: Dedifferentiation and decreased expression of adhesion molecules, E-cadherin and ZO-1, in colorectal cancer are closely related to liver metastasis. *J Exp Clin Cancer Res* 2003, 22(1):117-123.
6. Boivin GP, Groden J: Mouse Models of Intestinal Cancer. *Comp Med* 2004, 54(1):15-18.
7. Boivin GP, Washington K, Yang K, Ward JM, Pretlow TP, Russell R, Besselsen DG, Godfrey VL, Doetschman T, Dove WF *et al*: Pathology of mouse models of intestinal cancer: consensus report and recommendations. *Gastroenterology* 2003, 124(3):762-777.
8. Moser AR, Pitot HC, Dove WF: A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science* 1990, 247(4940):322-324.
9. Logan CY, Nusse R: The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* 2004, 20:781-810.
10. Oving IM, Clevers HC: Molecular causes of colon cancer. *Eur J Clin Invest* 2002, 32(6):448-457.
11. Bissahoyo A, Pearsall RS, Hanlon K, Amann V, Hicks D, Godfrey VL, Threadgill DW: Azoxymethane Is a Genetic Background-Dependent Colorectal Tumor Initiator and Promoter in Mice: Effects of Dose, Route, and Diet. *Toxicol Sci* 2005.

12. Grady WM, Myeroff LL, Swinler SE, Rajput A, Thiagalingam S, Lutterbaugh JD, Neumann A, Brattain MG, Chang J, Kim SJ *et al*: Mutational inactivation of transforming growth factor beta receptor type II in microsatellite stable colon cancers. *Cancer Res* 1999, 59(2):320-324.
13. Engle SJ, Ormsby I, Pawlowski S, Boivin GP, Croft J, Balish E, Doetschman T: Elimination of colon cancer in germ-free transforming growth factor beta 1-deficient mice. *Cancer Res* 2002, 62(22):6362-6366.
14. Zhu Y, Richardson JA, Parada LF, Graff JM: Smad3 mutant mice develop metastatic colorectal cancer. *Cell* 1998, 94(6):703-714.
15. Park YK, Franklin JL, Settle SH, Levy SE, Chung E, Jeyakumar LH, Shyr Y, Washington MK, Whitehead RH, Aronow BJ *et al*: Gene expression profile analysis of mouse colon embryonic development. *Genesis* 2005, 41(1):1-12.
16. Reichling T, Goss KH, Carson DJ, Holdcraft RW, Ley-Ebert C, Witte D, Aronow BJ, Groden J: Transcriptional profiles of intestinal tumors in Apc(Min) mice are unique from those of embryonic intestine and identify novel gene targets dysregulated in human colorectal tumors. *Cancer Res* 2005, 65(1):166-176.
17. van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Batlle E, Coudreuse D, Haramis AP *et al*: The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* 2002, 111(2):241-250.
18. Zhu P, Martin E, Mengwasser J, Schlag P, Janssen KP, Gottlicher M: Induction of HDAC2 expression upon loss of APC in colorectal tumorigenesis. *Cancer Cell* 2004, 5(5):455-463.
19. Watson JD, Oster SK, Shago M, Khosravi F, Penn LZ: Identifying genes regulated in a Myc-dependent manner. *J Biol Chem* 2002, 277(40):36921-36930.
20. Katz JP, Perreault N, Goldstein BG, Actman L, McNally SR, Silberg DG, Furth EE, Kaestner KH: Loss of Klf4 in mice causes altered proliferation and differentiation and precancerous changes in the adult stomach. *Gastroenterology* 2005, 128(4):935-945.
21. Lamhonwah AM, Ackerley C, Onizuka R, Tilups A, Lamhonwah D, Chung C, Tao KS, Tellier R, Tein I: Epitope shared by functional variant of organic cation/carnitine transporter, OCTN1, *Campylobacter jejuni* and *Mycobacterium paratuberculosis* may underlie susceptibility to Crohn's disease at 5q31. *Biochem Biophys Res Commun* 2005, 337(4):1165-1175.
22. Cragg RA, Phillips SR, Piper JM, Varma JS, Campbell FC, Mathers JC, Ford D: Homeostatic regulation of zinc transporters in the human small intestine by dietary zinc supplementation. *Gut* 2005, 54(4):469-478.
23. Jenny M, Uhl C, Roche C, Duluc I, Guillermin V, Guillemot F, Jensen J, Kedinger M, Gradwohl G: Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium. *Embo J* 2002, 21(23):6338-6347.
24. Kruhoffer M, Jensen JL, Laiho P, Dyrskjot L, Salovaara R, Arango D, Birkenkamp-Demtroder K, Sorensen FB, Christensen LL, Buh1 L *et al*: Gene expression signatures for colorectal cancer microsatellite status and HNPCC. *Br J Cancer* 2005, 92(12):2240-2248.

