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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 November 2012 (22.11.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/156529 A1

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

(21) International Application Number:
PCT/EP2012/059339

(22) International Filing Date:
21 May 2012 (21.05.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11166816.6 19 May 2011 (19.05.2011) EP

(71) Applicant *(for all designated States except US)*: **SIGNATURE DIAGNOSTICS AG** [DE/DE]; Hermannswerder 20a, 14473 Potsdam (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ADAMS, Hans-Peter** [DE/DE]; Gerlachstr. 41, 14478 Potsdam (DE). **HINZMANN, Bernd** [DE/DE]; Saupéweg 10, 13127 Berlin (DE). **MAYR, Tobias** [DE/DE]; Schönhauser Allee 162, 10435 Berlin (DE). **ROSENTHAL, André** [DE/DE]; Gröbener Dorfstr. 25, 14194 Ludwigsfelde (DE).

(74) Agent: **HOPPE, Georg, J.**; Beuckestr. 20, 14163 Berlin (DE).

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

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

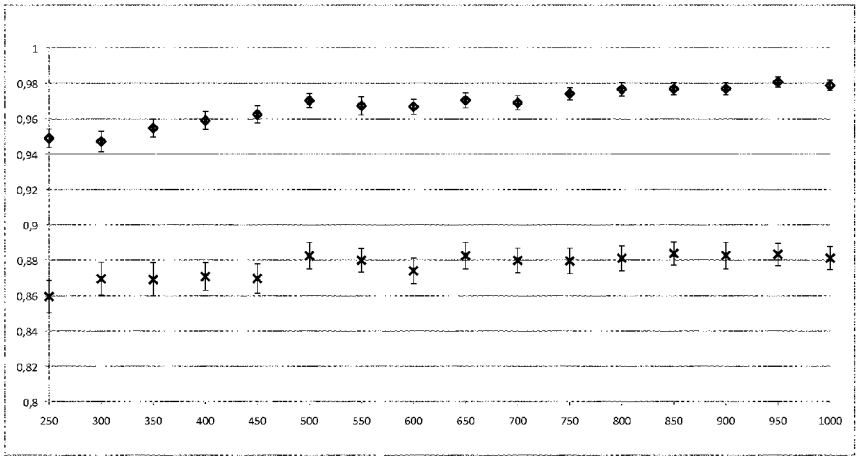
Published:

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

— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title**: METHODS AND KITS FOR DIAGNOSING COLORECTAL CANCER

Figure 1



(57) **Abstract**: The invention pertains to a method for early detection and screening of colorectal cancer in human subjects based on RNA isolated from blood obtained from said subject. According to the invention, the abundance of at least 250 RNAs listed in tables 1 to 4 is measured. Using the invention, an accurate and noninvasive screening and diagnosis tool for colorectal cancer is provided with a sensitivity of at least 75 % and a specificity of 85 % that has high clinical utility and the potential for broad adoption.

WO 2012/156529 A1

Methods And Kits For Diagnosing Colorectal Cancer

The invention pertains to a method for diagnosing or detecting colorectal cancer in human subjects based on ribonucleic acid (RNA), in particular based on RNA from blood.

Background

Colorectal cancer (CRC) is the second-leading cause of cancer-related deaths in the United States. Each year, approximately 150,000 people are diagnosed with CRC and almost 60,000 people die from the disease.

CRC arises from the mucosa forming the inner lining of colon and rectum. Like any other mucosa, it needs to be regenerated and proliferates at a high rate (about one third of all fecal matter are mucosa cells), and is thus susceptible to abnormal growth, *i.e.*, neoplasia and/or dysplasia. In fact, abnormal colonic mucosal growth can be detected in about 40 % of all persons over the age of 55 years. The development of neoplasia into cancer is a well-established concept in the biomedical sciences; and is termed adenoma-carcinoma-sequence (ACS).

Pathologists classify abnormal mucosal growth into four categories with increasing severity: 1) Low-grade intraepithelial neoplasia (LIEN) or adenoma, which occurs in more than 30 %; 2) high-grade intraepithelial neoplasia (HIEN) or advanced adenoma, occurring in more than 2 %; 3) carcinoma in situ (CIS or pTis), where the cancerous growth is still confined to the mucosa; and 4) CRC, where the cancerous growth has invaded the submucosa. CRC is diagnosed with an incidence rate of about 1 % in persons over the age of 55 years with an average risk for the disease. The lifetime risk of developing CRC is estimated to be 1 in 18 persons (Cancer Statistics 2009: A Presentation From the American Cancer Society; 2009, American Cancer Society, Inc.).

After primary diagnosis of CRC, the spread/stage of the disease is classified according to the guidelines set forth by the "Union International Contre le Cancer" (UICC). UICC-stage 0 includes CIS only. UICC-stages I and II are comprised of the localized stages, whereas UICC-stage III describes CRC where tumor cells have metastasized into regional lymph

nodes. The worst case is UICC-stage IV; it describes CRC which has metastasized into other organ(s), usually liver (~75 %), peritoneum (~22 %), and/or lung (~21 %).

In 2008, the cancer registry in the state of Brandenburg/Germany documented 1591 patients with newly diagnosed CRC and stage information. They were staged into UICC-stage I: 22.6 %, UICC-stage II: 29.2 %, UICC-stage III: 28.9 %, and UICC-stage IV: 19.0 %. Relative five year survival-rates by UICC-stage were: I: 90.5%, II: 78.8%, III: 60.6%, and IV: 9.3%.

The U.S. National Institutes of Health (http://seer.cancer.gov) reported for the period 1999 to 2006 216,332 patients diagnosed with CRC with localized disease (UICC-stage I and II): 39 %, regional disease (UICC stage III): 37 % and distant disease (UICC-stage IV): 19 %. Relative five year survival-rates by stage were: localized (UICC-stages I and II): 90.4 %, regional (UICC stage III): 60.5 %, and distant (UICC-stage IV): 11.6 %. However, the statistics of the U.S. National Institutes of Health do not cover the U.S. population, while the data from the cancer registry in the state of Brandenburg/Germany do.

Current technologies to detect mucosal neoplasia (polyps/adenoma) and CRC can be categorized into three classes:

I) ***in-vitro* diagnostics** (IVDs) – a specimen/sample (blood, stool, or urine) is taken from the test person and analyzed for one or more biomarkers as surrogate markers for colorectal neoplasia/cancer. Exemplary tests include guaic fecal occult blood test (gFOBT) or immunological fecal occult blood test (iFOBT), detection of tumor DNA-chains (deoxyribonucleic acid chains) in stool samples, detection of specific methylated tumor DNA-chains in stool samples, detection of specific free methylated DNA-chains in blood plasma, detection of elevated and/or lowered amounts of specific proteins in blood samples, or detection of elevated and/or lowered amounts of specific RNA-chains in blood samples;

II) **imaging methods without interventional capabilities** such as X-ray, double contrast barium enema (DCBE), video capsule endoscopy, or computed tomographic colonography;

III) **imaging methods with interventional capabilities** such as flexible sigmoidoscopy, colonoscopy, laparoscopy, or open surgery.

To obtain a definitive diagnosis of colorectal neoplasia/cancer, an invasive procedure is typically required. The procedure requires taking a sample of the visibly abnormal tissue growth (neoplasia/cancer) and having a person of skill in the art of pathology examine this sample, who will then decide (diagnose) whether this sample was taken from a neoplasia/cancer or not (Sternberg's Diagnostic Surgical Pathology (5th Edition). Mills SE, Carter D, Greenson JK, Reuter V, Stoler MH. Lippincott Williams & Wilkins (LWW), 2009).

In response to the high incidence and mortality rate of patients with CRC, the American Cancer Society issued the following statement: "There are significant updates to the guidelines for colorectal cancer screening. Two new tests are now recommended as options for colorectal cancer screening. They are stool DNA (sDNA) and computerized tomographic colonography (also known as "virtual colonoscopy"). For the first time, screening tests are grouped into categories based on performance characteristics: those that primarily detect cancer early and those that can also detect precancerous polyps. Tests that primarily detect cancer early are fecal (stool) tests, including guaiac-based and immunochemical-based fecal occult blood tests (gFOBT & FIT), and stool DNA tests (sDNA). Tests that detect both precancerous polyps and cancer include flexible sigmoidoscopy, colonoscopy, the double contrast barium enema, and computerized tomographic colonography (also known as virtual colonoscopy). It is the strong opinion of the expert panel that colon cancer prevention should be the primary goal of colorectal cancer screening. Exams that are designed to detect both early cancer and precancerous polyps should be encouraged if resources are available and patients are willing to undergo an invasive test." (Cancer Statistics 2009: A Presentation from the American Cancer Society; ©2009, American Cancer Society, Inc.). A review of current CRC screening in Europe can be found in: Zavoral M, Suchanek S, Zavada F, Dusek L, Muzik J, Seifert B, Fric P. Colorectal cancer screening in Europe. World J Gastroenterol. 2009 Dec 21; 15: 5907-15.

However, each of the tests for the detection of mucosal neoplasias (polyps/adenomas) and CRC has limitations.

For example, the imaging methods (with or without interventional capabilities) require preparation time for the test subject, specialized equipment, and specialized medical personnel. Therefore, colonoscopy and flexible sigmoidoscopy are used only in wealthy

economies such as the U.S., Switzerland, and Germany as primary screening tools for early detection of CRC. Even in the U.K., France, Italy or Spain, IVDs, in most instances gFOBT, are used as a primary screening tool for colorectal cancer. Only patients with a positive IVD test result are referred to colonoscopy.

Recently, a screening program for CRC using gFOBT was initiated in the United Kingdom. All eligible persons were contacted via mail, and a test kit was delivered. Yet, just about 50 % of all contacted people complied. The willingness of patients to undergo gFOBT testing in Germany has dropped from 8.2 million tests in 2001 to 4.6 million tests in 2007 (Projekt wissenschaftliche Begleitung von Früherkennungs-Koloskopien in Deutschland, Berichtszeitraum 2008 - 6. Jahresbericht, im Auftrag des GKV - Spitzenverbands und der Kassenärztlichen Bundesvereinigung Version 1.1, Stand: 19. Februar 2010, Zentralinstitut für die kassenärztliche Versorgung in der Bundesrepublik Deutschland). In the U.S., about 24.01 % of all eligible patients underwent gFOBT-screening in 2000, in 2005 the rate dropped to 17.07 % (U.S. National Cancer Institute, http://progressreport.cancer.gov).

Thus, the clinical utility of all stool-based CRC-screening is limited because individuals in the CRC screening population are simply unwilling to take the test repeatedly, unless they have no other choice.

The U.S. National Institutes of Health reported that compliance with endoscopy (flexible sigmoidoscopy or colonoscopy) is dependent on the education and income of the population; by 2005 37.66 % of persons with less than high school education, 46.27 % of persons with high school education, and 57.52 % of persons with higher than high school education had ever had an endoscopy (not defined to CRC screening purposes).

Colonoscopy is an invasive procedure, which is not only inconvenient but may be associated with health risks. Approximately 3 % of the individuals over 55 years undergoing colonoscopy for screening purposes have heavy bleeding incidences. Additionally, in 2 of 1,000 individuals perforation of the colon occurs. Emergency operations must be performed to correct both heavy bleeding and perforation. As a result, 2 of 10,000 individuals who undergo colonoscopy will die from these complications. The relatively high rate of accidents in combination with the time consuming bowel cleaning procedure has led to a low adoption

of colonoscopy as a screening tool even in those countries where colonoscopy is paid by the health insurances.

Thus, the overall clinical utility of all endoscopy-based CRC-screening is also limited, because it is only offered in a few countries, a high percentage of the CRC screening population is unwilling to take the test, and because of the complications associated with the test.

Therefore, the clinical utility of a test for detection of colorectal neoplasia depends not only on its performance characteristics, *i.e.*, sensitivity and specificity, but on acceptance by the patients, the medical community, and, of course, the private or public health care system that has to pay for the test.

A blood test would have the highest chance of acceptance by patients, at least in Europe, the U.S., and Japan. In terms of the medical community, a blood test would also have the highest chance of acceptance, in particular if sensitivity and specificity are convincingly high, if there is no need for preparation time, if the blood need not be processed immediately but can be shipped to a central laboratory, if the test is accepted by local regulatory authorities, and if the test is commercially available. A high level of acceptance of a test can only be achieved if the test is endorsed by CRC screening guidelines and by the general health care system.

Although blood-based colon cancer screening has been attempted, each previously reported test is inherently limited in its respective specificity and sensitivity.

For example, Han et al., (Clin Cancer Res 2008; 14, 455-460; also: WO 2007/048074 A1) reported the discovery and validation of a five-gene expression (messenger RNA) signature for detection of colorectal carcinoma. Basically, the 37 candidate genes for the signature were selected from microarray data of 16 CRC cases and 15 controls. These 37 candidate genes were evaluated on a second set of 115 samples (58 CRC, 57 controls) using quantitative real-time PCR, validating 17 genes as differentially expressed. A further gene selection step using the PCR-results revealed the 5 gene signature, which was validated on a third set of 102 samples. The predictive power of these five genes, which was evaluated using a fourth set of 92 samples, correctly identifying 88 % (30 of 34) of CRC samples and only 64 % (27 of 42)

of non-CRC samples. The intermediate zone contained 16 samples. The performance parameters are compiled into table A.

Table A: Estimates and Exact Confidence Limits of GeneNews ColonSentry™ Test

Performance Parameter	N_C	N	Estimate	Exact Two-Sided 95 % CI-Limits	
				Lower	Upper
Sensitivity	30	34	0.88	0.725	0.967
Specificity	27	42	0.64	0.480	0.784
Positive Predictive Value	30	45	0.67	0.510	0.800
Negative Predictive Value	27	31	0.87	0.702	0.964
Correct Classification Rate	57	76	0.75	0.637	0.842

N_C = Number of correctly classified cases; CI = Confidence interval; Exact confidence limits were computed using the proc FREQ of the statistics program SAS.

Provided that patients of the last validation set were a random selection of the screening population, applying the performance on a hypothetical set of 10,000 patients with an incidence of one percent and computing the performance parameters of this test yields the results shown in table B.

Table B: Estimates and Exact Confidence Limits of GeneNews ColonSentry™ Test applied to a Hypothetical Set of 10000 Persons

Performance Parameter	N_C	N	Estimate	Exact Two-Sided 95 % CI-Limits	
				Lower	Upper
Sensitivity	73	100	0.73	0.632	0.814
Specificity	6979	9900	0.70	0.696	0.714
Positive Predictive Value	73	2994	0.02	0.019	0.031
Negative Predictive Value	6979	7006	1.00	0.994	0.997
Correct Classification Rate	7052	10000	0.71	0.696	0.714

Nc = Number of correctly classified cases; CI = Confidence interval; Exact confidence limits were computed using the proc FREQ of the statistics program SAS.

However, based on the data provided by Han, 1,739 of 10,000 patients would have an "intermediate result". In clinical practice, this would mean that these 1,739 patients would have to undergo colonoscopy to clarify their state. However, in this computation, these 1,739 patients were regarded as having been predicted as low risk. The main disadvantage of the ColonSentry test is its relatively low sensitivity of 73 % and its low specificity of 70 %.

Applied to a screening population of 1 million individuals this means that 2,700 individuals with undetected CRC will not be detected by the test. In addition, 300,000 individuals (30 %) are diagnosed as false positive, which need to be followed up by colonoscopy. The combination of a relatively high false negative rate of 27 % with a high false positive rate of 30 % reduces the clinical utility of this test and impedes acceptance by the medical community and the screening population itself.

Epigenomics AG, Germany, has a CE-marked test, Epi proColon®, in the market that measures the methylation status of the Septin-9 gene and is based on detection of free somatic tumor DNA in blood serum.

Table C: Estimates and Exact Confidence Limits of Epi proColon® Test

Performance Parameter	N_C	N	Estimate	Exact Two-Sided 95% CI-Limits	
				Lower	Upper
Sensitivity	69	103	0.67	0.570	0.759
Specificity	135	154	0.88	0.814	0.924
Positive Predictive Value	69	88	0.78	0.684	0.865
Negative Predictive Value	135	169	0.80	0.730	0.856
Correct Classification Rate	204	257	0.79	0.739	0.842

N_C = Number of correctly classified cases; CI = Confidence interval; Exact confidence limits were computed using the proc FREQ of the statistics program SAS.

The product performance figures of Epi proColon® displayed in table C are cited from the companies' website (www.epigenomics.com). Table D shows the figures when the performance of the test is applied to a hypothetical screening population. Though the Epi proColon® test performs better than GeneNews' test in some performance parameters its overall sensitivity for all four stages of CRC is only 67 %. This means that if 10,000 individuals are screened and the prevalence of CRC in the screening population is approximately 1 %, so that 33 individuals with CRC will be missed by the test. Applied to a screening population of 1 million individuals (3.7 % of the German screening population or 1.3% of the US screening population) 3,300 individuals with CRC will not be detected by the test. This high false negative rate limits significantly the clinical utility of the Epi proColon® test. The false negative rate of patients with early stage CRC (UICC I and II) that will be missed is even higher.

Table D: Estimates and Exact Confidence Limits of Epi proColon® Test applied to a Hypothetical Set of 10,000 Persons

Performance Parameter	N_C	N	Estimate	Exact Two-Sided 95% CI-Limits	
				Lower	Upper
Sensitivity	67	100	0.67	0.569	0.761
Specificity	8679	9900	0.88	0.870	0.883
Positive Predictive Value	67	1288	0.05	0.041	0.066
Negative Predictive Value	8679	8712	1.00	0.995	0.997
Correct Classification Rate	8746	10000	0.87	0.868	0.881

N_C = Number of correctly classified cases; CI = Confidence interval; Exact confidence limits were computed using the proc FREQ of the statistics program SAS.

Another blood-based test developed by OncoMethylome Science (Liege, Belgium) measures the methylation status of two Genes FOXE1 and SYNE1. The sensitivity of this two-marker test for all four stages of CRC is 56 %, while the specificity is 91 % (ESMO meeting, Berlin, Germany, September 2009)

All three blood tests have a significant false negative rate and do not detect a significant number of patients with CRC. The ColonSentry test has a low specificity of 70 % and burdens colonoscopy facilities with a high number of false-positive test results.

Thus, there is a clear clinical need for an improved blood-based test for screening, detecting, or diagnosing colorectal cancer with high sensitivity and high specificity, which is minimally invasive so as to permit more widespread testing of the population to indicate the presence of colorectal cancer with high accuracy and therefore with a high clinical utility, and to ensure greater adherence to recommended protocols. Further, the identification of biomarkers, such as RNAs for use in such a minimally-invasive test would fulfill a longstanding need in the art.

Brief description of the invention

The present invention provides methods and kits for diagnosing, detecting, and screening for colorectal cancer. Particularly, the invention provides for preparing RNA expression profiles of patient blood samples, the RNA expression profiles being indicative of the presence or

absence of colorectal cancer. The invention further provides for evaluating the patient RNA expression profiles for the presence or absence of one or more RNA expression signatures that are indicative of colorectal cancer.

The inventors have surprisingly found that a sensitivity of at least 79 %, and a specificity of at least 89 %, is reached if and only if at least 250 RNAs are measured that are chosen from the RNAs listed in table 1. In other words, measuring 250 RNAs is necessary and sufficient for the detection of colon cancer in a human subject based on RNA from a blood sample obtained from said subject by measuring the abundance of at least 250 RNAs in the sample, that are chosen from the RNAs listed in table 1, and concluding based on the measured abundance whether the subject has colon cancer or not.