25. Giacomini CP, Leung SY, Chen X, Yuen ST, Kim YH, Bair E, Pollack JR: A gene expression signature of genetic instability in colon cancer. *Cancer Res* 2005, 65(20):9200-9205.
26. Hu M, Shivdasani RA: Overlapping gene expression in fetal mouse intestine development and human colorectal cancer. *Cancer Res* 2005, 65(19):8715-8722.
27. He TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B, Kinzler KW: Identification of c-MYC as a target of the APC pathway. *Science* 1998, 281(5382):1509-1512.
28. Feng XH, Liang YY, Liang M, Zhai W, Lin X: Direct interaction of c-Myc with Smad2 and Smad3 to inhibit TGF-beta-mediated induction of the CDK inhibitor p15(Ink4B). *Mol Cell* 2002, 9(1):133-143.
29. Frederick JP, Liberati NT, Waddell DS, Shi Y, Wang XF: Transforming growth factor beta-mediated transcriptional repression of c-myc is dependent on direct binding of Smad3 to a novel repressive Smad binding element. *Mol Cell Biol* 2004, 24(6):2546-2559.
30. Warner BJ, Blain SW, Seoane J, Massague J: Myc downregulation by transforming growth factor beta required for activation of the p15(Ink4b) G(1) arrest pathway. *Mol Cell Biol* 1999, 19(9):5913-5922.
31. Maggio-Price L, Treuting P, Zeng W, Tsang M, Bielefeldt-Ohmann H, Iritani BM: Helicobacter infection is required for inflammation and colon cancer in SMAD3-deficient mice. *Cancer Res* 2006, 66(2):828-838.
32. Reya T, Clevers H: Wnt signalling in stem cells and cancer. *Nature* 2005, 434(7035):843-850.
33. Korinek V, Barker N, Moerer P, van Donselaar E, Huls G, Peters PJ, Clevers H: Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat Genet* 1998, 19(4):379-383.
34. Pinto D, Gregorieff A, Begthel H, Clevers H: Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev* 2003, 17(14):1709-1713.
35. Kuhnert F, Davis CR, Wang HT, Chu P, Lee M, Yuan J, Nusse R, Kuo CJ: Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. *Proc Natl Acad Sci U S A* 2004, 101(1):266-271.
36. Kwong KY, Bloom GC, Yang I, Boulware D, Coppola D, Haseman J, Chen E, McGrath A, Makusky AJ, Taylor J *et al*: Synchronous global assessment of gene and protein expression in colorectal cancer progression. *Genomics* 2005, 86(2):142-158.
37. Zeng W, Wharton KA, Jr., Mack JA, Wang K, Gadbaw M, Suyama K, Klein PS, Scott MP: naked cuticle encodes an inducible antagonist of Wnt signalling. *Nature* 2000, 403(6771):789-795.
38. Smit L, Baas A, Kuipers J, Korswagen H, van de Wetering M, Clevers H: Wnt activates the Tak1/Nemo-like kinase pathway. *J Biol Chem* 2004, 279(17):17232-17240.

39. Byun T, Karimi M, Marsh JL, Milovanovic T, Lin F, Holcombe RF: Expression of secreted Wnt antagonists in gastrointestinal tissues: potential role in stem cell homeostasis. *J Clin Pathol* 2005, 58(5):515-519.
40. Yang Q, Bermingham NA, Finegold MJ, Zoghbi HY: Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. *Science* 2001, 294(5549):2155-2158.
41. Batlle E, Bacani J, Begthel H, Jonkheer S, Gregorieff A, van de Born M, Malats N, Sancho E, Boon E, Pawson T *et al*: EphB receptor activity suppresses colorectal cancer progression. *Nature* 2005, 435(7045):1126-1130.
42. Batlle E, Henderson JT, Begthel H, van den Born MM, Sancho E, Huls G, Meeldijk J, Robertson J, van de Wetering M, Pawson T *et al*: Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* 2002, 111(2):251-263.
43. Engle SJ, Hoying JB, Boivin GP, Ormsby I, Gartside PS, Doetschman T: Transforming growth factor beta1 suppresses nonmetastatic colon cancer at an early stage of tumorigenesis. *Cancer Res* 1999, 59(14):3379-3386.
44. Sodir NM, Chen X, Park R, Nickel AE, Conti PS, Moats R, Bading JR, Shibata D, Laird PW: Smad3 Deficiency Promotes Tumorigenesis in the Distal Colon of ApcMin/+ Mice. *Cancer Res* 2006, 66(17):8430-8438.
45. Taketo MM, Takaku K: Gastro-intestinal tumorigenesis in Smad4 mutant mice. *Cytokine Growth Factor Rev* 2000, 11(1-2):147-157.
46. Munoz NM, Upton M, Rojas A, Washington MK, Lin L, Chytil A, Sozmen EG, Madison BB, Pozzi A, Moon RT *et al*: Transforming growth factor beta receptor type II inactivation induces the malignant transformation of intestinal neoplasms initiated by Apc mutation. *Cancer Res* 2006, 66(20):9837-9844.
47. Plebani M, Basso D, Panozzo MP, Fogar P, Del Favero G, Naccarato R: Tumor markers in the diagnosis, monitoring and therapy of pancreatic cancer: state of the art. *Int J Biol Markers* 1995, 10(4):189-199.
48. Zbar AP: The immunology of colorectal cancer. *Surg Oncol* 2004, 13(2-3):45-53.
49. Valk-Lingbeek ME, Bruggeman SW, van Lohuizen M: Stem cells and cancer; the polycomb connection. *Cell* 2004, 118(4):409-418.
50. Kim DH, Kim M, Kwon HJ: Histone deacetylase in carcinogenesis and its inhibitors as anti-cancer agents. *J Biochem Mol Biol* 2003, 36(1):110-119.
51. Sansom OJ, Meniel VS, Muncan V, Phesse TJ, Wilkins JA, Reed KR, Vass JK, Athineos D, Clevers H, Clarke AR: Myc deletion rescues Apc deficiency in the small intestine. *Nature* 2007, 446(7136):676-679.
52. Abdel-Rahman WM, Ollikainen M, Kariola R, Jarvinen HJ, Mecklin JP, Nystrom-Lahti M, Knuutila S, Peltomaki P: Comprehensive characterization of HNPCC-related colorectal cancers reveals striking molecular features in families with no germline mismatch repair gene mutations. *Oncogene* 2005, 24(9):1542-1551.

53. Kho AT, Zhao Q, Cai Z, Butte AJ, Kim JY, Pomeroy SL, Rowitch DH, Kohane IS: Conserved mechanisms across development and tumorigenesis revealed by a mouse development perspective of human cancers. *Genes Dev* 2004, 18(6):629-640.
54. Lepourcelet M, Tou L, Cai L, Sawada J, Lazar AJ, Glickman JN, Williamson JA, Everett AD, Redston M, Fox EA *et al*: Insights into developmental mechanisms and cancers in the mammalian intestine derived from serial analysis of gene expression and study of the hepatoma-derived growth factor (HDGF). *Development* 2005, 132(2):415-427.
55. Lee JS, Chu IS, Mikaelyan A, Calvisi DF, Heo J, Reddy JK, Thorgeirsson SS: Application of comparative functional genomics to identify best-fit mouse models to study human cancer. *Nat Genet* 2004, 36(12):1306-1311.
56. Haigis KM, Hoff PD, White A, Shoemaker AR, Halberg RB, Dove WF: Tumor regionality in the mouse intestine reflects the mechanism of loss of Apc function. *Proc Natl Acad Sci U S A* 2004, 101(26):9769-9773.
57. Mithani SK, Balch GC, Shiou SR, Whitehead RH, Datta PK, Beauchamp RD: Smad3 has a critical role in TGF-beta-mediated growth inhibition and apoptosis in colonic epithelial cells. *J Surg Res* 2004, 117(2):296-305.
58. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I: Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001, 125(1-2):279-284.
59. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001, 25(4):402-408.
60. Kaiser S., et al.: Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer. *Genome Biology* 2007; 8(7):R131.

## CLAIMS

What is claimed is :

1. A method for making a diagnosis of a colorectal cancer in a subject, comprising:
  - (a) providing a biological sample from the subject;
  - (b) determining an amount in the sample of at least one peptide selected from the peptides set forth in Table A; and
  - (c) comparing the amount of the at least one peptide in the sample, if present, to a control level of the at least one peptide, wherein the subject is diagnosed as having colorectal cancer or a risk thereof if there is a measurable difference in the amount of the at least one peptide in the sample as compared to the control level.
2. The method of claim 1, further comprising determining an amount in the sample of at least two peptides selected the peptides set forth in Table A.
3. The method of claim 1, further comprising determining an amount in the sample of at least five peptides selected from the peptides set forth in Table A.
4. The method of claim 1, further comprising determining an amount in the sample of at least ten peptides selected from the peptides set forth in Table A.
5. The method of claim 1, wherein the biological sample comprises blood, plasma, or serum.
6. The method of claim 1, wherein the subject is human.
7. The method of claim 1, wherein determining the amount in the sample of the at least one peptide comprises determining the amount in the sample of the at least one peptide using mass spectrometry (MS) analysis, immunoassay analysis, or both.

8. The method of claim 7, wherein the immunoassay analysis comprises an enzyme-linked immunosorbent assay (ELISA).
9. The method of claim 1, further comprising selecting a treatment or modifying a treatment for the colorectal cancer based on the determined amount of the at least one peptide.
10. The method of claim 1, further comprising determining an amount in the sample of at least one additional peptide is selected from the peptides set forth in Table B.
11. The method of claim 10, wherein the biological sample comprises blood, plasma, or serum.
12. The method of claim 10, wherein the subject is human.
13. The method of claim 10, wherein determining the amount in the sample of the at least one peptide comprises determining the amount in the sample of the at least one peptide using mass spectrometry (MS) analysis, immunoassay analysis, or both.
14. The method of claim 10, wherein the immunoassay analysis comprises an enzyme-linked immunosorbent assay (ELISA).
15. The method of claim 10, further comprising selecting a treatment or modifying a treatment for the colorectal cancer based on the determined amount of the at least one peptide.
16. The method of claim 1, further comprising determining an amount in the sample of at least one additional peptide is selected from the peptides set forth in Table C.
17. The method of claim 16, wherein the biological sample comprises blood, plasma, or serum.
18. The method of claim 16, wherein the subject is human.