In one aspect, the invention provides a method for preparing RNA expression profiles that are indicative of the presence or absence of colorectal cancer. The RNA expression profiles are prepared from patient blood samples. The number of transcripts in the RNA expression profile may be selected so as to offer a convenient and cost effective means for screening samples for the presence or absence of colorectal cancer with high sensitivity and high specificity. Generally, the RNA expression profile includes the expression level or "abundance" of from 250 to about 1500 transcripts. In certain embodiments, the expression profile includes the RNA levels of 1000 transcripts or less, 800 transcripts or less, 500 transcripts or less.

In such embodiments, the profile may contain the expression level of at least 250 RNAs that are indicative of the presence or absence of colorectal cancer, and specifically, as selected from table 1, or may contain the expression level of at least 250, at least 500, at least 800, or of 1000 RNAs selected from table 1. Where larger profiles are desired, the profile may contain the expression level or abundance of at least about 1200 RNAs that are indicative of the presence or absence of colorectal cancer, and such RNAs may be selected from table 1. The identities and/or combinations of genes and/or transcripts that make up or are included in expression profiles are disclosed in table 1. In particular embodiments, the genes or transcripts include those listed in tables 2 to 4.

Such RNA expression profiles in accordance with this aspect may be evaluated for the presence or absence of an RNA expression signature indicative of colorectal cancer. Generally, the sequential addition of transcripts from table 1 to the expression profile provides for higher sensitivity and/or specificity for the detection of colorectal cancer. For example, the sensitivity of the methods provided herein may be at least 75 %, or at least 80 %, or at least 85 %, or at least 90 %. The specificity of the method may be at least 85 %, or at least 90 %, or at least 95%.

In a second aspect, the invention provides a method for detecting, diagnosing, or screening for colorectal cancer. In this aspect the method comprises preparing an RNA expression profile by measuring the abundance of at least 250, at least 500, or at least 800 RNAs in a patient blood sample, where the abundance of such RNAs are indicative of the presence or absence of colorectal cancer. The RNAs may be selected from the RNAs listed in table 1, and exemplary sets of such RNAs are disclosed in tables 2 to 4. The method further comprises evaluating the profile for the presence or absence of an RNA expression signature indicative of colorectal cancer, to thereby conclude whether the patient has or does not have colorectal cancer. The method generally provides a sensitivity for the detection of colorectal cancer of at least about 75 %, while providing a specificity of at least about 85 %.

In various embodiments, the method comprises determining the abundance of at least 250 RNAs, at least 500 RNAs, at least 800 RNAs, or of at least 1000 RNAs chosen from the RNAs listed in table 1, and as exemplified in tables 2 to 4, and classifying the sample as being indicative of colorectal cancer, or not being indicative of colorectal cancer.

In other aspects, the invention provides kits and custom arrays for preparing the gene expression profiles, and for determining the presence or absence of colorectal cancer.

Detailed description of the invention

The invention provides methods and kits for screening, diagnosing, and detecting colorectal cancer in human patients (subjects). "Colorectal cancer" (CRC) refers to both colorectal adenoma and colorectal carcinoma.

A colorectal adenoma is characterized by atypical growth of epithelial cells in mucosal tissue, i.e. neoplasia. Hypercellularity with enlarged, hyperchromatic nuclei, varying degrees of nuclear stratification, nuclear polymorphisms, and loss of polarity are the histologically defining features. In colorectal adenoma, this abnormal growth is confined to the mucosa; a synonym of adenoma is intraepithelial neoplasia (IEN). If this atypical growth of epithelial cells extends/invades through the muscularis mucosae, the muscle layer under the mucosa, with destruction of the usual anatomical wall, the pathologist terms this atypical growth a colorectal carcinoma.

The distinction between high- (HIEN) and low-grade (LIEN) intraepithelial neoplasia refers to the extent of the defining features.

A patient with CRC is traditionally defined as having undergone surgery/resection of colon and/or rectum for CRC and whose resection specimen has undergone examination by a board certified pathologist, who has diagnosed a colorectal carcinoma as defined above. A patient with CRC may have undergone complete colonoscopy of colon and rectum during which the examinating physician has taken a sample of suspect tissue, which in turn has undergone examination by a board-certified pathologist, who has diagnosed a colorectal carcinoma as defined above. A synonym for a patient with CRC is "CRC-case" or simply "case."

A patient without CRC is traditionally a person that has undergone complete colonoscopy during which the examining physician, an endoscopist, has noted no abnormal tissue growth. A synonym for a patient without CRC is "non-CRC-case" or "control." This however does not exclude that this person has any other carcinoma.

A patient with HIEN is traditionally a person that has undergone surgery/resection of colon and/or rectum for suspected CRC and whose resection specimen has undergone examination

by a board certified pathologist, who has diagnosed a high-grade intraepithelial neoplasia as defined above. Alternatively, a patient with HIEN may be a person that has undergone complete colonoscopy of colon and rectum during which the examinating physician has taken a sample of suspect tissue, which in turn has undergone examination by a board certified pathologist, who has diagnosed a high-grade intraepithelial neoplasia as defined above. A synonym for a person with HIEN is "HIEN-case" or "HIEN."

A patient with LIEN is traditionally a person that has undergone surgery/resection of colon and/or rectum for suspected CRC and whose resection specimen has undergone examination by a board certified pathologist, who has diagnosed a high-grade intraepithelial neoplasia. Alternatively, a patient with LIEN is a person that has undergone complete colonoscopy of colon and rectum during which the examinating physician has taken a sample of suspect tissue, which in turn has undergone examination by a board certified pathologist, who has diagnosed a low-grade intraepithelial neoplasia as defined above. A synonym for a person with LIEN "is LIEN-case" or "LIEN."

As disclosed herein, the present invention provides methods and kits for screening patient samples for those that are positive for CRC, that is, in the absence of colonoscopy and/or surgery or resection of colon or rectum with pathologist's examination.

The invention relates to the determination of the abundance of RNAs to detect a colorectal cancer in a human subject, wherein the determination of the abundance is based on RNA obtained (or isolated) from whole blood of the subject or from blood cells of the subject. For example, the sample may be obtained using PAXgene (QIAGEN) or an equivalent RNA isolation system. The measurement of the abundance of RNAs in the sample is preferably performed together, i.e. sequentially or preferably simultaneously. The blood sample preferably does not contain cancer cells. Preferably, the sample comprises or consists of white blood cells.

In various aspects, the invention involves preparing an RNA expression profile from a patient sample. The method may comprise isolating RNA from whole blood, and detecting the abundance or relative abundance of selected transcripts. As used herein, the terms RNA "abundance" and RNA "expression" are used interchangeably. The "RNAs" may be defined

by reference to an expressed gene, or by reference to a transcript, or by reference to a particular oligonucleotide probe for detecting the RNA (or cDNA derived therefrom), each of which is listed in table 1 for 1000 RNAs that are indicative of the presence or absence of colorectal cancer. Specifically, table 1 lists such RNAs by probe ID, gene symbol, and nucleotide sequences, which are publicly accessible.

The number of transcripts in the RNA expression profile may be selected so as to offer a convenient and cost effective means for screening samples for the presence or absence of colorectal cancer with high sensitivity and high specificity. For example, the RNA expression profile may include the expression level or "abundance" of from 250 to about 1200 transcripts. In certain embodiments, the expression profile includes the RNA levels of 1000 transcripts or less, 800 transcripts or less, or 500 transcripts or less. Such profiles may be prepared, for example, using custom microarrays or multiplex gene expression assays as described in detail herein.

In such embodiments, the profile may contain the expression level of at least 250 RNAs that are indicative of the presence or absence of colorectal cancer, and specifically, as selected from table 1, or may contain the expression level of at least 500, 800, or 1000 RNAs selected from table 1. Where larger profiles are desired, the profile may contain the expression level or abundance of at least 1200 RNAs that are indicative of the presence or absence of colorectal cancer, and such RNAs may be selected from table 1. Such RNAs may be defined by gene, or by transcript ID, or by probe ID, as set forth in table 1.

The identities of genes and/or transcripts that make up, or are included in exemplary expression profiles are disclosed in tables 1 to 4. As shown herein, profiles selected from the RNAs of table 1 support the detection of colorectal cancer with high sensitivity and high specificity.

The present invention provides an in-vitro diagnostic test system (IVD) that is trained (as described further below) for the detection of a colorectal cancer. For example, in order to determine whether a patient has colorectal cancer, reference RNA abundance values for colon cancer positive and negative samples are determined. The RNAs can be quantitatively measured on an adequate set of training samples comprising cases and controls, and with

adequate clinical information on carcinoma status, applying adequate quality control measures, and on an adequate set of test samples, for which the detection is yet to be made. With such quantitative values for the RNAs and the clinical data for the training samples, a classifier can be trained and applied to the test samples to calculate the probability of the presence or non-presence of the colorectal carcinoma.

Various classification schemes are known for classifying samples between two or more classes or groups, and these include, without limitation: Naïve Bayes, Support Vector Machines, Nearest Neighbors, Decision Trees, Logistics, Articifial Neural Networks, and Rule-based schemes. In addition, the predictions from multiple models can be combined to generate an overall prediction. Thus, a classification algorithm or "class predictor" may be constructed to classify samples. The process for preparing a suitable class predictor is reviewed in R. Simon, Diagnostic and prognostic prediction using gene expression profiles in high-dimensional microarray data, British Journal of Cancer (2003) 89, 1599-1604, which review is hereby incorporated by reference.

In this context, the invention teaches an in-vitro diagnostic test system (IVD) that is trained in the detection of a colorectal cancer referred to above, comprising at least 250 RNAs, which can be quantitatively measured on an adequate set of training samples comprising cases and controls, with adequate clinical information on carcinoma status, applying adequate quality control measures, and on an adequate set of test samples, for which the detection yet has to be made. Given the quantitative values for the RNAs and the clinical data for the training samples, a classifier can be trained and applied to the test samples to calculate the probability of the presence or absence of the colorectal carcinoma.

The present invention provides methods for detecting, diagnosing, or screening for colorectal cancer in a human subject with a sensitivity and specificity not previously described for a blood-based method (see Fig. 1). Specifically, the sensitivity of the methods provided herein is at least 75 %, at least 80 %, at least 85 %, or at least 90 %. The specificity of the methods is at least 85 %, at least 90 %, or at least 95%, for example, when determined with samples of at least 52 patients with CRC and adequate samples (e.g., at least 64) of normal individuals without CRC are tested.

Without wishing to be bound by any particular theory, the above finding may be due to the fact that an organism such as a human systemically reacts to the development of a colorectal tumor by altering the expression levels of genes in different pathways. The formation of cancerous tumor cells from a nonmalignant adenoma or nonmalignant polyps, the formation of high-grade intraepithelial neoplasias and the further growth and development of cancer of different stages may trigger differential expression of genes in white blood cells that are involved in both adaptive and innate immune responses, for example wound healing, inflammatory response and antibody production pathways. Although the change in expression (abundance) might be small for each gene in a particular signature, measuring a set of at least 250 genes, preferably even larger numbers such as 500, 800 or even more RNAs, for example at least 1000 RNAs at the same time allows for the detection of colorectal cancer in a human with high sensitivity and high specificity.

In this context, an RNA obtained from a subject's blood sample, i.e. an RNA biomarker, is an RNA molecule with a particular base sequence whose presence within a blood sample from a human subject can be quantitatively measured. The measurement can be based on a part of the RNA molecule, namely a part of the RNA molecule that has a certain base sequence, which allows for its detection and thereby allows for the measurement of its abundance in a sample. The measurement can be by methods known in the art, for example analysis on a solid phase device, or in solution (for example, by RT-PCR). Probes for the particular RNAs can either be bought commercially, or designed based on the respective RNA sequence.

In the method of the invention, the abundance of several RNA molecules (e.g. mRNA or pre-spliced RNA, intron-lariat RNA, micro RNA, small nuclear RNA, or fragments thereof) is determined in a relative or an absolute manner, wherein an absolute measurement of RNA abundance is preferred. The RNA abundance is, if applicable, compared with that of other individuals, or with multivariate quantitative thresholds.

The determination of the abundance of the RNAs described herein is performed from blood samples using quantitative methods. In particular, RNA is isolated from a blood sample obtained from a human subject that is to undergo CRC testing. Although the examples described herein use microarray-based methods, the invention is not limited thereto. For example, RNA abundance can be measured by in situ hybridization, amplification assays such

as the polymerase chain reaction (PCR), sequencing, or microarray-based methods. Other methods that can be used include polymerase-based assays, such as RT-PCR (e.g., TAQMAN), hybridization-based assays, such as DNA microarray analysis, as well as direct mRNA capture with branched DNA (QUANTIGENE) or HYBRID CAPTURE (DIGENE).

In certain embodiments, the invention employs a microarray. A "micoroarray" includes a specific set of probes, such as oligonucleotides and/or cDNAs (*e.g.*, expressed sequence tags, "ESTs") corresponding in whole or in part, and/or continuously or discontinuously, to regions of RNAs that can be extracted from a blood sample of a human subject. The probes are bound to a solid support. The support may be selected from beads (magnetic, paramagnetic, etc.), glass slides, and silicon wafers. The probes can correspond in sequence to the RNAs of the invention such that hybridization between the RNA from the subject sample (or cDNA derived therefrom) and the probe occurs. In the microarray embodiments, the sample RNA can optionally be amplified before hybridization to the microarray. Prior to hybridization, the sample RNA is fluorescently labeled. Upon hybridization to the array and excitation at the appropriate wavelength, fluorescence emission is quantified. Fluorescence emission for each particular nucleic acid is directly correlated with the amount of the particular RNA in the sample. The signal can be detected and together with its location on the support can be used to determine which probe hybridized with RNA from the subject's blood sample.

Accordingly, in certain aspects, the invention is directed to a kit or microarray for detecting the level of expression or abundance of RNAs in the subject's blood sample, where this "profile" allows for the conclusion of whether the subject has colorectal cancer or not (at a level of accuracy described herein). In another aspect, the invention relates to a probe set that allows for the detection of the RNAs associated with CRC. If these particular RNAs are present in a sample, they (or corresponding cDNA) will hybridize with their respective probe (i.e, a complementary nucleic acid sequence), which will yield a detectable signal. Probes are designed to minimize cross reactivity and false positives. In one embodiment, the probes used are given e.g. in table 1 as so-called NimbleGen probe set ID numbers. A NimbleGen probe set ID number is an identifier that refers to a set of probes selected to represent expressed sequences on an array. A NimbleGen probe set ID number identifies each probe present on the array, as known to a person of skill in the art. From the sequence defined by an NimbleGen probe set ID number, the sequence of a nucleic acid, such as an RNA or a

molecule derived therefrom (such as cDNA) hybridizing with the probe can be deduced.

Thus, the invention in certain aspects provides a microarray, which generally comprises a solid support and a set of oligonucleotide probes. The set of probes generally contains from 250 to about 3,000 probes, including at least 250 probes selected from table 1. In certain embodiments, the set contains 1200 probes or less, or 800 probes or less. In various embodiments, at least 250 probes are listed in table 1. The set of probes includes probes that hybridize to a combination of RNAs exemplified in any one of table 2, table 3, or table 4. The microarray may comprise, e.g., about 100.000 probes, some of which, usually about 10.000 to 100.000 probes, may be probes for providing reference data.

The conclusion whether the subject has colorectal cancer or not is preferably reached on the basis of a classification algorithm, which can be developed using e.g. a random forest method, a support vector machine (SVM), or a K-nearest neighbor method (K-NN), such as a 3-nearest neighbor method (3-NN), as known in the art.

From the cross-classification of the true disease state (Positive = patient with CRC and Negative = patient without CRC) as determined by a physician and the test result as determined by the classification algorithm, the following measures for binary tests can be derived (Sullivan MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press, 2003), see table E. An example is given in table F.

Table E: Cross-Classification of True Disease State by Test Result

Test Result	True Disease State		Total
	Negative	Positive	
Negative	n_{11}	n_{12}	$n_{1\Sigma}$
Positive	n_{21}	n_{22}	$n_{2\Sigma}$
Total	$n_{\Sigma 1}$	$n_{\Sigma 2}$	$n_{\Sigma\Sigma}$

Table F: Example of a Cross-Classification of True Disease State by Test Result

Test Result	True Disease State		Total
	Negative	Positive	
Negative	30	10	40
Positive	20	70	90
Total	50	80	130

"Sensitivity" (S^+ or true positive fraction (TPF)) refers to the count of positive test results among all true positive disease states divided by the count of all true positive disease states; in terms of table E this reads: $S^+ = n_{22} / n_{\sum 2}$; the result form table F would read: $S^+ = 70 / 80 = 0.875$. "Specificity" (S^- or true negative fraction (TNF)) refers to the count of negative test results among all true negative disease states divided by the count of all true negative disease states; in terms of table E this reads: $S^+ = n_{11} / n_{\sum 1}$; the result form table F would read: $S^- = 30 / 50 = 0.6$. "Correct Classification Rate" (CCR or true fraction (TF)) refers to the sum of the count of positive test results among all true positive disease states and count of negative test results among all true negative disease states divided by all the sum of all cases; in terms of table E this reads: $CCR = (n_{11} + n_{22}) / n_{\sum\sum}$; the result form table F would read: $CCR = (30 + 70) / 130 \approx 0.769230769$. The measures S^+, S^-, and CCR address the question: To what degree does the test reflect the true disease state?

"Positive Predictive Value" (PV^+ or PPV) refers to the count of true positive disease states among all positive test results dived by the count of all positive test results; in terms of table E this reads: $PV^+ = n_{22} / n_{2\sum}$; the result form table F would read: $PV^+ = 70 / 90 \approx 0.777777778$. "Negative Predictive Value" (PV^- or NPV) refers to the count of true negative disease states among all negative test results dived by the count of all negative test results; in terms of table E this reads: $PV^- = n_{11} / n_{1\sum\sum}$; the result form table F would read: $PV^- = 30 / 40 = 0.75$. The predictive values address the question: How likely is the disease given the test results?

Exact or asymptotic confidence limits (CI) for these rates or fractions can be computed using the commercially available software package SAS (SAS Institute Inc., Cary, NC, USA; www.sas.com) or the publicly available software package R (www.r-project.org) (for literature reference see: Agresti A, Caffo B. Simple and effective confidence intervals for proportions and differences of proportions from adding two successes and two failures. The American Statistician: 54: 280-288, 2000).

The preferred RNA molecules that can be used in combinations described herein for diagnosing and detecting colorectal cancer in a subject according to the invention can be found in table 1. The inventors have shown that the selection of at least 250 or more RNAs of the markers listed in table 1 can be used to diagnose or detect colorectal cancer in a subject using a blood sample from that subject. The RNA molecules that can be used for detecting, screening and diagnosing colorectal cancer are selected from the RNAs provided in table 2, 3, or 4. Also, the RNAs (e.g., at least 250, at least 500, at least 800, or more) can be selected from table 3.

Specifically, the method of the invention comprises at least the following steps: measuring the abundance of at least 250 RNAs (preferably 500 RNAs) in the sample, that are chosen from the RNAs listed in table 1, and concluding, based on the measured abundance, whether the subject has colorectal cancer or not. Measuring the abundance of RNAs may comprise isolating RNA from blood samples as described, and hybridizing the RNA or cDNA prepared therefrom to a microarray. Alternatively, other methods for determining RNA levels may be employed.