19. The method of claim 16, wherein determining the amount in the sample of the at least one peptide comprises determining the amount in the sample of the at least one peptide using mass spectrometry (MS) analysis, immunoassay analysis, or both.

20. The method of claim 16, wherein the immunoassay analysis comprises an enzyme-linked immunosorbent assay (ELISA).

21. The method of claim 16, further comprising selecting a treatment or modifying a treatment for the colorectal cancer based on the determined amount of the at least one peptide.

22. A system for making a diagnosis of a colorectal cancer in a subject, comprising:  
probes for selectively binding each of one or more peptide markers, wherein the peptide markers are selected from the peptides set forth in Table A; and  
means for detecting the binding of said probes to said one or more peptide markers.

23. The system of claim 22, further comprising probes for selectively binding each of one or more peptide markers selected from the peptides set forth in Table B.

24. The system of claim 22, further comprising probes for selectively binding each of one or more peptide markers selected from the peptides set forth in Table C.

25. A method for making a diagnosis of a colorectal cancer in a subject, comprising:  
(a) providing a biological sample from the subject;  
(b) determining an amount in the sample of at least one peptide selected from the peptides set forth in Table B; and  
(c) comparing the amount of the at least one peptide in the sample, if present, to a control level of the at least one peptide, wherein the subject is diagnosed as having colorectal cancer or a risk thereof if there is a measurable difference in the amount of the at least one peptide in the sample as compared to the control level.

26. The method of claim 25, further comprising determining an amount in the sample of at least two peptides selected the peptides set forth in Table B.

27. The method of claim 25, further comprising determining an amount in the sample of at least five peptides selected from the peptides set forth in Table B.

28. The method of claim 25, further comprising determining an amount in the sample of at least ten peptides selected from the peptides set forth in Table B.

29. The method of claim 25, wherein the biological sample comprises blood, plasma, or serum.

30. The method of claim 25, wherein the subject is human.

31. The method of claim 25, wherein determining the amount in the sample of the at least one peptide comprises determining the amount in the sample of the at least one peptide using mass spectrometry (MS) analysis, immunoassay analysis, or both.

32. The method of claim 31, wherein the immunoassay analysis comprises an enzyme-linked immunosorbent assay (ELISA).

33. The method of claim 25, further comprising selecting a treatment or modifying a treatment for the colorectal cancer based on the determined amount of the at least one peptide.

34. A system for making a diagnosis of a colorectal cancer in a subject, comprising:  
probes for selectively binding each of one or more peptide markers, wherein the peptide markers are selected from the peptides set forth in Table B; and  
means for detecting the binding of said probes to said one or more peptide markers.

35. A method for making a diagnosis of a colorectal cancer in a subject, comprising:  
(a) providing a biological sample from the subject;

(b) determining an amount in the sample of at least one peptide marker from at least one group of markers, selected from

(i) the group of markers set forth in Tables H and I, which are associated with a first type of colorectal cancer;

(ii) the group of markers set forth in Tables J and K, which are associated with a second type of colorectal cancer;

(iii) the group of markers set forth in Tables L and M, which are associated with a third type of colorectal cancer; and

(iv) the group of markers set forth in Tables N and O, which are associated with a fourth type of colorectal cancer; and

(c) comparing the amount of the at least one peptide marker in the sample, if present, to a control level of the at least one peptide marker; wherein the subject is diagnosed as having the type of colorectal cancer associated with the group of markers of which the at least one peptide marker is a member if there is a measurable difference in the amount of the at least one peptide in the sample as compared to the control level.

36. The method of claim 35, further comprising determining the amount in the sample of at least one peptide marker from at least two groups of markers.

37. The method of claim 35, further comprising determining the amount in the sample of at least one peptide marker from at least three groups of markers.

38. The method of claim 35, further comprising determining the amount in the sample of at least one peptide marker from four of the groups of markers.

39. The method of claim 35, wherein the at least one group of markers is selected from

(i) the group of markers set forth in Table I;

(ii) the group of markers set forth in Table K;

(iii) the group of markers set forth in Table M; and

(iv) the group of markers set forth in Table O.

40. The method of claim 35, further comprising determining the amount in the sample of at least one peptide marker from at least two groups of markers.

41. The method of claim 35, further comprising determining the amount in the sample of at least one peptide marker from at least three groups of markers.

42. The method of claim 35, further comprising determining the amount in the sample of at least one peptide marker from each of the groups of markers.

43. A system for making a diagnosis of a colorectal cancer in a subject, comprising:  
probes for selectively binding each of one or more peptide markers, wherein the peptide markers are selected from at least one group of markers, selected from  
(i) the group of markers set forth in Tables H and I, which are associated with a first type of colorectal cancer;  
(ii) the group of markers set forth in Tables J and K, which are associated with a second type of colorectal cancer;  
(iii) the group of markers set forth in Tables L and M, which are associated with a third type of colorectal cancer; and  
(iv) the group of markers set forth in Tables N and O, which are associated with a fourth type of colorectal cancer; and  
means for detecting the binding of said probes to said one or more peptide markers.

44. The system of claim 43, wherein the at least one group of markers is selected from  
(i) the group of markers set forth in Table I;  
(ii) the group of markers set forth in Table K;  
(iii) the group of markers set forth in Table M; and  
(iv) the group of markers set forth in Table O.

45. A method for making a diagnosis of a colorectal cancer in a subject, comprising:  
(a) providing a biological sample from the subject;

(b) determining an amount in the sample of at least one peptide selected from the peptides set forth in Table C; and

(c) comparing the amount of the at least one peptide in the sample, if present, to a control level of the at least one peptide, wherein the subject is diagnosed as having colorectal cancer or a risk thereof if there is a measurable difference in the amount of the at least one peptide in the sample as compared to the control level.

46. The method of claim 45, further comprising determining an amount in the sample of at least two peptides selected the peptides set forth in Table C.

47. The method of claim 45, further comprising determining an amount in the sample of at least five peptides selected from the peptides set forth in Table C.

48. The method of claim 45, further comprising determining an amount in the sample of at least ten peptides selected from the peptides set forth in Table C.

49. The method of claim 45, comprising determining an amount in the sample of at least one peptide selected from the peptides set forth in Table D.

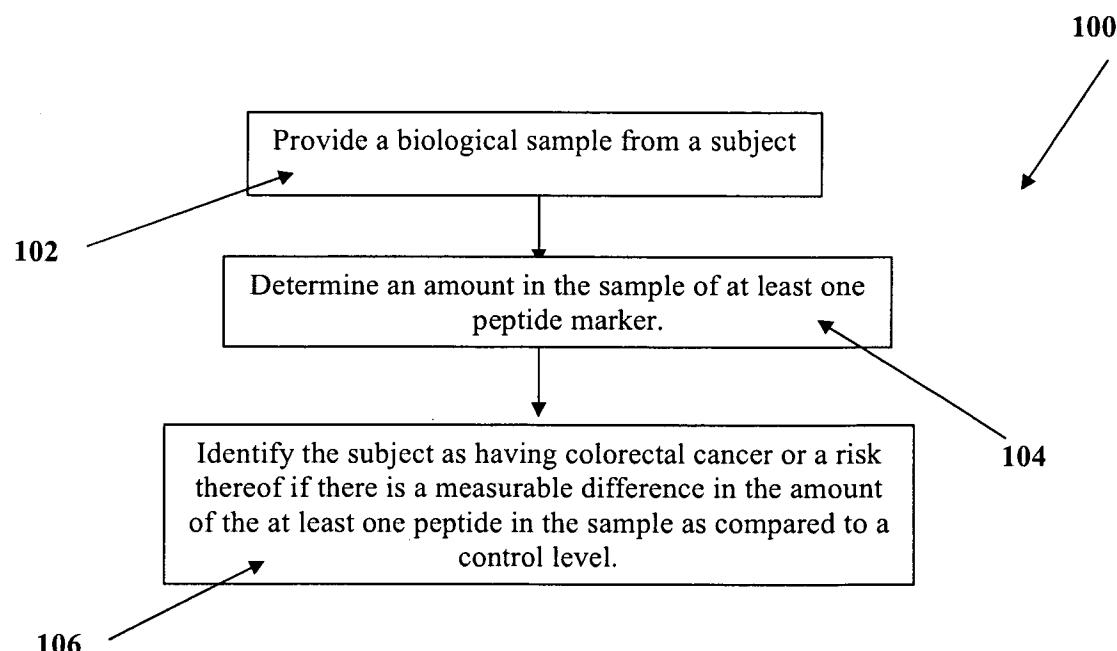
50. The method of claim 45, comprising determining an amount in the sample of at least one peptide selected from the peptides set forth in Table E.

51. The method of claim 45, comprising determining an amount in the sample of at least one peptide selected from the peptides set forth in Table F.

52. The method of claim 45, comprising determining an amount in the sample of at least one peptide selected from the peptides set forth in Table G.

53. The method of claim 45, wherein the biological sample comprises blood, plasma, or serum.

54. The method of claim 45, wherein the subject is human.
55. The method of claim 45, wherein determining the amount in the sample of the at least one peptide comprises determining the amount in the sample of the at least one peptide using mass spectrometry (MS) analysis, immunoassay analysis, or both.
56. The method of claim 55, wherein the immunoassay analysis comprises an enzyme-linked immunosorbent assay (ELISA).
57. The method of claim 45, further comprising selecting a treatment or modifying a treatment for the colorectal cancer based on the determined amount of the at least one peptide.
58. A system for making a diagnosis of a colorectal cancer in a subject, comprising:  
probes for selectively binding each of one or more peptide markers, wherein the peptide markers are selected from the peptides set forth in Table C; and  
means for detecting the binding of said probes to said one or more peptide markers.
59. The system of claim 58, wherein the peptide markers are selected from the peptides set forth in Table D.
60. The system of claim 58, wherein the peptide markers are selected from the peptides set forth in Table E.
61. The system of claim 58, wherein the peptide markers are selected from the peptides set forth in Table F.
62. The system of claim 58, wherein the peptide markers are selected from the peptides set forth in Table G.