Examples for sets of 250 RNAs that are measured together, i.e. sequentially or preferably simultaneously, are shown in table 2. The sets of 250 RNAs of table 2 are defined by a common threshold of sensitivity of at least 84 % and specificity of at least 93 %.

In a preferred embodiment of the invention as mentioned herein, the abundance of at least 500 RNAs (preferably 800 RNAs) in the sample is measured, wherein the at least 500 RNAs are chosen from the RNAs listed in table 1. Examples for sets of 500 RNAs that can be measured together, i.e. sequentially or preferably simultaneously, to detect colorectal cancer in a human subject are shown in table 3. The sets of 500 RNAs of table 3 are defined by a common threshold of sensitivity of at least 86 % and specificity of at least 95 %.

Similarly, the abundance of at least 800 RNAs (preferably up to 1000 RNAs) that are chosen from the RNAs listed in table 1 can be measured in the method of the invention. The sets of 800 RNAs of table 4 are defined by a common threshold of sensitivity of at least 87 % and specificity of at least 97 %.

Tables 2 through 4 show examples of combinations from table 1, varying in size from 250 to 800 sequences. Examples 1, 2, and 3 each consist of 250 sequences (table 2). Examples 4, 5, and 6 each consist of 500 sequences (table 3). Examples 7, 8, and 9 (table 4) each consist of 800 sequences.

The performance of the examples 1 to 9 was evaluated by applying them to the discovery set, split randomly into a learning and a test set, 1000 times. Each time, a random 85 % of the discovery set were used to train a classifier based on the expression profile of the example sequences. This classifier was then applied to the remaining 15 % as a test set. The results across all repetitions were averaged and are given in the following table:

	Example 1	Example 2	Example 3
Sensitivity	87.0%	86.9%	84.6%
Specificity	95.4%	95.8%	93.2%
	Example 4	Example 5	Example 6
Sensitivity	90.1%	86.6%	86.4%
Specificity	96.0%	95.5%	95.3%
	Example 7	Example 8	Example 9
Sensitivity	88.2%	88.2%	87.7%
Specificity	98.0%	97.8%	97.8%

Examples 1, 2, 4, 5, 7, and 8 were chosen based on their performance in this retrospective evaluation. Examples 3, 6, and 9 were chosen during the discovery procedure as consensus over all outer loops. In that procedure, a prospective estimate for sequences of length 250 was observed of 79.3% sensitivity and 89.2% specificity; for length 500: 79.5% sensitivity and 90.6% specificity; for length 800: 79.1% sensitivity and 89.6% specificity.

When all 1000 sequences are evaluated in this retrospective fashion, a sensitivity of 88.2 % and a specificity of 97.7 % were seen. In the discovery procedure, prospective estimates of 78.8 % sensitivity and 90.2 % specificity were observed.

When the wording "at least a number of RNAs" is used, this refers to a minimum number of RNAs that are measured. It is possible to use up to 10,000 or 20,000 genes in the invention, a fraction of which can be RNAs listed in table 1. In preferred embodiments of the invention, abundance of up to 5,000, 2,500, 2,000, 1,000 RNAs of randomly chosen RNAs that are not listed in table 1 is measured in addition to RNAs of table 1 (or subsets thereof).

In a preferred embodiment, only RNAs that are mentioned in table 1 are measured.

The expression profile or abundance of RNA markers for colorectal cancer, for example the at least 250 RNAs described above, (or more RNAs as disclosed above and herein), is determined preferably by measuring the quantity of the transcribed RNA of the marker gene. This quantity of the mRNA of the marker gene can be determined for example through chip technology (microarray), (RT-) PCR (for example also on fixated material), Northern hybridization, dot-blotting, sequencing, or in situ hybridization.

The microarray technology, which is most preferred, allows for the simultaneous measurement of RNA abundance of up to many thousand RNAs and is therefore an important tool for determining differential expression (or differences in RNA abundance), in particular between two biological samples or groups of biological samples. In order to apply the microarray technology, the RNAs of the sample need to be amplified and labeled and the hybridization and detection procedure can be performed as known to a person of skill in the art.

As will be understood by those of ordinary skill in the art, the analysis can also be performed through single reverse transcriptase-PCR, competitive PCR, real time PCR, differential display RT-PCR, Northern blot analysis, sequencing, and other related methods. In general, the larger the number of markers is that are to be measured, the more preferred is the use of the microarray technology. However, multiplex PCR, for example, real time multiplex PCR is known in the art and is amenable for use with the present invention, in order to detect the presence of 2 or more genes or RNA simultaneously.

The RNA whose abundance is measured in the method of the invention can be mRNA, cDNA, unspliced RNA, or its fragments. Measurements can be performed using the

complementary DNA (cDNA) or complementary RNA (cRNA), which is produced on the basis of the RNA to be analyzed, e.g. using microarrays. A great number of different arrays as well as their manufacture are known to a person of skill in the art and are described for example in the U.S. Patent Nos. 5,445,934; 5,532,128; 5,556,752; 5,242,974; 5,384,261; 5,405,783; 5,412,087; 5,424,186; 5,429,807; 5,436,327; 5,472,672; 5,527,681; 5,529,756; 5,545,331; 5,554,501; 5,561,071; 5,571,639; 5,593,839; 5,599,695; 5,624,711; 5,658,734; and 5,700,637, each of which is hereby incorporated by reference in its entirety.

Preferably the decision whether the subject has colon cancer comprises the step of training a classification algorithm on an adequate training set of cases and controls and applying it to RNA abundance data that was experimentally determined based on the blood sample from the human subject to be diagnosed. The classification method can be a random forest method, a support vector machine (SVM), or a K-nearest neighbor method (K-NN), such as 3-NN.

For the development of a model that allows for the classification for a given set of biomarkers, such as RNAs, methods generally known to a person of skill in the art are sufficient, i.e., new algorithms need not be developed.

The major steps of such a model are:
1) condensation of the raw measurement data (for example combining probes of a microarray to probeset data, and/or normalizing measurement data against common controls);
2) training and applying a classifier (i.e. a mathematical model that generalizes properties of the different classes (carcinoma vs. healthy individual) from the training data and applies them to the test data resulting in a classification for each test sample.

For example, the raw data from microarray hybridizations can be.quantile-normalized after forming the median of multiple identical probes and subtracting the local background (see Bolstad,Irizarry,Anstrand, Speed: A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics, 19 (2):185-193). Alternative methods for condensation can be used, such as: FARMS as shown by Hochreiter (2006, Bioinformatics 22(8): 943-9), RMA, GC-RMA, see Irizarry et al (2003). Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. Biostatistics. *4*, 249–264.). Similar to condensation, classification of the test data set through a

support-vector-machine or other classification algorithms is known to a person of skill in the art, like for example classification and regression trees, penalized logistic regression, sparse linear discriminant analysis, Fisher linear discriminant analysis, K-nearest neighbors, shrunken centroids, and artificial neural networks (see Wladimir Wapnik: The Nature of Statistical Learning Theory, Springer Verlag, New York, NY, USA, 1995; Berhard Schölkopf, Alex Smola: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, Cambridge, MA, 2002; S. Kotsiantis, Supervised Machine Learning: A Review of Classification Techniques, Informatica Journal 31 (2007) 249-268).

The key component of these classifier training and classification techniques is the choice of RNA biomarkers that are used as input to the classification algorithm.

In a further aspect, the invention refers to the use of a method as described above and herein for the detection of colorectal cancer in a human subject, based on RNA from a blood sample.

In a further aspect, the invention also refers to the use of a microarray for the detection of colorectal cancer in a human subject based on RNA from a blood sample. According to the invention, such a use can comprise measuring the abundance of at least 250 RNAs (or more, as described above and herein) that are listed in table 1. Accordingly, the microarray comprises at least 250 probes for measuring the abundance of the at least at least 250 RNAs. It is preferred that the microarray has a set of 11 probes for each RNA, and 1 or 3 probes for each RNA are also preferred. Commercially available microarrays, such as from Roche NimbleGen, Inc., may be used. Alternatively, at most 250, at most 500, or at most 800 RNAs are measured in a sample, in order to detect or diagnose CRC.

In another embodiment, the abundance of the at least 250 RNAs is measured by multiplex RT-PCR. In a further embodiment, the RT-PCR includes real time detection, e.g., with fluorescent probes such as Molecular beacons or TaqMan® probes.

In a preferred embodiment, the microarray comprises probes for measuring only RNAs that are listed in table 1 (or subsets thereof).

In yet a further aspect, the invention also refers to a kit for the detection of colorectal cancer in a human subject based on RNA obtained from a blood sample. Such a kit comprises a means for measuring the abundance of at least 250 RNAs that are chosen from the RNAs listed in table 1. In a further embodiment, the at least 250 RNAs are chosen from the RNAs listed in any of the tables provided herein, for example, the RNAs are chosen from tables 2 to 4. The means for measuring expression can be probes that allow for the detection of RNA in the sample or primers that allow for the amplification of RNA in the sample. Ways to devise probes and primers for such a kit are known to a person of skill in the art.

Further, the invention refers to the use of a kit as described above and herein for the detection of colorectal cancer in a human subject based on RNA from a blood sample comprising means for measuring the abundance of at least 250 RNAs that are chosen from the RNAs listed in table 1. Such a use may comprise the following steps: contacting at least one component of the kit with RNA from a blood sample from a human subject, measuring the abundance of at least 250 RNAs (or more as described above and herein) that are chosen from the RNAs listed in table 1 using the means for measuring the abundance of at least 250 RNAs, and concluding, based on the measured abundance, whether the subject has colorectal cancer.

In yet a further aspect, the invention also refers to a method for preparing an RNA expression profile that is indicative of the presence or absence of colorectal cancer, comprising: isolating RNA from a whole blood sample, and determining the level or abundance of from 250 to about 3000 RNAs, including at least 250 RNAs selected from table 1.

Preferably, the expression profile contains the level or abundance of 800 RNAs or less, of 500 RNAs or less. Further, it is preferred that at least 250 RNAs, at least 500 RNAs, at least 800 RNAs are listed in table 1 or in tables 2 to 4.

Another preferred embodiment of the method comprises determining the presence or absence of an RNA expression signature indicative of colorectal cancer.

In yet a further aspect, the invention also refers to a microarray, comprising a solid support and a set of oligonucleotide probes, the set containing from 250 to about 3,000 probes, and including at least 250 probes selected from table 1. Preferably, the set contains 2000 probes

or less, 1000 probes or less. At least 250 probes can be those listed in table 1 or table 2. At least 500 probes can be those listed in table 1 or table 3. In another embodiment, at least 800 probes are listed in table 1 or table 4.

Features of the invention that were described herein in combination with a method, a microarray, a kit, or a use also refer, if applicable, to all other aspects of the invention.

Tables and Figures

Table 1 shows a list of 1000 RNAs that are differentially expressed in several human subjects with colorectal cancer in comparison to subject without colorectal cancer. Each marker is characterized by a SEQ ID NO., by a NimbleGen probe set ID, which is also the primary transcript. Seq ID.

Table 1A (shows the changes of expression level (abundance) between cases and controls for all of 1000 RNAs listed in table 1. The data was derived from 116 samples, 52 CRC cases and 64 controls. In the second column, the log-2 fold change is shown, and in the third column, the non-log fold change is shown. The numbers in the second column represent the differences of cases (events) and controls (non-envents) in log2 steps. For example, a value of 0.08 for the first probe set (SEQ ID NO. 1) means an expression increase by 2^0.08-fold for events vs. non-events. In other words, the events had a 106 % expression (see third column) with respect to non-events, i.e. 6 % more RNA (0.06 = 2^0.08-1, 1.08 = 2^0.08).

Tables 2 to 4 show examples of combinations from table 1, varying in size from 250 to 800 sequences. Examples 1, 2, and 3 each consist of 250 sequences (table 2). Examples 4, 5, and 6 each consist of 500 sequences (table 3). Examples 7, 8, and 9 (table 4) each consist of 800 sequences.

Figure 1 is a graph showing the performance of various RNA colorectal cancer signatures as a function of their length. The x-axis shows the number of RNAs in each particular signature subset. The subset length is varied along the x-axis from 250 to 1000 (the full set). The y-axis shows performance of the subset. Performance is shown in terms of sensitivity (percentage of real carcinomas that were properly classified; lower values (crosses)) and specificity (percentage of real healthy controls that were properly classified; upper values (diamonds)) scaled from 80 % to 100 %, with standard errors. These are retrospective examinations of the 1000 RNA full set. As the figure shows, reduced sets compared to the full set appear to have excellent performance. For each subset size, the performance was measured for 100 randomly chosen subsets based on leave-15 %-out runs: the subset is trained on 85 % and applied to the remaining 15 %. The average performance over all the 15 % tests is shown.

Examples

Materials and Methods

Study Protocol

All results described herein are based on a prospective, clinical-diagnostic study protocol entitled Früherkennung kolorektaler Karzinome mit Hilfe von RNA-basierten Expression-Signaturen aus Blut - Eine multizentrische, diagnostische Studie zur Entwicklung und Validierung von Genexpressions-Signaturen zur Früherkennung kolorektaler Karzinome - Version CRC.SCR.2.BR.1 vom 06. Januar 2009" in accordance with the Guideline for Good Clinical Practice (Directive 75/318/EEC) July 1996, version July 2002 (http://www.emea.eu.int/pdfs/human/ich/013595en.pdf). This study protocol was reviewed and approved by the local ethics authority, the "Ethik-Kommission der Landesärztekammer Brandenburg" on January 14th 2009. All persons entered into this study gave written informed consent that their blood samples and associated clinical data could be used for this research endeavor. Moreover, the persons gave written, informed consent that samples and clinical data could be audited by regulatory authorities if the research results would be used in a performance evaluation study (according to German law (Gesetz über Medizinprodukte)).

This study was designed as a cohort study. One cohort, the colonoscopy cohort, included persons undergoing colonoscopy. The second cohort, the surgery cohort, included patients scheduled for surgery for suspected colorectal carcinoma.

The inclusion criteria for the colonoscopy cohort were: 1) Willingness to undergo complete colonoscopy; 2) At least 55 years of age; 3) Ability to give written, informed consent; 4) Written informed consent after information about the study was given by a physician. The exclusion criteria for the colonoscopy cohort were: 1) Rectoscopy, sigmoidoscopy, or colonoscopy during the last five years prior to inclusion into the study; 2) Treatment of a malignant disease during the last five years prior to inclusion into the study, except for malignoma with low metastatic potential such as basalioma in situ of the skin.

The inclusion criteria for the surgery cohort were: 1) Age at initial diagnosis at least 18 years of age; 2) Ability to give written, informed consent; 3) (Suspected) Diagnosis of colorectal

carcinoma UICC-stage I to IV; 4) Surgery is planned in such a fashion that staging according to UICC-criteria is feasible; 5) No treatment prior to surgery; 6) No treatment for a malignant disease during the last five years prior to inclusion into the study; 7) No other disease that lowers life expectancy below one year; 8) Regular follow-up examinations have to be possible; 9) Written informed consent after information about the study was given by a physician.

Blood Drawing and Blood Sample Storage

In the colonoscopy cohort, blood was drawn after written, informed consent and prior to bowel cleaning and change of medication if necessary. In the surgery cohort, blood was drawn after written, informed consent and prior to removal of the resected tissue from the body of the patient by the surgeon. In both cohorts, colonoscopy cohort and surgery cohort, blood was drawn into PAXgene™ blood tubes (Becton Dickinson). The tubes were stored within four hours in a freezer at -20 °C until transport to the laboratory on dry ice, where the tubes were stored again in a freezer at -20 °C until RNA-extraction.

Sample Sizes

The study protocol initially stipulated the use of 220 blood samples from the colonoscopy cohort and 220 blood samples from the surgery cohort for the discovery of the signatures. However, the study protocol was open to changes depending on the results of the discovery process. Additionally, the study protocol initially stipulated the use 220 blood samples from the colonoscopy cohort and 220 blood samples from the surgery cohort for prospective performance evaluation purposes. Again, these samples sizes were open and amenable to change.

Quality Control of Clinical Data

All clinical data of all included persons from both cohorts were checked directly from the patient's medical records and entered into study databases, as prescribed in the study protocol. Each item in the study databases were checked independently by two persons. Patients of both cohorts together with their blood samples were withdrawn if any violation of the inclusion of exclusion criteria cited above was detected. In particular for the colonoscopy cohort, all samples and their associated clinical data were excluded from all analyses if the

colonoscopy was found to be not complete. Moreover, blood samples were destroyed and clinical data were deleted if the patient decided to withdraw his/her informed consent.

Recruitment

499 persons scheduled for colonoscopy were recruited into the study; RNA blood samples were taken prior to colonoscopy. The recruitment period for the non-CRC patients lasted from February 9, 2009 until April 3, 2009. In parallel, RNA blood samples were taken from patients with (suspected) diagnosis of CRC prior or during surgery for CRC.

Sample Selection for RNA-Extraction

The selection criteria of CRC-cases included availability of UICC stage information and the check that the patients did not fulfill any exclusion criteria for the MSKK-study. At this time clinical data of the non-CRC case were not available. Hence, the first 240 patients of the study were selected.

Altogether, 480 PAX-RNA samples, 240 from non-CRC persons and 240 from CRC patients, were randomized into the discovery set. Additionally, the order of processing by the laboratory was randomized. A drop-out rate of 10 %, or 20 cases in each group, was assumed.

RNA extraction

Total RNA extraction from blood was performed using the QIAGEN PAXgene Blood miRNA Kit together with the QIAcube® robot according to the manufacturer's instructions. The RNA obtained is therefore sometimes referred to as PAX-RNA.

Before starting RNA extraction, the buffers BM3 and BM4 were mixed with the specified amount of 96-100 % Ethanol. Furthermore, the DNAse I (1500 Kunitz units) was dissolved in 550 µl of RNAse free water.

After thawing, the PAX-blood tubes were turned several times to ensure proper mixing and left over night (or minimal two hours) at room temperature. Then, the tubes were centrifuged for 10 minutes at 4000 x g using a swing out bucket. Next, the supernatant was carefully removed from each tube, and the pellets were washed (vortexing until complete resuspension) with 4 ml RNAse free water. After centrifuging again for 10 minutes at 4000 x g, the pellets

were resuspended in 350 µl buffer BR1 and transferred to 2 ml processing tubes. The opened tubes were then loaded into the rotor adapters of the centrifuge. Subsequently, all buffers were placed on the respective spots of the reagent bottle holder. After closing the lid, the PAXgene Blood miRNA Part A protocol was started. When it was finished, the instrument door was opened, the RNA containing tubes were closed and subsequently placed on the shaker adaptor. After closing the instrument door again, the PAXgene Blood miRNA Part B protocol was started. When the program was finished, concentration was determined by UV-absorption measurement and the samples were stored at -70 °C until use.

For understanding of underlying principles of the automatic procedure the manual protocol is briefly described below. There is no difference between the two methods until the resuspension of the pellet.