**FIG. 1**

2/6

## Identification of Differentially Regulated Genes

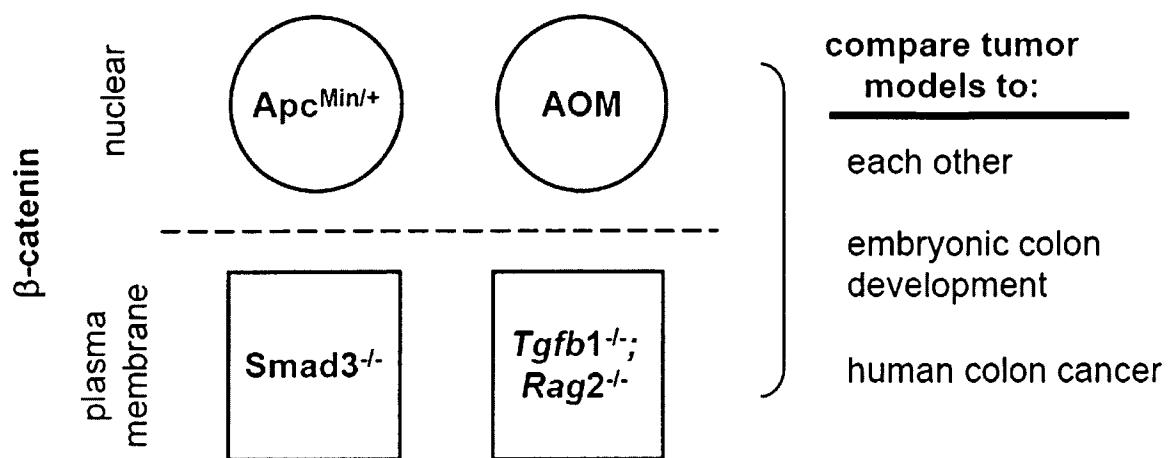


FIG. 2

3/6

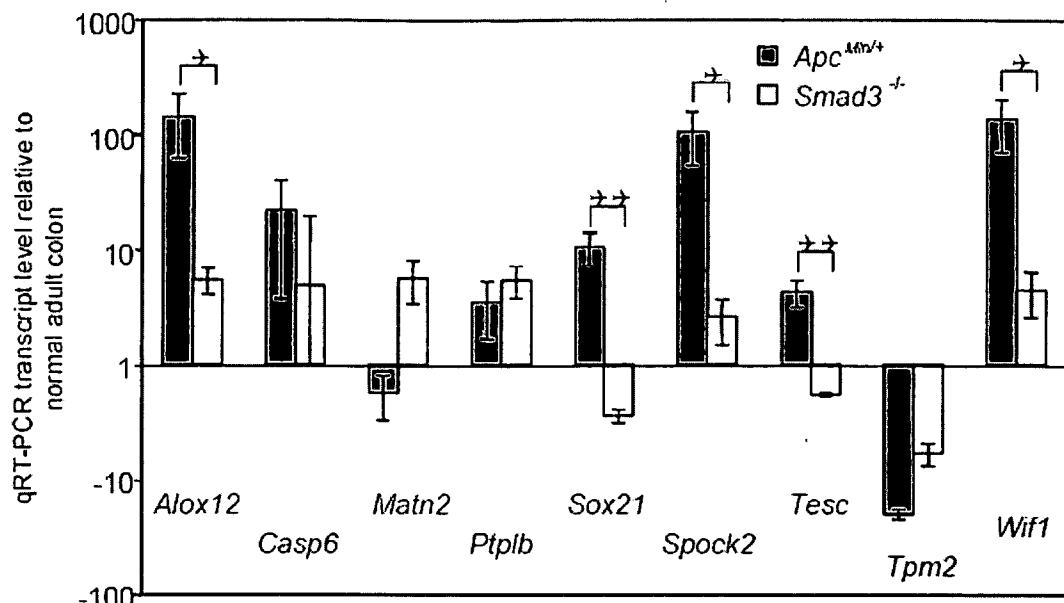


FIG. 3

Embryonic Colon					
Murine Tum Models	over-expressed		under-expressed		expected
	observed	expected	observed	expected	
	total feature count	50393	5796	na	3541
over-expressed	8804	4693*	2502	588***	1529
under-expressed	8018	2004**	2278	3282****	1392

FIG. 4

Embryonic Colon					
human CRCs	over-expressed		under-expressed		expected
	observed	expected	observed	expected	
	ortholog genome	8621	2718	na	2365
over-expressed	2212	1080*	697	549***	607
under-expressed	737	156**	232	431****	202