To the resupended pellet 300 µl binding buffer (BR2) and 40 µl proteinase K solution was added and mixed by vortexing for 5 seconds. Incubation follows for 10 minutes at 55 °C using a shaker–incubator at 400–1400 rpm. After incubation, the temperature of the shaker–incubator is set to 65 °C. The lysate is pipetted directly into a PAXgene Shredder spin column (lilac) placed in a 2 ml processing tube and centrifuged for 3 minutes at maximum speed (don't not to exceed 20,000 x g). The entire supernatant of the flow-through fraction was carefully transferred the to a fresh 1.5 ml microcentrifuge tube without disturbing the pellet in the processing tube. 350 µl ethanol (96–100%, purity grade p.a.) was added and mixed by vortexing. The mixture is briefly (1–2 seconds at 500–1000 x g) centrifuged to remove drops from the inside of the tube Lid. 700 µl of the sample is pipetted into the PAXgene RNA spin column (red) placed in a 2 ml processing tube, and centrifuged for 1 minute at 8,000–20,000 x g. The spin column was placed in a new 2 ml processing tube (PT), and the old processing tube containing flow-through discarded. Subsequently, the remaining sample was pipetted into the PAXgene RNA spin column, and centrifuged for 1 minute at 8,000–20,000 x g. The spin column was placed in a new 2 ml processing tube and the old processing tube containing flow-through again discarded. Subsequently, 350 µl wash buffer 1 (BR3) was pipetted into the PAXgene RNA spin column. After centrifugation for 1 minute at 8000–20,000 x g the spin column was placed in a new 2 ml processing tube and the old processing tube containing the flow-through again discarded. 10 µl DNase I stock solution is added to 70 µl DNA digestion buffer in a 1.5 ml microcentrifuge tube and mixed by gently flicking the

tube followed by a brief centrifugation to collect residual liquid from the sides of the tube. The DNase I incubation mix (80 µl) was pipetted directly onto the PAXgene RNA spin column membrane, and placed on the benchtop (20–30 °C) for 15 minutes. After Incubation, 350 µl wash buffer 1 (BR3) is pipetted into the PAXgene RNA spin column and centrifuged for 1 minute at 8000–20,000 x g. The spin column was placed in a new 2 ml processing tube and the old processing tube containing flow-through again discarded. 500 µl wash buffer 2 (BR4) was pipetted into the PAXgene RNA spin column and centrifuged for 1 minute at 8,000–20,000 x g. The spin column was placed in a new 2 ml processing tube and the old processing tube containing flow-through again discarded. Another 500 µl wash buffer 2 (BR4) was added to the PAXgene RNA spin column and centrifuged for 3 minutes at 8,000–20,000 x g. The spin column is placed in a new 2 ml processing tube and the old processing tube containing flow-through again discarded. The column is centrifuged for 1 minute at 8,000–20,000 x g. The processing tube is discarded and the column was placed on a 1.5 ml microcentrifuge tube. 40 µl elution buffer (BR5) directly pipetted onto the PAXgene RNA spin column membrane and subsequntly centrifuged for 1 minute at 8,000–20,000 x g to elute the RNA. The elution step is repeated using again 40 µl elution buffer (BR5) and the same 1.5 ml microcentrifuge tube. The RNA is denatured for 5 minutes at 65 °C in the shaker–incubator (see above) without shaking.

Results of RNA-Extraction

The quality control is performed on the Agilent Bioanalyzer.

From the RNA-extraction samples, 14 samples (2.92 %) showed RNA integrity numbers (RIN) lower than 3.5, which rendered these samples unfit for microarray hybridization.

Selection of Samples for Microarray Hybridization

Based on the results of RNA extraction and clinical information about UICC-stages and complete colonoscopy in case of non-CRC cases, 117 samples were authorized for microarray hybridization. The vast majority of drop-outs were incomplete colonoscopies.

Labeling and Chip Hybridization

Total RNA (100 ng) was amplified with the Ovation® Whole Blood system (Nugen, San Carlos, USA):

First Strand cDNA Synthesis

1. First Strand reagents A2 & A3 (stored at –20°C) and A1 (stored at –80°C) were obtained.
2. First Strand Enzyme Mix (blue: A3) was placed on ice.
3. First Strand Primer Mix (blue: A1) and First Strand Buffer Mix (blue: A2) were thawed at room temperature. Once thawed, reagents were placed on ice.
4. The contents of A1 were mixed by vortexing for two seconds, and then spun in a microcentrifuge for two seconds.
5. A 96 well PCR plate or 8-strip tubes were placed on ice.
6. The concentration of all RNA samples was adjusted so that the total amount in 5 µl was 50 ng. 5 µl of target RNA was aliquoted into each well of the 96 well plate (or 8-strip tubes) and kept on ice.
7. A1 was aliquoted into 8-strip tubes and kept on ice. See table below for volumes added to each tube of the 8-strip reservoir for the appropriate number of reactions.

Reagents	8 reactions = 1 strip	16 reactions = 2 strips	24 reactions = 3 strips	48 reactions = 6 strips	96 reactions = 12 strips
Volume of A1 primer to add per tube of mini 8-strip reservoir (in mL)	3	6	9	16	32

8. 2 µl of A1 was added per sample into the first column of the RNA reaction plate (or 8-strip tubes) using an 8-channel pipette. The reaction was mixed by pipetting up and down 3-4 times. The reaction mix was completely transferred from the pipette tips.
9. New tips were obtained, and the procedure was repeated for each column of the PCR plate (or 8-strip tubes) containing samples.
10. The PCR seal was placed on the plate in the correct orientation, then the plate (or 8-strip tubes) was spun briefly.
11. Plate was placed in a pre-warmed thermal cycler programmed to run:
12. The plate was incubated for five minutes at 65 °C.
13. While samples were incubating, the contents of A2 were mixed by vortexing for two seconds and then spun for two seconds and finally, placed on ice.
14. Shortly before the Primer Annealing step was complete, a Mastermix was prepared by combining A2 and A3 in an appropriately sized tube, using the volumes listed in the table below for various numbers of reactions. **Note:** The Mastermix was mixed by pipetting up and down 3-4 times to ensure the enzyme was mixed well.

Mastermix Calculations for First Strand Mastermix

(in ml)

Reagents	1 reaction	8 reactions = 1 strip	16 reactions = 2 strips	24 reactions = 3 strips	48 reactions = 6 strips	96 reactions = 12 strips
A2 (buffer)	12	120	240	348	665	1311
A3 (enzyme)	1	10	20	29	55	109
Total Volume of Mastermix	13	130	260	377	720	1420
Volume of Mastermix to add per tube of mini 8-strip reservoir		15.5	30	45	88	175

15. When the primer annealing was completed, the plate (or 8-strip tubes) was removed from the thermal cycler, then spun briefly and placed on ice.
16. #The thermal cycler was pre-warmed for the next incubation (see program 2 below).
17. 13 µl of the First Strand Mastermix was added per sample to the first column of the reaction plate, using an 8-channel pipette, mixed by pipetting up and down 3-4 times. The reaction mix was completely transferred from the pipette tips.
18. New tips were obtained, and the procedure was repeated for each column.
19. The PCR seal was placed on the plate in the correct orientation, then the plate (or 8-strip tubes) was spun briefly.
20. The plate (or 8-strip tubes) was placed in a pre-warmed thermal cycler programmed to run:
21. The plate (or 8-strip tubes) was incubated at 48 °C for 60 minutes.
22. The plate (or 8-strip tubes) was heated at 70 °C for 15 minutes.
23. Then the plate (or 8-strip tubes) was cooled to 4 °C.
24. While Program 2 was running, the Second Strand Buffer (yellow: B1) and Second Strand Enzyme Mix (yellow: B2) stored at –20°C was obtained. B2 was placed on ice, and B1 thaw was thawed at room temperature. (in preparation for Second Strand cDNA Synthesis, below)
25. Once the thermal cycler temperature had reached 4°C in Program 2, the plate (or 8-strip tubes) was removed from the thermal cycler, then spun briefly and placed on ice.
26. Second strand cDNA synthesis was immediately started.

Second Strand cDNA Synthesis

1. The contents of B2 were mixed by flicking the tube 6-8 times. Then the contents were spun in a microcentrifuge for two seconds, and placed on ice.

2. The contents of B1 were mixed by vortexing for two seconds. Then the contents were spun for two seconds in a microcentrifuge and placed on ice.
3. A Mastermix was made by combining B1 and B2 in an appropriately sized tube using the volumes listed in table below for the various numbers of reactions. Note: The Mastermix was mixed by pipetting up and down several times to ensure the enzyme was mixed well.

Mastermix Calculations for Second Strand Mastermix (in ml)

Reagents	1 reaction	8 reactions = 1 strip	16 reactions = 2 strips	24 reactions = 3 strips	48 reactions = 6 strips	96 reactions = 12 strips
B1 (buffer)	18	180	360	522	972	1944
B2 (enzyme)	2	20	40	58	108	216
Total Volume of Mastermix	20	200	400	580	1080	2160
Volume of Mastermix to add per tube of mini 8-strip reservoir		23	46	68	132	260

4. A strip of 8 tubes was filled with Second Strand Mastermix and kept on ice. See table above for volumes added to each tube of the 8-strip reservoir.
5. 20 µl of Second Strand master was added per sample to the first column of reaction plate (or 8-strip tubes), using an 8-channel pipette.
6. The reagents in that column were mixed by pipetting up and down 3-4 times, using the same pipette, volume setting, and tips.
7. New tips were obtained, and steps 5 & 6 were repeated for each column was spun briefly.
8. The PCR seal was placed on the plate, then the plate (or 8-strip tubes) briefly.
9. The plate was placed in a pre-warmed thermal cycler programmed to run:
10. The plate was incubated at 37°C for 30 minutes.
11. The plate was heated at 75°C for 15 minutes.
12. The plate was cooled to 4°C.
13. The SPIA Buffer Mix (red: C2), SPIA™ Enzyme Mix (red: C3), and water (green: D1) were obtained from the components stored at –20°C and the SPIA™ Primer Mix (red: C1) stored at –80°C. The enzyme mix was kept on ice. The remaining components were thawed at room temperature and then placed on ice. (in preparation for SPIA Amplification, below)
14. Once the temperature had reached 4°C, the plate was removed from the thermal cycler and placed on ice.
15. The SPIA™ amplification was immediately started.

SPIA Amplification

1. The contents of C3 were mixed by inverting gently 5 times. The enzyme was well mixed without introducing bubbles, and placed on ice.
2. The contents of C1 were mixed by vortexing for two seconds, and then spun in a microcentrifuge for two seconds, and finally, placed on ice.
3. The contents of C2 were mixed by inverting gently 5 times and placed on ice.
4. ASPIA Mastermix was made by combining C2, C1, water, and C3 in an appropriately sized tube and kept on ice using the volumes listed in table below for the various numbers of reactions. Note: The Mastermix was mixed by pipetting up and down several times to ensure the enzyme is mixed well. The Mastermix was used immediately.

Mastermix Calculations for SPIA Mastermix (in ml)

Reagents	1 reaction	8 reactions = 1 strip	16 reactions = 2 strips	24 reactions = 3 strips	48 reactions = 6 strips	96 reactions = 12 strips
C1 (primer)	2	18	36	54	106	201
C2 (buffer)	72	648	1296	1944	3816	7237
D1(water)	4	36	72	108	212	402
C3 (enzyme)	40	360	720	1080	2120	4020
Total Volume of Mastermix	118	1062	2124	3186	6254	11860
Volume of Mastermix to add per tube of large 8-strip reservoir		130	260	390	770	1475

5. A strip of 8 tubes was filled with SPIA Mastermix and kept on ice. See table above for volumes added to each tube of the 8-strip reservoir.
6. 118 µl of the SPIA™ Mastermix was added per sample to the first column of plate (or 8-strip tubes), using an 8-channel pipette.
7. The reagents in the column were mixed by pipetting up and down 3-4 times, using the same pipette, volume setting, and tips.
8. New tips were obtained, and the procedure repeated for each column.
9. At this stage, the SPIA reaction was divided into two plate wells (or 8-strip tubes) for incubation, and during this step the potential for sample cross contamination was high; therefore, care was taken in pipetting the correct columns.
10. 79 µl of the SPIA™ reaction mixture was transferred from each column of the reaction plate to the new plate wells (or 8-strip tubes), using an 8-channel pipette.

    Note: This was done to accommodate the 100ul max volume recommendation for heating in the thermal cyclers.
11. Using fresh tips, this was repeated for each column.

12. A PCR seal was placed on the plate (or close caps to 8-strip tubes), then the plates (or 8-strip tubes) were spun for 2 seconds.
13. The plate was placed in a pre-warmed thermal cycler according to Program SPIA™ Amplification 1:

Program SPIA™ Amplification 1	48°C for 30 minutes, then 4°C forever

14. The plate was incubated at 48°C for 30 minutes.
15. Once the temperature had reached 4°C, the plate was removed from the thermal cycler and placed on ice.
16. 3 ul Whole Blood Reagent (C4) was added to each well. Each well was mixed by pipetting up and down 3-4 times with a large volume (>40 ul).
17. The plate (tubes) was resealed and returned to pre-warmed thermal cycler running Program SPIA™ Amplification 2:

Program SPIA™ Amplification 2	48°C for 30 minutes, then 95°C for 5 minutes, then 4°C forever

18. The plate (tubes) was incubated at 48°C for 30 minutes.
19. The samples were heated to 95°C for 5 minutes.
20. The samples were cooled to 4°C.
21. The plate (or 8-strip tubes) was removed from the thermal cycler and placed on ice for immediate use in purification or at -20°C (to store for purification at a later time).

After SPIA™ Amplification, the cDNA was purified with a QIAquick PCR purification spin column (QIAGEN, Hilden) using the Qiacube robot. One microgram of purified cDNA was then subjected to Cy3 labeling with the NimbleGen One-Color DNA Labeling Kits (Roche NimbleGen, Inc.; Madison, WI, USA.) following the manufacturers instructions:

1. The following solution was prepared:

   Random Primer Buffer (vial 2) 1100 µl

   β-Mercaptoethanol 2 µl

   (Fresh buffer was prepared each time the primers were resuspended.)

2. Cy3 Random Nonamers (vial 3) were briefly centrifuged because some of the product could have dislodged during shipping. The primer was diluted in 1,050 µl of Random Primer Buffer with β-Mercaptoethanol. 40 µl individual reaction volumes were aliquoted in 0.2 ml thin-walled PCR tubes and store at -15 to -25°C, protected from light.

3. The following components were assembled in separate 0.2 ml thin-walled PCR tubes:

   cDNA 1 µg
   Diluted Cy3 Random Nonamers from step 2 40 µl
   PCR Grade Water (vial 1) fill to volume 80 µl

4. The samples were heat-denature in a thermocycler at +98°C for 10 minutes. Quick-chilled in an ice-water bath for 2 minutes.

5. The following dNTP/Klenow master mix was prepared for each sample prepared in step 4. All reagents and dNTP/Klenow master mix were kept on ice. After the addition of Klenow, the samples were no longer vortexed.

   dNTP Mix (10 mM each dNTP) (vial 5) 10 µl
   PCR Grade Water (vial 1) 8µl
   Klenow Fragment (3'->5' exo-) 50U/µl (vial 4) 2µl

6. 20 µl of the dNTP/Klenow master mix prepared in step 5 was added to each of the denatured samples prepared in step 4 and kept on ice. Total volume was now 100 µl

7. The samples were mixed well by pipetting up and down 10 times. No vortexing took place after the addition of Klenow.

8. The samples were spun quickly to collect contents in bottom of the tube.

9. The samples were incubated for 2 hours at +37°C in a thermocycler with a heated lid, and protected from light.

10. The reaction was stopped by adding 21,5µl of the Stop Solution (vial 6). Total volume was now 121,5 µl

11. The samples were vortexed briefly, spun, and the entire contents were transferred to a 1.5 ml tube containing 110 µl isopropanol. Total volume was now 231,5 µl

12. The samples were vortexed well and incubated for 10 minutes at +15 to +25°C, protected from light.

13. The samples were centrifuged at 12,000 x g for 10 minutes. The supernatant was removed with a pipette. The pellet was pink.

14. The pellet was rinsed with 500 µl 80% ice-cold ethanol. The pellet was dislodged from tube wall by pipetting a few times.

15. The pellet was centrifuged at 12,000 x g for 2 minutes. The supernatant was removed with a pipette.
16. The contents were dried in a DNA vacuum concentrator on low heat until dry (approximately 5 minutes), protected from light.
17. The tubes were spun briefly prior to opening. Each pellet was rehydrated in 25 µl PCR Grade Water (vial 1).
18. The pellet was vortexed for 30 seconds, quick-spun, and the contents were collected in the bottom of the tube. Either vortexing was continued or the samples were left at +15 to +25°C, protected from light, for approximately 5 minutes or until the pellet was completely rehydrated, then vortexed again and quick-spun.
19. Each sample was quantitated using the following formula:

    Concentration (µg/ml) = A260 x 50 x Dilution Factor. Based on the concentration, calculate the volume of Cy3-labeled cDNA sample required for each hybridization to obtain 4µg.

Four micrograms of the labeled cDNA was then hybridized onto NimbleGen Human Gene Expression 12x135K Arrays (Roche NimbleGen, Inc.; Madison, WI, USA) according to the manufacturer's instructions:

Sample preparation

1. The Hybridization System was set to +42°C. With the cover closed, at least 3 hours were allowed for the temperature to stabilize.
2. The dried sample pellet was resuspended in 5µl water and 3.3 µl of the appropriate Sample Tracking Control. Each sample to be hybridized an array was resuspended in a unique STC. STCs used for each sample was recorded. If Sample Tracking Controls were not used, then the dried sample pellet was resuspended in 8.3 µl water.
3. The samples was vortexed well and spun to collect contents at bottom of the tube.
4. Using components from a NimbleGen Hybridization Kit, the hybridization solution master mix was prepared.

2X Hybridization Buffer (vial 1)	88.5 µl
Hybridization Component A (vial 2)	35.4 µl

Alignment Oligo (vial 3) 3.6 µl

The amount was sufficient to hybridize all arrays on one slide. To hybridize multiple slides, the amounts were adjusted accordingly.

5. 8,7 µl of hybridization solution was added to the resuspended pellet (3,3 µl) from step 2.
6. The sample was vortexed well (approximately 15 seconds) and spun to collect contents in bottom of the tube. The sample was incubated at +95°C for 5 minutes, protected from light.
7. The tubes were placed at +42°C (in the Hybridization System sample block or heat block) for at least 5 minutes or until ready for sample loading. The tubes were vortexed and spun prior to loading.

The Preparation of Mixers

1. The HX12 mixer was located and removed from its package. A compressed gas nozzle was used to gently blow compressed nitrogen or argon gas across the mixer and moved slowly to remove any dust or debris. Canned compressed air for this purpose was not used. Samples were loaded within 30 minutes of opening the vacuum-packaged mixer to prevent the formation of bubbles during loading and/or hybridization.
2. The Precision Mixer Alignment Tool (PMAT) was positioned with its hinge on the left. The PMAT was opened.
3. The mixer was snapped onto the two alignment pins on the lid of the PMAT with the tab end of the mixer toward the inside hinge and the mixer's adhesive gasket facing outward.
4. While pushing back the plastic spring with a thumb, the slide was placed in the base of the PMAT so that the barcode was on the right and the corner of the slide sat against the plastic spring. The NimbleGen logo and barcode number was readable. The thumb was removed when the spring had engaged the corner of the slide and the entire slide was registered to the edge of the PMAT to the rightmost. It was ensured that the slide was lying flat against the PMAT. Compressed nitrogen or argon gas was gently blown across the mixer and slide to remove dust.
5. Using forceps, the backing was removed from the adhesive gasket of the mixer, and the lid of the PMAT was closed so that the gasket makes contact with the slide.

6. The lid was lifted by grasping the long edges of the PMAT while simultaneously applying pressure with a finger through the window in the lid of the PMAT to free the mixer-slide assembly from the alignment pins.
7. The mixer-slide assembly from the PMAT was removed.
8. The mixer-slide assembly was placed on the back of a +42°C heating block for 5 minutes to facilitate complete adhesion of the mixer to the slide.
9. The Mixer Brayer rubbed over the mixer with moderate pressure to adhere the adhesive gasket and to remove any bubbles. A corner of the Mixer Brayer was used first to rub the borders between the arrays and then was used to rub around the outside of the arrays. The adhesive gasket became clear when they were fully adhered to both surfaces.
10. The mixer-slide assembly was placed in the slide bay of the Hybridization System.

Loading & Hybridizing Samples

1. The manufacturers loading diagram was referred to when loading samples to localize the appropriate fill and vent ports.
2. The following was kept in mind when the samples were loaded:

   • When pipetting the sample before loading, residual volume was left in the sample tube to avoid introducing bubbles. The volumes needed (6 µl) allowed for a residual volume.

   • After aspirating the designated sample volume, the pipette tip was inspected for air bubbles. If bubbles existed, the pipette was dispensed and reloaded.

   The following was kept in mind when the samples were loaded:

   • The pipette tip was kept perpendicular to the slide to avoid possible leakage at the fill port.

   • Gentle pressure of the tip was applied into the port to ensure a tight seal while loading the sample.
3. The sample was loaded into a fill port. Due to the close proximity of the fill and vent ports, care was taken not to overfill the arrays. The sample was loaded until it entered the vent port channel. The sample was not allowed to come to the surface of the HX12 mixer. Any overflow from the fill and vent ports was dried with a cotton swab after loading the

array. Small bubbles occasionally formed in the corners of the mixer-slide assembly during loading. These bubbles dissipated upon mixing and did not compromise the data.

4. One mixer port seal was used to cover both the fill and vent ports on the mixers, one chamber at a time was filled and sealed. The mixer port seal was pressed, using uniform pressure across the seal to adhere.
5. Forceps were used to press the mixer port seal around the fill and vent ports to ensure it had adhered in those areas
6. The bay clamp was closed.
7. The Mixing Panel was turned on in the Hybridization System, the mix mode was set to B, and the mix button was pressed to start mixing. That the Hybridization System recognizes the slide in each occupied bay (its indicator light becomes green) was confirmed.
8. Approximately 10 minutes after the Hybridization System was started:
   - The setting of the mix mode to B was verified.
   - The display of a green light for all occupied stations was verified.
9. The sample was hybridized at +42°C to the array(s) for 16 - 20 hours.

Washing of Hybridized Arrays

1. The components of the NimbleGen Wash Buffer Kit and NimbleGen Array Processing Accessories were located. Prior to the first use of the Wash Buffer Kit, the DTT was reconstituted. In a fume hood, 1 M DTT solutions were prepared by adding 1.2 ml of water (vial 5) to each tube of dry DTT (vial 4). After reconstitution, the 1 M DTT solutions were stored at -15 to -25°C.
2. Before removing the mixer-slide assemblies from the Hybridization System, Washes I, II, and III were prepared by adding 10X Wash Buffer I, II, or III (vial 1, 2, or 3) in each case to 243 ml water and 27 ml DTT solution from step1. Two containers of Wash I were prepared.
3. To facilitate the removal of the mixer, a shallow dish containing Wash I to +42°C was heated. The temperature of Wash I was measured at every use. The remaining three wash solutions were kept at +15 to +25°C.

4. The Mixer Disassembly Tool was inserted into the shallow dish containing warm Wash I. When multiple slides were washed, a slide rack was inserted into the wash tank containing Wash I at +15 to +25°C.

5. The mixer-slide assembly was removed from the Hybridization System and was loaded into the Mixer Disassembly Tool, which was immersed in the shallow dish containing warm Wash I.

6. The mixer-slide assembly was not allowed to cool before removing the mixer. Power was kept on to the Hybridization System's heat block and mixer system during mixer-slide disassembly, and each mixer-slide assembly was transferred one at a time to Wash I for immediate removal of the mixer.

7. With the mixer-slide assembly submerged, the mixer was carefully peeled off the slide. The Mixer Disassembly Tool was held flat while removing the mixer to avoid any horizontal movement or scraping with the mixer across the slide. Care was taken not to touch the array surface of the slide.

8. As the mixer is extremely flexible, care was taken to peel the mixer off slowly to avoid breaking the slide.

   The mixer was discarded, and the slide removed from the Mixer Disassembly Tool.

9. The slide was gently agitated for 10 - 15 seconds in the shallow dish containing warm Wash I to quickly remove the hybridization buffer. To achieve good array uniformity, the hybridization buffer was quickly and evenly washed off the slide surface as soon as the mixer was removed.

   When washing multiple slides, the slide was transferred with the barcode at the top into a slide rack in the wash tank that contained Wash I. When washing one slide, the slide was transferred into a slide container that contained Wash I. It was agitated vigorously for 10 - 15 seconds.

   In order to ensure high quality data, the microarray area of the slide was kept wet at all times during all wash steps.

10. When a NimbleGen Microarray Dryer or other microarray dryer that dries multiple slides at a time was used, steps 4 - 9 were repeated until all the mixer was removed from all slides to wash. Each slide was loaded into the slide rack with the array facing the same direction.

11. The slides were washed for an additional 2 minutes in Wash I with vigorous, constant agitation. When multiple slides were washed, the rack was moved up and down with enough agitation to make foam appear. When one slide was washed, the slide container was shaken at least 20 times every 10 seconds.

    Several times during the wash, the wash tank was rocked so the wash solution covered and cleaned the top of the slides.

12. The rack was quickly blotted several times using paper towels to minimize buffer carryover. The slide(s) was transferred to Wash II and washed for 1 minute with vigorous, constant agitation. When multiple slides were washed, the wash tank was rocked so the wash solution covered and cleaned the tops of the slide(s).

    The slides were not allowed to dry between the wash steps.

13. The slide(s) was transferred to Wash III and washed for 15 seconds with vigorous, constant agitation. When multiple slides were washed, the wash tank was rocked so the wash solution covered and cleaned the tops of the slide(s).

14. The slide(s) was removed from Wash III. The slide(s) was spin-dried in a NimbleGen Microarray Dryer or other microarray dryer per the manufacturer's recommendation. For a NimbleGen Microarray Dryer, the drying time was 2 minutes (120 seconds).

15. The slide(s) was removed from the NimbleGen Microarray Dryer or other microarray dryer. The edges were blot-dried with lint-free paper to remove any residual moisture.

    The steps for scanning the array(s) were carried out immediately thereafter.

16. When the slides could not be scanned immediately after washing, they were kept in their original slide case in a dark desiccator until they were ready to be scanned.

The array was then scanned with the NimbleGen MS 200 Microarray Scanner and Hybridization signals were collected with the NimbleGen the MS 200 Data Collection Software, using the default settings. Text files were exported and imported into R (Bioconductor package) for further analysis.

Microarray Quality Control

Chip images were visually inspected for surface errors. All chips passed this test. The microarray data were quality controlled using spearman correlation with a cutoff set to 0.85. One chip was excluded from further analysis.

Discovery Procedure

In brief, from all raw measurement data from microarray hybridizations scheduled for the discovery (i.e., from samples selected for discovery, and measurements that passed quality control) the median value of the three probesets existing for each transcript was used and the logarithmic values were quantile normalized. The statistical software systems R (version 2.9.0) and Bioconductor (version 1.6) (for this software, see http://cran.r-project.org/, http://www.r-project.org/ and http://www.bioconductor.org, also Ross Ihaka and Robert Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3): 299-314, 1996; Gentleman et al 2005: Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Springer, New York, NY) were used for anlysis.

The resulting normalized array data were then partitioned into groups of samples in a double nested bootstrap approach (Efron (1979) Bootstrap Methods--Another Look at the Jackknifing, Ann. Statist. 7, 1-6). In the outer loop of this bootstrap, the samples were partitioned into an outer test set and an outer training set. In the inner bootstrap loop, this outer training set was partitioned again into an inner training set and an inner test set.

On the inner training set, probeset relevance was estimated through a decision-tree-analysis. The influence of each feature was determined from its contribution to the classification error: In case the error of a probeset increases due to the permutation of the values of a probeset while the values of all other probesets remain unchanged, this probeset is weighted more strongly.

The relevance evaluations of the individual inner loops were combined for each external loop and the chosen probesets were used to train a support vector machine (Bernhard Schölkopf, Alex Smola: *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA, 2002, also see citations above). This classifier was

trained on the outer training set (which was already used for its feature selection) and applied to the outer test set of samples. For this classification, the same methods and parameters were use as for the intended application on external data (for example as part of validation): the details are described below as application for the RNAs of the invention. The whole external loop of the bootstrap procedure is a simulation of later classification of unknown data. Its average performance over all external loops gives a prospective estimate of the performance of the classification procedure (mainly based on the chosen biomarker set). The common mistake of overfitting, i.e., the overly optimistic evaluation of a classification method on its discovery set, is thereby avoided.

Finally, the results of all inner loops and of all external loops were combined to form a common biomarker set with prospective estimates of performance.

Application of the Invention

A number of RNAs that result from the discovery above can be used with state of the art classification methods as described in the cited literature and as known by a person of skill in the art. As any classification, it will require a representative training set, which the inventors obtained through a clinical study fulfilling the requirements described above. Part of the training set is the necessary clinical information (carcinoma patient or healthy control, as defined by the clinical study). Similar to the clinical requirements, this description of the algorithmic part of the application of the invention also presupposes the described lab process and quality controls were applied.

The training set of microarray raw data was condensed by the same method as the discovery set in the section "Discovery Procedure" above: the median value of the three probesets existing for each transcript was used and the logarithmic values were quantile normalized. A standard implementation of support vector machines was used, such as the one contained in Bioconductor's (see references above) package for machine learning "e1071" with parameter "kernel" set to "linear" (preferential version for this package is 1.5-19).

It is important not to present the classification information on the training data as numeric data, but as categorical data. This can be ensured by passing the corresponding arguments as

an R "factor" (for example as in "svm(..., as.factor(clinicalData),...)". Otherwise this svm algorithm will use the wrong type of classification.

To apply this svm, using the same software package, to any kind of new microarray data, a condensation step is necessary just as for the discovery data. It is possible to simply reapply the condensation method above to each individual new sample in combination with the whole discovery set. This first method approximates the preferential method well, which is to explicitly compute a condensation model or parameter based on the discovery data and apply it to the test data. This can be done with the software package on.farms, preferentially version 1.4.1, available from the author (see references above). The application of the described svm to the thus condensed test data produces the desired decision value.

The embodiments illustrated and discussed in this specification are intended only to teach those skilled in the art the one way to make and use the invention. Modifications and variation of the above-described embodiments of the invention are possible without departing from the invention, as appreciated by those skilled in the art in light of the above teachings. It is therefore understood that, within the scope of the claims and their equivalents, the invention may be practiced otherwise than as specifically described.

Table 1 SEQ ID NO	Nimblegene Probe ID / Primary Sequence ID	Gene Symbol
1	NM_203281	BMX
2	NM_001013725	LOC441268
3	NM_032415	CARD11
4	NM_001236	CBR3
5	NM_144646	IGJ
6	NM_014230	SRP68
7	AK026965	N/A
8	NM_021244	RRAGD
9	BC045532	LSM8
10	NM_004972	JAK2
11	NM_015336	ZDHHC17
12	NM_001017523	BTBD11
13	XM_933163	LOC644019
14	XM_291663	LIPL3
15	Z22970	CD163
16	XM_933170	LOC644019
17	BC051905	ZNF281
18	NM_001039796	FLJ35776
19	NM_001617	ADD2
20	NM_006135	CAPZA1
21	AF329495	FCRLM1
22	NM_012161	FBXL5
23	NM_001012707	C21orf114
24	NM_145800	38961
25	BC007537	CSRP2BP
26	NM_144570	C16orf34
27	BC064557	LOC653866
28	BC020653	ARG1
29	XM_930401	LOC400927
30	AK127341	ARMC8
31	BC050434	FLJ10154
32	BC063635	PDSS1
33	NM_018327	C20orf38
34	BC069006	CBWD3
35	AY382899	N/A
36	NM_152581	MOSPD2
37	NM_024610	HSPBAP1
38	XM_043653	BEXL1
39	NM_133279	FCAR
40	BC000819	RP11-82K18.3
41	NM_013314	BLNK
42	NM_145647	WDR67
43	NM_014767	SPOCK2
44	AY116204	NALP12
45	NM_052903	TUBGCP5
46	NM_001025108	AFF3
47	AF116273	BAG1
48	NM_177980	CDH26
49	NM_030814	C9orf45
50	BC002362	LDHB
51	BC035854	DSCR1L2
52	NM_022746	MOSC1
53	NM_001012713	C10orf31
54	NM_016250	NDRG2
55	BC047681	S100A9
56	NM_002197	ACO1
57	NM_020690	MASK-BP3
58	NM_080747	K6IRS2
59	NM_031988	MAP2K6
60	BC093977	IL18R1
61	XM_932570	LOC644993
62	BC002448	ABLIM1
63	BC044574	ECHDC2
64	BX648869	PARP9
65	BC045629	DOCK8
66	BC013876	OPTN
67	XM_933190	LOC644019
68	NM_012347	FBXO9
69	BC012009	MAP2K6
70	XM_376576	LOC401308
71	BC058031	HP
72	BC035582	TRIM22
73	NM_005232	EPHA1
74	NM_024829	FLJ22662
75	BC039237	NEFL
76	NM_014207	CD5
77	NM_018284	GBP3
78	AK096196	N/A
79	NM_001012761	RGMB
80	BC092516	NCOA3
81	NM_001716	BLR1
82	XM_926338	LOC642956
83	CR456539	PES1
84	BC036511	MEIS1
85	NM_001034843	N/A

86	NM_080913	ASGR2
87	NM_181804	PKIG
88	XM_927709	LOC653491
89	AK092324	ADAMTSL3
90	NM_201453	CBWD3
91	NM_001540	HSPB1
92	XM_496277	GTSCR1
93	NM_024662	NAT10
94	BC070140	ARL6IP6
95	NM_173644	FLJ33860
96	XM_377476	MGC57346
97	NM_014184	CNIH4
98	BC070387	TRBV3-1
99	AX772832	N/A
100	AJ617684	BPHL
101	XM_165511	FABP5L3
102	NM_001009566	CLSTN1
103	NM_024838	THNSL1
104	NM_004032	DDO
105	NM_032738	FCRLM1
106	NM_001012991	LOC400506
107	NM_001007273	DUSP13
108	AK055262	N/A
109	NM_004392	DACH1
110	CR612818	NUDC
111	NM_153611	CYBASC3
112	BC037545	PARP1
113	XM_933184	LOC644019
114	NM_001034842	PTCHD3
115	NM_001039954	LOC645745
116	BC010926	HIST1H4H
117	NM_020414	DDX24
118	AK026323	ALPK1
119	XM_933364	LOC653458
120	NM_000322	RDS
121	BC014958	BRDG1
122	XM_926748	LOC643416
123	NM_020697	KCNS2
124	NM_002938	RNF4
125	BC069661	CD200R1
126	AB071605	HUWE1
127	NM_176870	MT1M
128	NM_000045	ARG1
129	NM_001013702	LOC440258

130	AK172753	LOC129285
131	CR749296	CTRC
132	NM_014462	LSM1
133	AY358410	N/A
134	NM_152914	MGC33894
135	BC047310	LOC346653
136	NM_001005409	SF3A1
137	AK093006	N/A
138	BC005831	TCL1A
139	NM_172249	CSF2RA
140	XM_370729	LOC387934
141	XM_930986	LOC642466
142	BC065009	ZNF343
143	AK127911	N/A
144	BC051725	N/A
145	NM_003202	TCF7
146	NM_002970	SAT
147	NM_198799	BCAS4
148	NM_016424	CROP
149	BC048769	TCF7
150	NM_018023	YEATS2
151	NM_133271	FCAR
152	NM_018206	VPS35
153	NM_144590	ANKRD22
154	NM_005091	PGLYRP1
155	BC017798	ANKRD49
156	NM_032295	SLC37A3
157	XM_927900	LOC653527
158	NM_006734	HIVEP2
159	BC009016	C1QC
160	NM_014641	MDC1
161	XM_932082	LOC339287
162	NM_019846	CCL28
163	NM_001012971	C20orf106
164	NM_004463	FGD1
165	NM_018668	VPS33B
166	XM_929621	LOC646675
167	NM_203414	C17orf81
168	NM_024525	TTC13
169	AK126355	ALDH18A1
170	BC052611	PXN
171	NM_173544	BCNP1
172	NM_000461	THRB
173	NM_014164	FXYD5

174	AF388365	MED12L
175	BX647724	RNASEN
176	NM_016951	CKLF
177	XM_928303	LOC645262
178	BC037960	CCR6
179	XM_934852	LOC129293
180	NM_014941	MORC2
181	NM_006159	NELL2
182	NM_031407	HUWE1
183	NM_024103	SLC25A23
184	BC017583	SCRG1
185	AY358341	VSIG4
186	BC071668	HIST1H2AM
187	NM_001237	CCNA2
188	NM_147200	TRAF3IP2
189	NM_022893	BCL11A
190	NM_000877	IL1R1
191	NM_172369	C1QC
192	XM_929287	LOC646358
193	BC017745	NUFIP1
194	NM_006029	PNMA1
195	BC021110	WSB1
196	NM_018660	ZNF395
197	AK098423	PAF1
198	BC007563	RPL13
199	BC020757	MT1G
200	AK127365	PDE4B
201	BC037897	LOC196394
202	NM_000633	BCL2
203	NM_021809	TGIF2
204	NM_001442	FABP4
205	NM_018385	LSG1
206	NM_017929	PEX26
207	NM_004044	ATIC
208	AK054879	N/A
209	BC004548	PRO1853
210	BC060324	HIST2H2AC
211	AK092231	KLHL25
212	NM_004753	DHRS3
213	NM_001012981	ZNF694
214	AK057340	SLC36A1
215	AY044431	DDX27
216	NM_052950	WDFY2
217	AK130657	CD19

218	BC006521	FCRLM1
219	AK026669	DYNLT1
220	AK025303	KIF13A
221	NM_012411	PTPN22
222	AY847220	PXK
223	NM_021966	TCL1A
224	NM_018491	CBWD1
225	NM_181641	CKLF
226	XM_926738	LOC643403
227	NM_015352	POFUT1
228	AF528079	MGC17330
229	NM_001014446	OCIAD2
230	BC022337	CDC42EP2
231	NM_020932	MAGEE1
232	AF010447	MR1
233	NM_015626	WSB1
234	AK127026	LOC340318
235	NM_018235	CNDP2
236	NM_020853	KIAA1467
237	AK125030	ELAC2
238	NM_015066	TRIM35
239	AK091509	RNF175
240	AY956764	HSP90AB3P
241	NM_015225	KIAA0367
242	NM_033419	PERLD1
243	AL834304	SFRS15
244	BC020238	SRP68
245	NM_022138	SMOC2
246	BC092450	NGFRAP1L1
247	NM_003121	SPIB
248	NM_004393	DAG1
249	NM_004281	BAG3
250	NM_022368	PJA1
251	BC027036	STAT5A
252	NM_144673	CMTM2
253	NM_004799	ZFYVE9
254	BC009573	CBWD1
255	NM_002119	HLA-DOA
256	NM_001012706	LOC283551
257	NM_003707	RUVBL1
258	NM_000313	PROS1
259	NM_199135	RP11-561O23.3
260	NM_001001433	STX16
261	NM_013378	VPREB3