FIG. 5

4/6

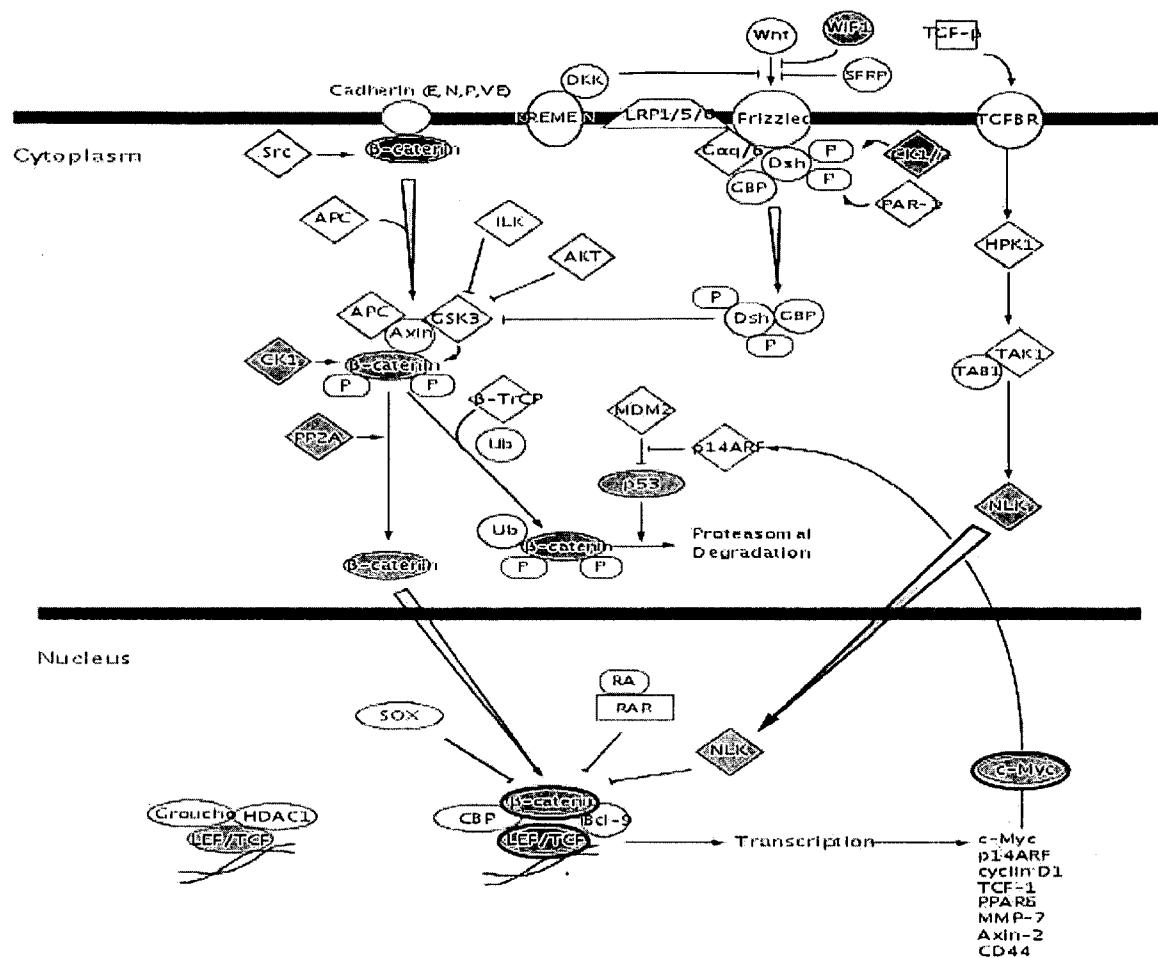


FIG. 6A

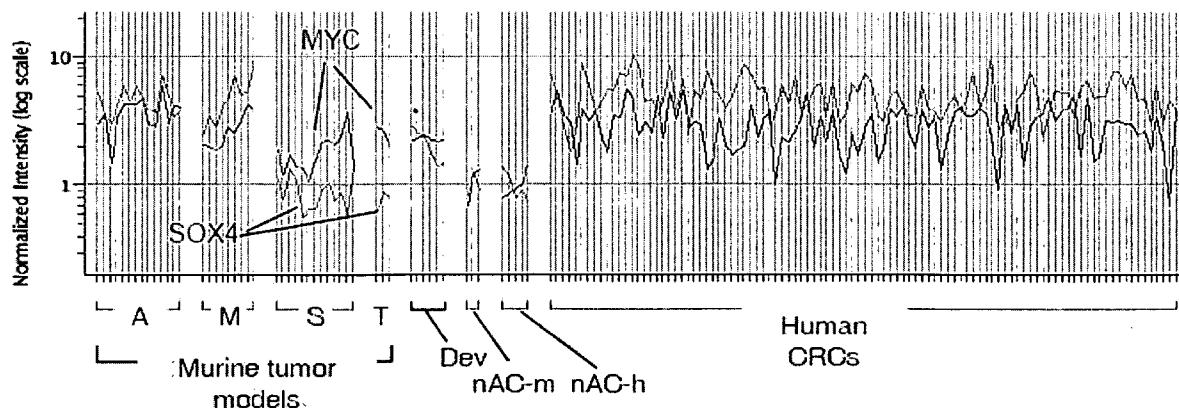