262	NM_000268	NF2
263	AM180330	BLNK
264	NM_001010867	C1orf69
265	NM_004226	STK17B
266	XM_497054	LOC653458
267	NM_001031716	OBFC2A
268	BC063430	CPXM
269	AK022948	FLJ12886
270	NM_014587	SOX8
271	NM_022351	EFCBP1
272	XM_498917	LOC440928
273	NM_133466	ZNF545
274	NM_000878	IL2RB
275	AY923176	MDM4
276	AK001930	UEV3
277	BC107723	OBFC2A
278	NM_000664	ACACA
279	NM_016113	TRPV2
280	AY358413	N/A
281	NM_173484	KLF17
282	AK022723	C13orf23
283	BC009569	ASB3
284	NM_015292	FAM62A
285	NM_153214	FLJ37440
286	NM_030640	DUSP16
287	NM_207388	FLJ31222
288	AK094105	MOSC1
289	XM_933343	LOC653458
290	NM_198521	C12orf42
291	NM_004183	VMD2
292	BC033896	AK5
293	NM_020156	C1GALT1
294	BC027920	LY9
295	XM_927534	LOC644389
296	NM_052837	SCAMP3
297	NM_032330	CAPNS2
298	NM_002883	RANGAP1
299	BC025708	MCTP2
300	NM_183416	KIF1B
301	NM_005653	TFCP2
302	NM_032318	HIATL2
303	NM_019088	PAF1
304	BC056147	FBXO38
305	XM_933148	LOC644019

306	BC010505	SCAMP3
307	NM_000491	C1QB
308	BC015831	DNASE1L3
309	NM_153047	FYN
310	NM_006573	TNFSF13B
311	NM_012216	MID2
312	NM_020164	ASPH
313	XM_933093	LOC644019
314	NM_015041	CLUAP1
315	AK124867	CBWD2
316	NM_016616	TXNDC3
317	NM_019008	RP5-1104E15.5
318	NM_181501	ITGA1
319	NM_024021	MS4A4A
320	BC006564	ST6GALNAC6
321	XM_929992	FLJ46120
322	NM_003152	STAT5A
323	NM_001039702	THEDC1
324	AY439213	KIAA0367
325	XM_372108	LOC389748
326	NM_138394	HNRPLL
327	XM_931904	LOC653210
328	BC100019	HMGB2
329	NM_015721	GEMIN4
330	NM_001031853	LOC387755
331	BC017701	C4orf18
332	NM_018127	ELAC2
333	NM_005035	POLRMT
334	NM_002167	ID3
335	BC106933	CCT4
336	BC067803	CBWD5
337	NM_153694	SYCP3
338	NM_014303	PES1
339	NM_001002021	PFKL
340	NM_022336	EDAR
341	BC067456	GPR52
342	NM_001037175	SUSD4
343	XM_925980	LOC442446
344	AK022576	CNOT10
345	AK092265	PRRT2
346	NM_030801	MAGED4
347	AK000110	USP48
348	BC088370	PRL
349	NM_000937	POLR2A

350	XM_930399	KIAA0565
351	NM_018324	THEDC1
352	BC020666	VPREB3
353	AK128707	N/A
354	AF447872	N/A
355	NM_080387	CLEC4D
356	NM_013235	RNASEN
357	AL832648	NDEL1
358	NM_015568	PPP1R16B
359	NM_177537	MAGED4
360	XM_933882	LOC646744
361	NM_022074	FAM111A
362	NM_000334	SCN4A
363	NM_001014368	HSP90AB2P
364	NM_004082	DCTN1
365	BC002807	MS4A1
366	NM_014740	DDX48
367	XM_926000	FAM7A3
368	AM180338	BLNK
369	NM_032228	MLSTD2
370	NM_004818	DDX23
371	BC068516	FLJ20581
372	AK022471	NGLY1
373	BC002701	ASCIZ
374	NM_147184	TP53I3
375	XM_929819	LOC646861
376	NM_173630	RTTN
377	NM_002064	GLRX
378	BC001401	ZAK
379	NM_018196	TMLHE
380	NM_173676	PNPLA1
381	NM_001164	APBB1
382	BC001139	AZI2
383	AK027414	NOD27
384	BC023570	YLPM1
385	BC028003	IL27RA
386	NM_001363	DKC1
387	BC103841	MT1M
388	NM_014706	SART3
389	XM_934960	NBPF1
390	U69668	TPR
391	NM_002965	S100A9
392	BC031797	SOX8
393	BC028391	SPPL2B

394	NM_020379	MAN1C1
395	XM_931779	LOC643754
396	BC014261	RHOH
397	NM_022006	FXYD7
398	BC025340	MGC39372
399	AY339061	CDC91L1
400	AK123693	CKAP4
401	XM_933210	LOC644019
402	NM_000901	NR3C2
403	NM_032132	HORMAD1
404	NM_003236	TGFA
405	NM_133274	FCAR
406	AX772826	N/A
407	U52682	IRF4
408	AK130452	N/A
409	BC029832	ZYG11B
410	BC025782	CD4
411	NM_199483	C20orf24
412	NM_020795	NLGN2
413	BC100963	IL15
414	BC027724	KIAA0241
415	XM_932385	LOC653840
416	NM_001008237	LOC130502
417	XM_932827	LOC653337
418	NM_001017980	LOC203547
419	NM_001008756	GCET2
420	BC003107	ID3
421	BC033094	IKBKAP
422	BC047030	UPP1
423	AF329491	FCRLM1
424	NM_002375	MAP4
425	NM_001018053	PFKFB2
426	AM180340	BLNK
427	NM_018011	FLJ10154
428	AF046024	UBE1C
429	NM_018014	BCL11A
430	XM_929509	LOC653761
431	BC069781	N/A
432	NM_002544	OMG
433	BC107704	C20orf106
434	NM_000273	GPR143
435	NM_006600	NUDC
436	XM_928492	LOC645460
437	NM_031892	SH3KBP1

438	NM_145038	MGC16372
439	NM_133269	FCAR
440	NM_030944	C15orf5
441	NM_014500	HTATSF1
442	BC054344	USP32
443	XM_928033	LOC644948
444	NM_178231	ALS2CR14
445	DQ008445	GADD45A
446	BC109069	CD180
447	NM_033051	TSCOT
448	XM_930969	FLJ31033
449	BC035554	CKAP5
450	NM_181805	PKIG
451	NM_000031	ALAD
452	BC026327	EPHA4
453	NM_001775	CD38
454	BC013437	MEF2A
455	BC073797	TRBC1
456	BC091490	ODF2L
457	NM_000386	BLMH
458	NM_173462	PAPLN
459	NM_021065	HIST1H2AD
460	NM_014872	ZBTB5
461	XM_928264	LOC645225
462	BC101707	WDR6
463	NM_002844	PTPRK
464	BC074728	SRR
465	BC018145	FLJ38663
466	XM_935143	C17orf60
467	BC071582	USP36
468	NM_005870	SAP18
469	BC029412	DYNLT1
470	CR606022	C2orf4
471	BC002812	CBR3
472	NM_152994	LOC129285
473	NM_004748	CCPG1
474	NM_006779	CDC42EP2
475	BC031056	TCF4
476	AK026162	KIAA0515
477	BC000616	SWAP70
478	NM_017784	OSBPL10
479	NM_003864	SAP30
480	BC000169	WBSCR22
481	NM_145239	PRRT2

482	NM_173799	VSIG9
483	XM_933156	LOC644019
484	NM_021724	NR1D1
485	XM_933196	LOC644019
486	NM_018128	TSR1
487	NM_052838	37135
488	NM_003530	HIST1H3D
489	NM_030798	WBSCR16
490	NM_018072	HEATR1
491	NM_004963	GUCY2C
492	AK128046	N/A
493	XM_370697	LOC387867
494	AK131437	SIRT7
495	AK021886	TANC2
496	NM_006257	PRKCQ
497	NM_001006109	C3orf37
498	NM_199485	C20orf24
499	AF087872	NDRG2
500	BC069668	ZNF415
501	BC017046	ANXA6
502	BC032454	SDCCAG8
503	AB051543	SYNE1
504	BC006447	NFX1
505	XM_933219	LOC644019
506	BC027922	N/A
507	BC032577	ZNF545
508	NM_018029	FLJ10213
509	NM_001515	GTF2H2
510	NM_022662	ANAPC1
511	AK021784	C15orf28
512	NM_003270	TSPAN6
513	XM_930891	FAM7A3
514	CR626109	SARS
515	NM_005762	TRIM28
516	BC112127	SMC1L1
517	NM_003253	TIAM1
518	AL136910	FIP1L1
519	NM_007066	PKIG
520	AF329494	FCRLM1
521	NM_015492	C15orf39
522	NM_000701	ATP1A1
523	NM_006993	NPM3
524	BX641086	SLC35B3
525	BC104978	TPARL

526	NM_144736	PRO1853
527	BC023558	SPOCK2
528	AF497245	DREV1
529	NM_004367	CCR6
530	NM_182827	FKBP9L
531	XM_933152	LOC644019
532	AK027305	TMEM25
533	NM_006549	CAMKK2
534	XM_933146	LOC645789
535	NM_005502	ABCA1
536	NM_005582	CD180
537	XM_930920	LOC642548
538	NM_031206	LAS1L
539	XM_929358	LOC644037
540	BC024016	FIP1L1
541	NM_182852	CCNB1IP1
542	NM_015147	CEP68
543	BC063384	IGHD
544	NM_005698	SCAMP3
545	NM_001013646	RP5-1153D9.3
546	BC036917	C6orf141
547	XM_926390	FAM7A1
548	NM_005214	CTLA4
549	NM_003534	HIST1H3G
550	BC068522	JUN
551	NM_012210	TRIM32
552	NM_207331	LOC153561
553	NM_182597	FLJ39575
554	AF212253	CBWD1
555	BC031618	C10orf30
556	NM_005407	SALL2
557	NM_001004441	DP58
558	NM_021114	SPINK2
559	XM_930928	LOC642548
560	XM_933238	LOC644019
561	NM_172000	TEDDM1
562	BC023998	MTSS1
563	NM_018475	TPARL
564	AB075864	KIAA1984
565	XM_933704	MEGF9
566	BC010931	MEF2B
567	NM_001008800	CCT3
568	NM_133280	FCAR
569	NM_014974	DIP2C

570	AB032251	FALZ
571	NM_148966	TNFRSF25
572	XM_933175	LOC644019
573	XM_928400	LOC653349
574	NM_182557	BCL9L
575	NM_002096	GTF2F1
576	NM_003097	SNRPN
577	BC093636	ARCN1
578	NM_001935	DPP4
579	NM_052961	SLC26A8
580	BC101104	NUP62
581	BX640829	PEO1
582	NM_005692	ABCF2
583	XM_931510	LOC389901
584	NM_015111	N4BP3
585	NM_006644	HSPH1
586	AK124794	N/A
587	AF329733	MAGED4
588	BC014900	MGC16824
589	AF293368	CBWD3
590	BC071758	TRA@
591	NM_016235	GPRC5B
592	AF294627	LEF1
593	BC039695	JAK2
594	NM_001002861	ARHGEF5
595	XM_933089	LOC644019
596	NM_018326	GIMAP4
597	NM_014258	SYCP2
598	BC010643	PFAAP5
599	NM_018441	PECR
600	BC062589	LY9
601	NM_178121	SBP1
602	BC006769	NOLC1
603	AY236962	HNRPLL
604	BC001872	IGHM
605	NM_001961	EEF2
606	XM_929349	LOC646421
607	NM_016355	DDX47
608	BC034966	N/A
609	BC059383	ANAPC4
610	NM_016337	EVL
611	NM_004843	IL27RA
612	BC106764	IL18RAP
613	NM_012228	MSRB2

614	NM_002000	FCAR
615	AK095649	NDRG1
616	NM_003038	SLC1A4
617	NM_004994	MMP9
618	BC001660	RNASET2
619	NM_052880	MGC17330
620	NM_005165	ALDOC
621	BC093975	IL18R1
622	AK056533	ABCA5
623	BC111987	PITRM1
624	AK024120	N/A
625	BC035774	SMCHD1
626	NM_005449	FAIM3
627	AK129657	N/A
628	BC070364	TRA@
629	BC043503	MEIS1
630	XM_926299	LOC653165
631	BC111024	MCTP2
632	NM_012367	OR2B6
633	BC082971	N/A
634	NM_005049	PWP2H
635	BC032805	ZNF616
636	NM_002158	HTLF
637	AK131383	N/A
638	NM_006162	NFATC1
639	NM_012108	BRDG1
640	NM_001008744	TDP1
641	XM_935181	PXDN
642	BX537840	CNOT1
643	AB063318	CCT6A
644	XM_930579	LOC653820
645	BC026092	ZNF165
646	NM_018590	GALNACT-2
647	NM_052963	TOP1MT
648	NM_003032	ST6GAL1
649	XM_933604	LOC644037
650	NM_181716	PRR6
651	NM_152683	FLJ33167
652	BX538238	MALAT1
653	NM_016122	CCDC41
654	D14689	NUP214
655	NM_020933	ZNF317
656	XM_928822	LOC644019
657	AK125658	PAPLN

658	BC051305	TRPV2
659	XM_928559	LOC653636
660	BC106925	ALDOC
661	NM_024523	GCC1
662	AB011163	KIF1B
663	NM_020536	CSRP2BP
664	NM_145795	DONSON
665	BC038381	BCAS4
666	NM_138609	H2AFY
667	CR601704	CBWD3
668	BC105106	MDM4
669	NM_000387	SLC25A20
670	NM_181717	HCG27
671	BC032941	EIF4ENIF1
672	NM_001001290	SLC2A9
673	XM_933367	LOC653458
674	NM_032328	EFCAB2
675	AK000925	N/A
676	NM_001008938	CKAP5
677	NM_031299	CDCA3
678	AK223148	TMEM19
679	NM_002646	PIK3C2B
680	NM_014755	SERTAD2
681	BC067299	MDM4
682	NM_017528	WBSCR22
683	AF064771	DGKA
684	NM_017935	BANK1
685	XM_933352	LOC653458
686	NM_007063	TBC1D8
687	AK026065	NMNAT1
688	NM_003320	TUB
689	NM_001155	ANXA6
690	NM_001003407	ABLIM1
691	NM_004697	PRPF4
692	XM_926051	LOC642548
693	AY437879	PXK
694	NM_000417	IL2RA
695	AF068181	BLNK
696	BC021276	IGHD
697	NM_000693	ALDH1A3
698	NM_004612	TGFBR1
699	NM_002660	PLCG1
700	BC029167	CD177
701	AF153604	USP15

702	AK001283	ASB3
703	BC041465	HCG27
704	NM_001001712	LCN10
705	NM_001979	EPHX2
706	NM_001031700	C4orf18
707	NM_052817	MID2
708	NM_004358	CDC25B
709	BC032756	HIST1H2AM
710	AF104304	ZFYVE9
711	AK092439	MOSC1
712	XM_934270	LOC647065
713	NM_015103	PLXND1
714	NM_015714	G0S2
715	AK097259	N/A
716	XM_927706	LOC644591
717	XM_933200	LOC644019
718	NM_080476	CDC91L1
719	BC021098	BCL11A
720	BC001976	SF3A1
721	XM_930964	FLJ31033
722	NM_138799	OACT2
723	XM_372780	LOC391045
724	XM_933226	LOC644019
725	AK001077	TANC2
726	BC101710	OR5P2
727	NM_032369	HVCN1
728	NM_015627	LDLRAP1
729	AK094682	N/A
730	NM_138450	ARL11
731	NM_004111	FEN1
732	NM_182568	FLJ36492
733	BC033095	HIST1H3D
734	NM_014760	TATDN2
735	NM_001783	CD79A
736	AF064770	DGKA
737	NM_001924	GADD45A
738	NM_001017396	ZNF2
739	BC051304	TNS1
740	NM_006803	AP3M2
741	NM_000443	ABCB4
742	NM_022805	SNRPN
743	NM_000574	CD55
744	NM_022806	SNRPN
745	BC044578	KIAA0748

746	BC008482	VDAC1
747	NM_024430	PSTPIP2
748	BC035161	CRY2
749	BC010125	C3orf37
750	NM_080926	LOC96610
751	NM_002291	LAMB1
752	BC001661	ABCF2
753	BC039322	RBM17
754	NM_006850	IL24
755	NM_001002796	MCTP1
756	NM_001013706	LOC440503
757	NM_138792	LEO1
758	NM_021602	CD79B
759	NM_002237	KCNG1
760	BC022571	KIAA0367
761	NM_007268	VSIG4
762	AK122891	ABLIM1
763	BC104256	PKIG
764	NM_012482	ZNF281
765	AK000173	TBC1D2B
766	BC054003	RNASEN
767	NM_002225	IVD
768	NM_207337	LOC196394
769	NM_021813	BACH2
770	BC013331	H2AFY
771	NM_020860	STIM2
772	NM_007355	HSP90AB1
773	NM_002053	GBP1
774	AK026467	FLJ22814
775	BC105000	EPHA4
776	NM_018334	LRRN3
777	NM_020406	CD177
778	NM_004881	TP53I3
779	BC037212	CBWD1
780	NM_175622	MT1JP
781	AK025997	MCTP1
782	BC007964	CD38
783	AK127541	AEBP1
784	AB209736	SNX9
785	NM_002807	PSMD1
786	BC096265	VNN1
787	AK074066	GEMIN5
788	BC042818	NGFRAP1L1
789	BC104920	ABCB5

790	NM_181879	LILRA5
791	NM_004666	VNN1
792	BC024767	HNRPU
793	NM_001273	CHD4
794	NM_016947	C6orf48
795	NM_017699	SIDT1
796	NM_173854	SLC41A1
797	NM_014900	COBLL1
798	BC012493	LCHN
799	AK098368	UBQLN4
800	BC103840	C20orf127
801	XM_211305	C17orf60
802	AK126264	DPY19L3
803	XM_034819	ZNF629
804	BC071860	LDHB
805	NM_178460	SIRPD
806	AK128204	N/A
807	BC039345	LOC162632
808	AK054645	RP11-138L21.1
809	NM_030917	FIP1L1
810	NM_006519	DYNLT1
811	NM_001039664	TNFRSF25
812	NM_000148	FUT1
813	NM_001012978	NGFRAP1L1
814	AK097639	DC36
815	XM_927976	LOC644891
816	NM_001024916	CBWD5
817	NM_198542	MGC4728
818	BC107141	C14orf29
819	AB023162	BAHD1
820	NM_015362	C17orf81
821	NM_004668	MGAM
822	BC003579	RFP2
823	XM_929502	MEGF9
824	XM_931897	LOC653210
825	NM_004741	NOLC1
826	NM_080672	PHACTR3
827	NM_173681	ATG9B
828	NM_145243	OMA1
829	AY462284	CDK5RAP1
830	NM_001457	FLNB
831	NM_138357	C10orf42
832	BC093784	SART3
833	BC018435	FLJ11259

834	NM_003853	IL18RAP
835	XM_932139	LOC644349
836	XM_928978	LOC646014
837	BC014479	PXK
838	NM_152272	CHMP7
839	XM_928052	LOC644980
840	NM_003670	BHLHB2
841	BX641121	N/A
842	NM_172175	IL15
843	XM_929002	LOC653458
844	NM_180991	SLCO4C1
845	BC000120	GTF2F1
846	NM_001039906	FLJ30064
847	XM_926024	LOC347364
848	BC027258	BCL2
849	AK058102	QRICH2
850	AK025211	ASAH1
851	BC074752	KCNJ1
852	XM_931886	LOC643680
853	AK025351	TRAF3IP2
854	NM_001004065	AKAP2
855	NM_003959	HIP1R
856	AK026076	ZNF395
857	AJ224442	WBSCR22
858	XM_930380	LOC645515
859	NM_001001586	ATP1A1
860	NM_000416	IFNGR1
861	NM_024708	ASB7
862	NM_004130	GYG1
863	NM_031458	PARP9
864	AY831680	N/A
865	NM_182909	DOC1
866	NM_024325	ZNF343
867	NM_003777	DNAH11
868	NM_014749	KIAA0586
869	BC069388	IRAK3
870	NM_015187	KIAA0746
871	XM_933204	LOC644019
872	BC000344	SAE1
873	NM_006640	40057
874	NM_006145	DNAJB1
875	NM_153187	SLC22A1
876	AF234262	N/A
877	NM_007115	TNFAIP6

878	NM_001715	BLK
879	AK026318	N/A
880	NM_033053	DMRTC1
881	BC033902	CLSTN1
882	AB040529	MAGED4
883	NM_001628	AKR1B1
884	XM_497141	LOC441511
885	NM_016226	VPS29
886	NM_000423	KRT2A
887	XM_932458	LOC654345
888	BC039820	L3MBTL
889	NM_130785	TPTE2
890	NM_001444	FABP5
891	XM_932805	LOC645315
892	XM_372262	LOC389901
893	XM_929611	LOC646668
894	AK027057	ZNF33A
895	AF329493	FCRLM1
896	BC106077	TRDV2
897	NM_004925	AQP3
898	AK128061	N/A
899	BC039871	CCPG1
900	BC052285	TOP1MT
901	NM_004665	VNN2
902	AK056172	LOC554203
903	NM_207325	DPY19L3
904	NM_015033	FNBP1
905	BC109306	CD22
906	AK097143	FLJ39824
907	AY254198	BCNP1
908	NM_002129	HMGB2
909	BC074735	FLJ20054
910	NM_020734	FAM80B
911	AK098633	HORMAD1
912	BC003006	MAL
913	XM_933103	LOC644019
914	NM_002015	FOXO1A
915	NM_021200	PLEKHB1
916	XM_371248	LOC374973
917	NM_016613	C4orf18
918	BC000903	HMGB2
919	NM_032127	C11orf56
920	NM_001035223	RGL3
921	NM_017895	DDX27

922	NM_006714	SMPDL3A
923	NM_000717	CA4
924	BC104902	ANAPC1
925	BC026930	TRIM38
926	BC106070	N/A
927	NM_030588	DHX9
928	NM_004893	H2AFY
929	NM_001039573	SEC14L1
930	BC106955	NGFRAP1L1
931	BC038289	CCL14
932	BC000609	KIAA0738
933	CR603690	WBSCR22
934	NM_021950	MS4A1
935	BC071759	FAM111A
936	BC008847	DDX24
937	NM_001838	CCR7
938	NM_025216	WNT10A
939	BC066647	KIAA0586
940	BC011240	NDRG2
941	NM_207585	IFNAR2
942	NM_181985	LILRA5
943	BC063502	FLJ37953
944	AK056657	STX6
945	BC012389	TSPAN6
946	M84810	TFCP2
947	AK092379	N/A
948	BC017377	MLSTD2
949	AF212234	PECR
950	NM_016327	UPB1
951	NM_006649	UTP14A
952	BC021742	ECRG4
953	BC008069	STAC3
954	NM_025030	RP4-692D3.1
955	NM_017487	ADD2
956	NM_000948	PRL
957	NM_005381	NCL
958	XM_933270	LOC653361
959	BC003070	GATA3
960	AK025521	CRTC3
961	BC101274	ANKDD1A
962	NM_015154	MESDC2
963	NM_020448	NPAL3
964	BC066596	SMARCA2
965	CR749466	CMAH

966	NM_022804	SNURF
967	NM_001039670	HOM-TES-103
968	NM_033445	HIST3H2A
969	XM_932315	GTF2IP1
970	BC000033	GYG1
971	NM_002371	MAL
972	XM_929153	LOC646197
973	NM_000578	SLC11A1
974	BC041387	MCTP2
975	NM_013399	C16orf5
976	BC009617	ECHDC3
977	NM_002685	EXOSC10
978	NM_001025	RPS23
979	AF070658	C17orf81
980	XM_293332	LOC340549
981	NM_001635	AMPH
982	NM_001721	BMX
983	BC034824	N/A
984	CR619928	MAP3K8
985	NM_018607	PRO1853
986	NM_012346	NUP62
987	BC093876	SMURF2
988	NM_020728	FAM62B
989	NM_001618	PARP1
990	AF151839	DREV1
991	NM_152852	MS4A6A
992	NM_001771	CD22
993	BC003549	ECHDC1
994	NM_032044	REG4
995	BC041802	SERHL2
996	XM_933222	LOC644019
997	AF057569	HUWE1
998	NM_032509	RBM13
999	AK123773	N/A
1000	AF305817	STX16

Table 1A

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
1	0.08	106%
2	0.04	103%
3	-0.04	97%
4	-0.07	95%
5	0.08	106%
6	-0.04	97%
7	0.05	104%
8	0.04	103%
9	0.06	104%
10	0.08	106%
11	0.09	106%
12	-0.05	97%
13	0.03	102%
14	0.05	103%
15	0.09	106%
16	0.03	102%
17	0.05	104%
18	0.08	106%
19	-0.07	96%
20	0.02	101%
21	-0.08	95%
22	0.03	102%
23	0.08	106%
24	-0.02	98%
25	-0.04	97%
26	-0.04	97%
27	0.04	103%
28	0.14	110%
29	0.06	104%
30	0.03	102%
31	0.05	104%
32	0.04	103%
33	0.05	104%
34	0.04	103%
35	-0.05	97%
36	0.03	102%
37	0.04	103%
38	-0.03	98%
39	0.12	109%
40	0.02	102%
41	-0.08	95%
42	-0.02	99%
43	-0.03	98%
44	0.02	102%
45	-0.07	96%
46	-0.08	95%
47	-0.04	97%
48	0.04	103%
49	-0.09	94%
50	-0.05	96%
51	-0.06	96%
52	0.06	104%
53	0.07	105%
54	-0.06	96%
55	0.03	102%
56	0.06	104%
57	-0.03	98%
58	-0.08	95%
59	0.09	107%
60	0.06	104%
61	-0.09	94%
62	-0.05	96%
63	-0.04	97%
64	0.03	102%
65	0.06	104%
66	-0.03	98%
67	0.04	103%
68	0.04	103%
69	0.06	104%
70	-0.04	97%
71	0.04	103%
72	0.06	104%
73	-0.08	94%
74	0.03	102%
75	-0.09	94%
76	-0.04	97%
77	0.06	104%
78	0.04	103%
79	-0.03	98%
80	0.05	104%
81	-0.09	94%
82	0.05	104%
83	-0.07	95%
84	-0.06	96%
85	0.05	104%
86	0.04	103%
87	-0.07	95%
88	0.03	102%
89	0.03	102%
90	0.03	102%
91	-0.04	97%
92	-0.05	97%
93	-0.06	96%
94	0.05	104%
95	0.10	107%
96	-0.03	98%
97	0.07	105%
98	-0.03	98%
99	0.03	102%
100	-0.04	97%
101	0.05	103%
102	-0.04	97%
103	-0.04	97%
104	0.02	101%
105	-0.08	94%
106	-0.04	97%
107	0.05	104%
108	0.10	108%
109	0.04	103%
110	-0.04	97%
111	-0.04	97%
112	-0.05	97%
113	0.04	103%
114	-0.03	98%
115	0.04	103%
116	0.05	103%
117	-0.06	96%
118	0.07	105%
119	-0.06	96%
120	0.05	103%
121	-0.06	96%
122	0.03	102%
123	0.02	102%
124	-0.03	98%
125	-0.02	98%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
126	-0.04	97%
127	0.06	104%
128	0.15	111%
129	0.03	102%
130	0.04	103%
131	0.03	102%
132	0.04	103%
133	0.07	105%
134	-0.02	99%
135	0.05	103%
136	-0.05	96%
137	0.05	103%
138	-0.09	94%
139	0.01	101%
140	0.05	103%
141	0.05	104%
142	-0.04	97%
143	0.03	102%
144	0.05	103%
145	-0.09	94%
146	0.01	101%
147	-0.06	96%
148	0.05	103%
149	-0.09	94%
150	-0.04	97%
151	0.12	109%
152	-0.03	98%
153	0.17	112%
154	0.06	105%
155	0.05	104%
156	0.09	107%
157	-0.10	93%
158	-0.03	98%
159	0.12	109%
160	-0.06	96%
161	0.04	103%
162	-0.06	96%
163	0.05	103%
164	-0.05	96%
165	-0.05	96%
166	-0.04	97%
167	-0.06	96%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
168	0.05	103%
169	-0.04	97%
170	0.07	105%
171	-0.09	94%
172	-0.05	97%
173	-0.03	98%
174	0.04	103%
175	-0.04	97%
176	0.03	102%
177	0.05	104%
178	-0.08	95%
179	-0.10	94%
180	-0.08	95%
181	-0.12	92%
182	-0.04	97%
183	-0.06	96%
184	0.06	104%
185	0.08	105%
186	0.05	103%
187	0.04	103%
188	-0.05	96%
189	-0.05	97%
190	0.06	104%
191	0.12	109%
192	0.07	105%
193	-0.05	97%
194	-0.03	98%
195	0.07	105%
196	-0.05	97%
197	-0.04	97%
198	-0.03	98%
199	0.05	104%
200	-0.01	99%
201	0.06	104%
202	-0.08	94%
203	-0.03	98%
204	0.06	104%
205	-0.05	96%
206	-0.04	98%
207	-0.07	95%
208	0.05	104%
209	0.05	104%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
210	0.03	102%
211	-0.02	98%
212	-0.07	96%
213	-0.03	98%
214	0.08	106%
215	-0.05	96%
216	0.05	104%
217	-0.06	96%
218	-0.09	94%
219	0.04	103%
220	0.06	104%
221	0.05	104%
222	0.04	103%
223	-0.08	94%
224	0.03	102%
225	0.03	102%
226	0.02	102%
227	-0.04	97%
228	-0.06	96%
229	-0.02	98%
230	-0.04	98%
231	-0.07	95%
232	0.03	102%
233	0.07	105%
234	-0.04	97%
235	-0.05	97%
236	-0.04	97%
237	-0.05	96%
238	-0.04	97%
239	0.04	103%
240	-0.07	96%
241	0.13	109%
242	-0.03	98%
243	-0.05	97%
244	-0.04	97%
245	-0.01	99%
246	-0.07	95%
247	-0.08	94%
248	-0.04	98%
249	-0.08	95%
250	-0.04	97%
251	-0.05	96%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
252	0.08	105%
253	-0.07	95%
254	0.04	103%
255	-0.05	97%
256	-0.05	97%
257	-0.05	96%
258	0.11	108%
259	0.03	102%
260	0.03	102%
261	-0.10	93%
262	-0.04	97%
263	-0.07	95%
264	-0.04	98%
265	0.05	103%
266	-0.05	97%
267	0.05	103%
268	-0.02	99%
269	-0.05	96%
270	-0.08	95%
271	0.05	104%
272	0.03	102%
273	-0.05	97%
274	-0.05	97%
275	0.06	104%
276	0.04	103%
277	0.05	103%
278	-0.04	98%
279	-0.04	97%
280	0.11	108%
281	0.09	106%
282	-0.05	97%
283	0.02	102%
284	-0.06	96%
285	-0.06	96%
286	-0.03	98%
287	0.08	106%
288	0.06	104%
289	-0.04	97%
290	-0.02	98%
291	0.04	103%
292	-0.10	94%
293	0.05	104%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
294	-0.03	98%
295	0.08	106%
296	-0.05	97%
297	0.10	107%
298	-0.06	96%
299	0.09	107%
300	0.09	106%
301	-0.05	97%
302	0.05	104%
303	-0.05	97%
304	0.04	103%
305	0.04	103%
306	-0.05	97%
307	0.09	106%
308	-0.05	97%
309	-0.03	98%
310	0.03	102%
311	-0.05	97%
312	0.07	105%
313	0.03	102%
314	-0.05	97%
315	0.03	102%
316	0.11	108%
317	-0.04	97%
318	0.06	104%
319	0.06	105%
320	-0.03	98%
321	0.06	104%
322	-0.02	98%
323	0.10	107%
324	0.10	107%
325	-0.07	95%
326	0.04	103%
327	-0.07	95%
328	0.05	104%
329	-0.03	98%
330	0.04	103%
331	0.07	105%
332	-0.05	97%
333	-0.03	98%
334	-0.07	95%
335	-0.04	97%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
336	0.04	102%
337	0.08	105%
338	-0.07	95%
339	-0.03	98%
340	-0.07	95%
341	0.06	104%
342	-0.09	94%
343	-0.04	97%
344	-0.04	97%
345	-0.07	95%
346	-0.06	96%
347	0.06	104%
348	0.06	104%
349	-0.04	97%
350	0.11	108%
351	0.14	110%
352	-0.08	95%
353	0.05	103%
354	0.09	106%
355	0.11	108%
356	-0.07	95%
357	0.06	104%
358	-0.06	96%
359	-0.06	96%
360	0.05	103%
361	0.02	101%
362	-0.03	98%
363	-0.05	97%
364	-0.05	97%
365	-0.05	97%
366	-0.04	98%
367	0.05	104%
368	-0.08	94%
369	0.06	104%
370	-0.05	97%
371	0.03	102%
372	0.05	103%
373	-0.03	98%
374	0.08	106%
375	0.09	106%
376	-0.04	97%
377	0.03	102%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
378	0.05	103%
379	0.05	103%
380	0.07	105%
381	-0.05	97%
382	0.06	104%
383	0.03	102%
384	-0.05	97%
385	-0.06	96%
386	-0.04	97%
387	0.05	103%
388	-0.04	97%
389	-0.05	97%
390	-0.03	98%
391	0.04	103%
392	-0.10	94%
393	-0.03	98%
394	-0.06	96%
395	0.09	107%
396	-0.05	97%
397	-0.03	98%
398	-0.06	96%
399	-0.04	98%
400	0.06	104%
401	0.04	103%
402	-0.08	95%
403	0.08	106%
404	0.08	105%
405	0.12	109%
406	0.13	110%
407	-0.04	97%
408	0.04	103%
409	0.06	104%
410	-0.07	95%
411	0.02	101%
412	-0.07	95%
413	0.04	103%
414	0.08	105%
415	0.05	103%
416	0.07	105%
417	0.05	104%
418	-0.04	97%
419	-0.04	97%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
420	-0.07	95%
421	-0.03	98%
422	0.05	103%
423	-0.09	94%
424	-0.05	96%
425	0.07	105%
426	-0.07	95%
427	0.03	102%
428	0.04	103%
429	-0.05	97%
430	-0.07	95%
431	0.05	104%
432	0.08	105%
433	0.06	104%
434	0.05	103%
435	-0.04	98%
436	0.03	102%
437	-0.02	99%
438	0.06	104%
439	0.13	109%
440	0.04	103%
441	-0.04	97%
442	0.10	107%
443	-0.03	98%
444	0.04	103%
445	0.07	105%
446	-0.07	95%
447	0.02	102%
448	0.04	103%
449	-0.05	97%
450	-0.07	95%
451	-0.07	95%
452	-0.09	94%
453	0.06	104%
454	0.04	103%
455	-0.03	98%
456	0.06	104%
457	-0.05	97%
458	0.06	104%
459	0.09	106%
460	-0.04	97%
461	0.07	105%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
462	-0.05	97%
463	-0.06	96%
464	-0.04	97%
465	-0.05	97%
466	0.07	105%
467	-0.03	98%
468	-0.03	98%
469	0.03	102%
470	0.02	102%
471	-0.07	95%
472	0.05	103%
473	0.05	104%
474	-0.03	98%
475	-0.07	95%
476	-0.05	97%
477	-0.06	96%
478	-0.08	95%
479	0.06	105%
480	-0.05	97%
481	-0.07	95%
482	-0.05	97%
483	0.04	103%
484	-0.06	96%
485	0.04	103%
486	-0.05	97%
487	-0.07	95%
488	0.07	105%
489	-0.04	97%
490	-0.06	96%
491	0.06	104%
492	0.05	104%
493	-0.02	99%
494	0.02	102%
495	0.10	107%
496	-0.04	98%
497	-0.04	97%
498	0.02	101%
499	-0.06	96%
500	-0.02	99%
501	-0.05	96%
502	0.04	103%
503	0.05	103%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
504	-0.05	97%
505	0.04	102%
506	0.06	105%
507	-0.04	97%
508	0.05	103%
509	0.04	103%
510	-0.04	97%
511	0.04	103%
512	-0.05	97%
513	0.05	103%
514	-0.04	97%
515	-0.06	96%
516	-0.07	95%
517	-0.08	95%
518	-0.02	99%
519	-0.04	97%
520	-0.06	96%
521	-0.05	97%
522	-0.04	98%
523	-0.03	98%
524	0.03	102%
525	0.06	104%
526	0.05	104%
527	-0.05	96%
528	0.04	103%
529	-0.08	94%
530	0.06	104%
531	0.03	102%
532	-0.04	97%
533	0.02	102%
534	0.01	101%
535	0.07	105%
536	-0.07	95%
537	-0.05	96%
538	-0.05	97%
539	-0.04	97%
540	-0.02	99%
541	-0.04	97%
542	-0.04	97%
543	-0.10	93%
544	-0.05	97%
545	0.06	104%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
546	0.02	102%
547	0.05	103%
548	-0.03	98%
549	0.04	103%
550	-0.05	97%
551	-0.05	97%
552	0.06	104%
553	0.06	104%
554	0.04	103%
555	0.09	107%
556	-0.04	97%
557	0.09	106%
558	-0.05	97%
559	-0.05	97%
560	0.04	103%
561	0.02	102%
562	-0.08	95%
563	0.05	104%
564	0.07	105%
565	0.06	104%
566	0.05	103%
567	-0.05	97%
568	0.11	108%
569	-0.05	97%
570	-0.02	99%
571	-0.01	99%
572	0.04	103%
573	-0.04	97%
574	-0.05	97%
575	-0.06	96%
576	-0.05	97%
577	-0.04	97%
578	-0.09	94%
579	0.08	106%
580	-0.05	96%
581	-0.05	96%
582	-0.06	96%
583	-0.03	98%
584	-0.04	97%
585	-0.05	96%
586	-0.05	97%
587	-0.06	96%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
588	-0.03	98%
589	0.03	102%
590	-0.07	96%
591	-0.11	93%
592	-0.10	93%
593	0.08	106%
594	-0.03	98%
595	0.03	102%
596	-0.02	99%
597	0.06	104%
598	0.04	103%
599	0.05	103%
600	-0.03	98%
601	-0.07	95%
602	-0.04	97%
603	0.04	103%
604	-0.08	94%
605	-0.04	97%
606	0.05	104%
607	-0.04	97%
608	0.09	106%
609	0.04	102%
610	-0.05	97%
611	-0.05	97%
612	0.07	105%
613	0.07	105%
614	0.12	109%
615	-0.02	99%
616	-0.04	97%
617	0.12	108%
618	-0.02	99%
619	-0.04	97%
620	-0.07	95%
621	0.10	107%
622	0.05	103%
623	-0.04	97%
624	0.02	101%
625	0.04	103%
626	-0.04	97%
627	0.06	105%
628	-0.08	94%
629	-0.06	96%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
630	0.04	103%
631	0.09	106%
632	0.09	107%
633	-0.06	96%
634	-0.05	97%
635	-0.06	96%
636	0.07	105%
637	0.06	104%
638	-0.07	95%
639	-0.04	97%
640	-0.06	96%
641	-0.02	99%
642	-0.03	98%
643	-0.02	99%
644	0.06	104%
645	-0.03	98%
646	0.07	105%
647	-0.04	97%
648	-0.05	97%
649	-0.04	97%
650	-0.06	96%
651	0.05	103%
652	0.05	104%
653	0.05	104%
654	-0.04	97%
655	-0.04	97%
656	0.04	103%
657	0.06	104%
658	-0.04	97%
659	0.07	105%
660	-0.08	95%
661	-0.05	96%
662	0.07	105%
663	-0.04	97%
664	0.02	101%
665	-0.06	96%
666	0.02	102%
667	0.04	103%
668	0.06	104%
669	0.05	104%
670	0.03	102%
671	-0.04	97%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
672	0.02	101%
673	-0.05	97%
674	0.03	102%
675	0.07	105%
676	-0.05	97%
677	0.02	101%
678	-0.05	96%
679	-0.07	96%
680	-0.04	97%
681	0.06	104%
682	-0.04	97%
683	-0.04	97%
684	-0.08	94%
685	-0.04	97%
686	0.04	103%
687	0.03	102%
688	-0.07	95%
689	-0.05	97%
690	-0.06	96%
691	-0.05	97%
692	-0.06	96%
693	0.02	101%
694	-0.07	95%
695	-0.07	95%
696	-0.11	92%
697	0.03	102%
698	0.04	103%
699	-0.05	97%
700	0.12	109%
701	0.02	102%
702	0.03	102%
703	0.03	102%
704	-0.04	98%
705	-0.08	95%
706	0.08	106%
707	-0.05	97%
708	-0.05	97%
709	0.05	103%
710	-0.07	95%
711	0.07	105%
712	0.04	103%
713	-0.04	97%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
714	0.12	109%
715	0.04	103%
716	0.06	104%
717	0.04	103%
718	-0.03	98%
719	-0.04	97%
720	-0.05	97%
721	0.08	105%
722	0.09	107%
723	0.05	104%
724	0.04	102%
725	0.08	106%
726	-0.01	100%
727	-0.03	98%
728	-0.07	95%
729	-0.09	94%
730	0.07	105%
731	-0.03	98%
732	0.11	108%
733	0.11	108%
734	-0.03	98%
735	-0.07	95%
736	-0.04	97%
737	0.04	103%
738	-0.05	97%
739	-0.04	97%
740	-0.05	97%
741	-0.08	95%
742	-0.04	97%
743	0.03	102%
744	-0.04	97%
745	-0.07	95%
746	-0.05	96%
747	0.08	106%
748	-0.05	96%
749	-0.05	96%
750	-0.05	97%
751	0.05	104%
752	-0.06	96%
753	-0.02	98%
754	-0.07	95%
755	0.03	102%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
756	0.05	104%
757	-0.05	97%
758	-0.06	96%
759	-0.04	98%
760	0.13	109%
761	0.08	106%
762	-0.05	97%
763	-0.08	95%
764	0.05	104%
765	0.05	103%
766	-0.07	95%
767	-0.03	98%
768	0.06	104%
769	-0.12	92%
770	0.02	101%
771	0.02	102%
772	-0.06	96%
773	0.06	104%
774	-0.05	97%
775	-0.08	95%
776	-0.13	92%
777	0.13	110%
778	0.08	106%
779	0.04	103%
780	0.04	103%
781	0.04	103%
782	0.07	105%
783	-0.09	94%
784	-0.04	97%
785	-0.04	97%
786	0.10	108%
787	-0.06	96%
788	-0.07	95%
789	-0.01	99%
790	0.04	103%
791	0.07	105%
792	-0.02	99%
793	-0.04	97%
794	-0.02	99%
795	-0.05	96%
796	-0.06	96%
797	-0.03	98%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
798	-0.04	97%
799	-0.05	97%
800	0.05	103%
801	0.07	105%
802	0.04	103%
803	-0.07	95%
804	-0.05	96%
805	0.04	103%
806	0.05	104%
807	0.06	104%
808	0.04	103%
809	-0.02	99%
810	0.03	102%
811	-0.04	97%
812	-0.08	94%
813	-0.07	95%
814	0.04	103%
815	0.05	103%
816	0.04	103%
817	-0.04	98%
818	0.06	104%
819	-0.04	97%
820	-0.04	97%
821	0.05	104%
822	0.01	101%
823	0.05	104%
824	-0.06	96%
825	-0.05	96%
826	0.05	104%
827	-0.11	92%
828	0.03	102%
829	-0.04	97%
830	-0.07	96%
831	0.03	102%
832	-0.03	98%
833	0.03	102%
834	0.07	105%
835	0.07	105%
836	0.05	103%
837	0.06	104%
838	-0.04	97%
839	-0.13	91%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
840	-0.06	96%
841	0.08	106%
842	0.04	103%
843	-0.06	96%
844	0.07	105%
845	-0.04	97%
846	0.08	106%
847	-0.02	98%
848	-0.04	98%
849	-0.01	99%
850	0.04	103%
851	0.05	103%
852	0.05	104%
853	-0.05	97%
854	-0.06	96%
855	-0.03	98%
856	-0.05	96%
857	-0.04	97%
858	-0.03	98%
859	-0.07	95%
860	0.06	104%
861	0.04	103%
862	0.04	103%
863	0.03	102%
864	0.06	104%
865	-0.04	98%
866	-0.03	98%
867	-0.06	96%
868	-0.03	98%
869	0.06	104%
870	-0.04	97%
871	0.04	103%
872	-0.04	97%
873	-0.05	97%
874	-0.02	99%
875	0.06	104%
876	0.10	108%
877	0.11	108%
878	-0.04	97%
879	0.06	104%
880	-0.08	95%
881	-0.04	97%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
882	-0.05	96%
883	-0.05	97%
884	0.03	102%
885	0.03	102%
886	-0.03	98%
887	-0.06	96%
888	0.01	101%
889	0.06	104%
890	0.05	103%
891	0.03	102%
892	-0.03	98%
893	0.07	105%
894	-0.03	98%
895	-0.04	97%
896	-0.08	94%
897	-0.09	94%
898	0.05	104%
899	0.06	104%
900	-0.04	97%
901	0.03	102%
902	0.05	104%
903	0.07	105%
904	-0.02	99%
905	-0.08	95%
906	0.03	102%
907	-0.09	94%
908	0.06	105%
909	0.04	103%
910	0.01	101%
911	0.09	107%
912	-0.09	94%
913	0.04	103%
914	-0.04	97%
915	-0.10	93%
916	0.06	104%
917	0.08	106%
918	0.07	105%
919	0.06	104%
920	0.04	103%
921	-0.05	96%
922	0.05	104%
923	0.06	104%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
924	-0.04	97%
925	0.01	101%
926	0.12	108%
927	-0.03	98%
928	0.02	101%
929	-0.01	99%
930	-0.07	96%
931	0.10	107%
932	-0.04	97%
933	-0.04	97%
934	-0.05	97%
935	0.02	102%
936	-0.06	96%
937	-0.08	95%
938	-0.09	94%
939	-0.04	97%
940	-0.05	97%
941	0.05	103%
942	0.09	106%
943	0.06	104%
944	0.03	102%
945	-0.08	94%
946	-0.03	98%
947	0.03	102%
948	0.08	106%
949	0.04	103%
950	0.06	105%
951	-0.05	97%
952	-0.09	94%
953	0.01	101%
954	0.05	104%
955	-0.02	98%
956	0.09	106%
957	-0.04	97%
958	0.04	103%
959	-0.07	96%
960	-0.06	96%
961	0.05	103%
962	-0.04	97%
963	-0.04	97%
964	-0.05	97%
965	0.07	105%