FIG. 6B

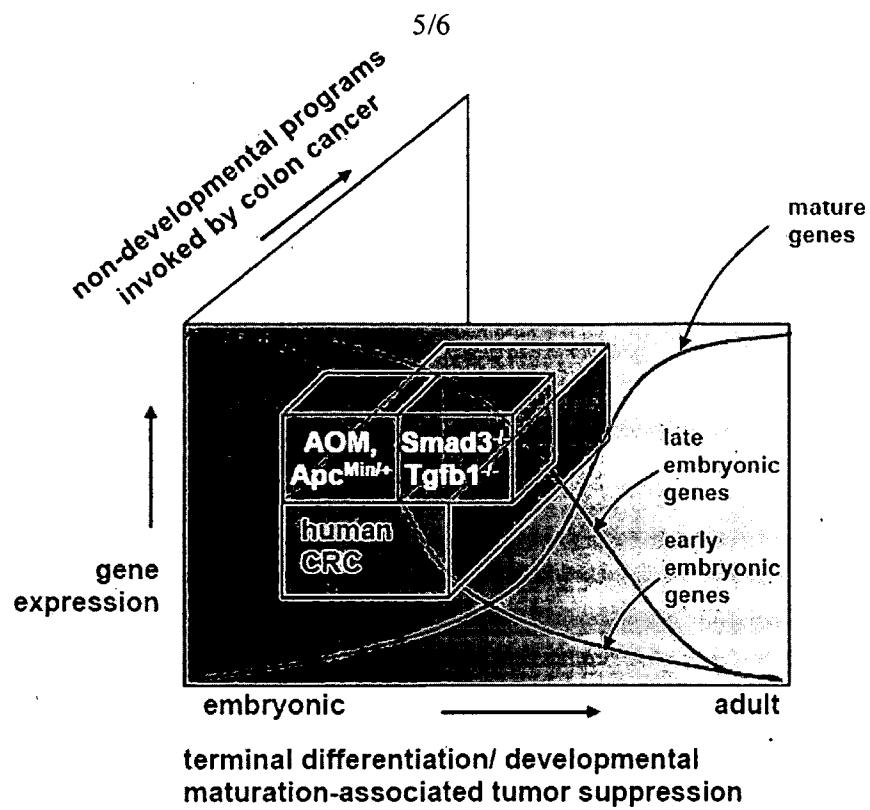


FIG. 7A

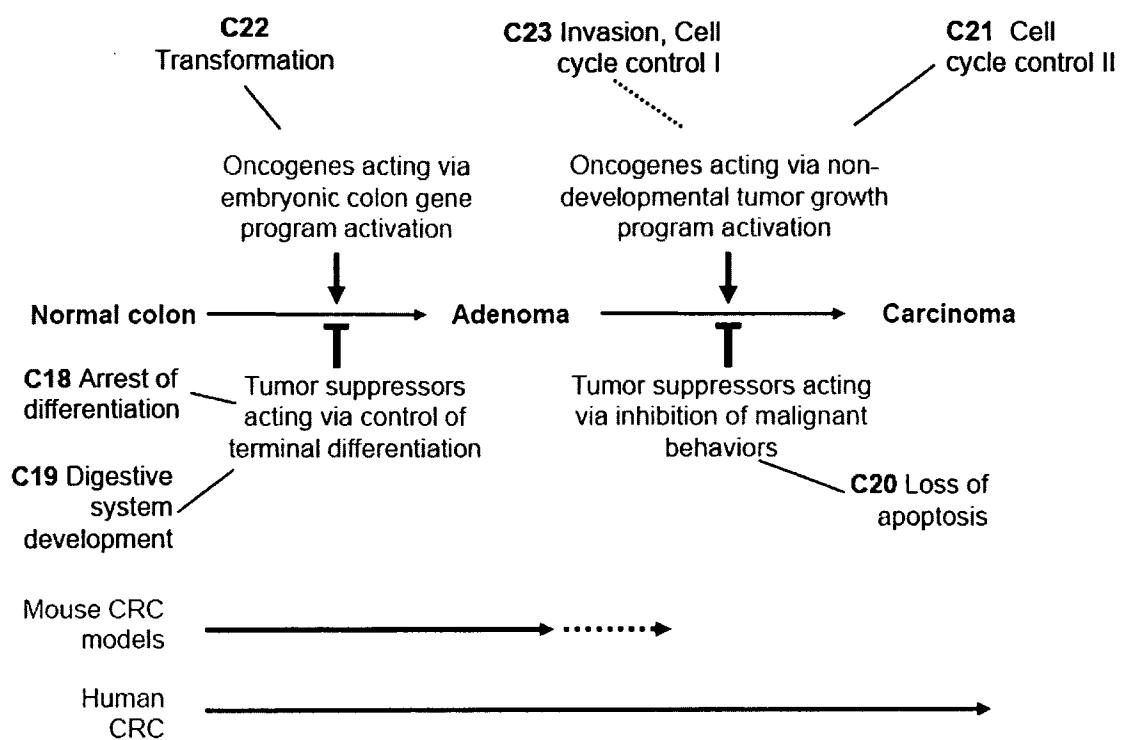


FIG. 7B