SEQ ID NO	Log-2 Fold Change	Non-Log Fold Change
966	-0.04	97%
967	0.03	102%
968	0.04	103%
969	0.09	106%
970	0.05	104%
971	-0.09	94%
972	-0.05	97%
973	0.03	102%
974	0.06	104%
975	-0.06	96%
976	0.07	105%
977	-0.04	97%
978	-0.07	95%
979	-0.05	97%
980	-0.02	99%
981	0.08	105%
982	0.09	106%
983	0.03	102%
984	0.06	104%
985	0.05	103%
986	-0.04	97%
987	0.03	102%
988	-0.04	97%
989	-0.04	97%
990	0.04	103%
991	0.02	101%
992	-0.09	94%
993	0.05	103%
994	-0.13	91%
995	-0.05	96%
996	0.04	103%
997	-0.03	98%
998	-0.02	98%
999	0.07	105%
1000	0.04	103%

SEQ ID Nos for Examples 1-3 of length 250		
Example 1	Example 2	Example 3
1	1	4
3	2	5
5	8	8
10	12	10
12	17	12
20	26	13
25	28	15
29	30	17
33	34	18
35	46	25
42	48	26
46	52	28
51	56	29
52	58	30
54	60	34
59	62	35
67	67	46
75	70	51
78	72	52
80	78	59
81	80	62
87	81	67
92	82	78
93	85	80
104	88	81
107	89	82
111	93	87
113	102	93
117	113	97
119	117	102
126	118	113
132	119	117
133	122	119
136	126	126
137	128	127
140	131	136
142	134	137
145	136	141
149	137	143
153	138	145
156	149	149

SEQ ID Nos for Examples 1-3 of length 250		
Example 1	Example 2	Example 3
163	153	153
164	156	156
169	160	158
170	164	163
176	166	166
178	168	168
179	170	170
181	171	178
190	179	179
191	188	180
192	195	181
195	197	188
197	199	195
202	203	197
206	204	202
218	209	204
233	215	208
237	218	209
253	220	214
259	227	215
265	233	216
266	235	219
270	240	224
280	254	225
297	255	227
299	259	233
302	263	235
305	265	246
306	266	259
313	267	265
314	270	266
318	280	270
319	287	275
325	297	280
329	298	287
330	302	288
334	305	297
338	306	298
352	311	299
353	313	302
356	319	305

SEQ ID Nos for Examples 1-3 of length 250		
Example 1	Example 2	Example 3
357	329	306
363	330	313
365	334	318
367	338	319
374	350	325
375	351	334
385	354	337
387	357	338
390	361	350
391	365	351
392	371	354
394	372	363
396	373	365
401	375	366
403	385	374
404	386	375
405	387	379
406	391	385
409	392	390
410	394	391
415	400	392
420	401	394
422	403	401
424	404	403
426	406	406
432	420	420
438	421	424
445	422	430
449	423	432
452	424	433
454	426	442
461	427	445
463	432	448
466	447	449
467	449	452
469	467	456
473	469	458
479	473	461
481	477	468
484	484	469
485	486	477

SEQ ID Nos for Examples 1-3 of length 250		
Example 1	Example 2	Example 3
488	499	483
489	511	484
499	512	488
501	515	499
508	516	501
510	517	505
515	522	512
516	526	515
518	527	516
526	528	522
530	529	526
544	541	527
546	542	529
547	544	536
549	546	537
550	547	538
554	549	550
555	554	552
557	558	554
559	559	555
560	562	557
561	564	559
563	568	560
570	574	564
571	575	565
575	576	570
576	585	571
580	592	572
582	593	574
585	601	575
588	602	576
590	608	582
592	611	584
593	614	592
599	620	593
600	623	601
602	625	602
608	628	608
609	630	628
614	634	630
627	641	632

SEQ ID Nos for Examples 1-3 of length 250		
Example 1	Example 2	Example 3
634	647	634
637	650	638
638	653	644
640	656	652
644	660	660
652	666	666
656	667	667
660	668	668
665	675	676
667	676	679
668	678	681
675	682	689
678	689	690
679	690	699
681	692	704
682	694	706
684	699	707
689	707	711
699	724	732
704	730	733
707	732	739
719	733	743
722	743	745
732	744	749
739	749	750
741	751	754
744	756	762
745	760	764
748	762	765
749	764	769
772	765	772
774	772	775
775	774	778
778	775	779
788	779	783
798	788	796
806	791	807
807	799	810
810	801	813
813	804	814
815	806	816

SEQ ID Nos for Examples 1-3 of length 250		
Example 1	Example 2	Example 3
816	807	823
823	810	827
825	811	830
826	813	834
830	816	839
834	824	841
839	827	846
841	830	850
846	839	856
847	844	859
849	845	861
850	861	862
851	862	864
859	867	871
862	871	874
863	874	876
871	877	877
874	881	883
876	886	887
877	889	889
883	893	890
886	897	893
888	902	897
889	907	902
893	908	911
894	911	915
897	915	917
906	917	921
931	921	926
936	931	928
937	936	931
940	940	936
947	945	937
948	954	940
950	956	947
954	958	948
956	969	957
963	971	958
976	972	982
981	982	985
982	989	990

990	995	993	994	996	996	996	998	999

Table 2

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
1	1	1
2	2	3
3	4	4
5	5	5
6	8	6
8	10	8
9	11	10
11	12	12
12	13	13
16	15	15
18	16	16
22	19	17
25	22	18
28	25	22
29	27	25
30	29	26
31	30	28
32	31	29
33	32	30
36	33	31
47	34	32
51	35	34
53	37	35
56	42	39
57	44	45
58	46	46
59	48	48
64	51	50
67	55	51
75	56	52
78	59	55
80	60	58
81	61	59
82	67	60
86	69	62
88	70	67
90	72	73
96	73	75
100	78	78
102	79	79
103	82	80

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
109	85	81
110	87	82
112	89	83
113	90	85
114	100	87
117	101	89
120	102	90
122	103	92
123	105	93
125	107	97
126	108	100
130	111	101
131	114	102
133	119	104
134	120	105
136	121	107
138	122	112
139	124	113
140	125	117
141	127	118
143	129	119
144	130	122
145	131	126
146	133	127
147	136	128
149	137	133
152	140	134
153	141	136
155	143	137
157	145	140
161	146	141
163	147	143
167	148	144
168	149	145
170	150	149
172	152	151
174	153	153
178	154	156
179	155	158
180	158	159
181	160	163

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
183	161	164
184	162	166
188	165	168
191	166	170
195	170	171
196	171	172
197	178	177
199	179	178
202	181	179
203	182	180
205	183	181
206	184	182
209	187	185
211	188	188
216	190	190
218	191	191
220	192	192
223	195	193
224	196	195
225	197	197
226	199	199
228	202	202
229	204	203
231	205	204
234	206	205
241	208	206
242	209	208
246	210	209
247	211	214
249	212	215
251	214	216
252	216	217
257	217	218
258	218	219
259	219	220
260	220	224
267	224	225
268	226	227
269	229	231
270	233	233
272	234	235

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
278	235	237
280	239	238
282	240	239
284	241	240
285	242	246
287	244	252
290	247	254
292	249	255
294	254	258
295	255	259
296	257	263
297	260	265
298	265	266
302	266	267
303	268	269
305	269	270
306	270	272
310	273	275
311	275	280
313	277	286
314	280	287
315	283	288
317	284	291
319	286	297
320	287	298
324	288	299
325	290	302
326	291	303
328	298	305
330	301	306
333	302	311
334	305	313
337	306	314
339	307	315
340	308	318
346	309	319
348	311	324
354	312	325
355	313	326
357	314	329
358	317	330

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
359	319	334
361	321	336
363	323	337
364	324	338
365	325	345
367	326	346
370	328	350
371	330	351
375	332	354
376	333	355
378	334	357
379	337	361
381	338	363
383	339	365
384	344	366
385	347	369
387	348	372
388	350	373
390	351	374
391	354	375
393	355	379
395	356	384
400	359	385
401	360	386
403	361	387
404	363	390
405	366	391
406	369	392
407	374	394
412	375	401
414	376	402
416	383	403
418	385	404
419	390	405
420	391	406
426	392	410
433	394	411
434	395	413
435	397	414
436	398	415
437	400	416

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
439	401	420
441	403	421
442	404	422
446	409	423
447	410	424
450	412	426
452	414	427
454	417	430
456	419	431
458	420	432
461	421	433
463	422	439
464	424	441
465	427	442
466	429	445
468	436	446
471	437	448
472	438	449
482	439	452
483	440	456
484	441	458
485	442	459
488	443	461
490	445	463
497	447	466
498	448	467
501	449	468
502	450	469
505	451	471
506	452	473
508	455	475
509	456	477
510	457	479
512	458	481
516	460	483
522	461	484
523	462	485
524	465	486
525	466	488
526	468	489
527	469	498

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
528	475	499
529	477	501
530	481	502
531	482	504
533	483	505
537	484	507
538	487	509
543	488	510
544	489	512
545	493	513
547	495	515
549	496	516
551	499	517
554	502	518
555	506	519
556	508	520
557	509	521
558	510	522
559	512	526
564	514	527
565	515	528
569	516	529
574	517	531
576	518	536
578	519	537
579	521	538
581	522	541
582	524	542
583	535	544
584	538	545
586	539	547
587	544	549
588	545	550
590	547	552
592	549	554
593	552	555
595	553	557
597	557	559
599	558	560
603	559	562
604	560	563

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
605	561	564
608	562	565
610	564	570
614	566	571
615	567	572
619	571	574
620	572	575
625	573	576
626	574	578
627	575	580
628	576	582
630	580	584
631	582	585
632	583	588
634	588	589
636	593	592
637	594	593
638	597	597
641	602	599
642	604	601
643	606	602
644	608	606
645	611	608
651	613	611
652	617	614
655	620	617
658	626	620
659	627	621
662	628	627
663	630	628
665	632	630
667	634	631
668	635	632
670	636	634
671	638	637
675	640	638
676	642	640
677	647	644
678	649	650
679	652	652
680	654	653

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
681	655	654
683	657	656
684	658	657
686	660	659
687	661	660
690	662	661
693	665	663
695	666	666
697	667	667
700	668	668
703	673	671
704	674	672
706	675	673
708	676	674
710	677	675
711	678	676
712	679	678
713	681	679
718	682	681
719	687	682
723	688	684
724	696	689
730	697	690
731	699	691
733	700	692
735	701	695
737	702	696
738	707	699
739	713	700
740	716	704
741	717	706
742	720	707
743	723	711
744	726	715
745	729	717
747	732	722
749	733	723
751	734	724
752	735	730
754	737	732
763	738	733

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
764	739	734
765	741	736
766	743	739
770	744	741
771	745	743
772	748	744
773	749	745
777	750	747
778	751	748
779	752	749
781	753	750
788	755	751
790	756	752
791	757	754
793	758	757
796	759	760
798	760	762
800	761	764
803	765	765
804	770	769
805	774	770
806	777	772
808	778	774
810	779	775
814	785	778
816	786	779
817	787	782
818	790	783
819	791	785
821	792	787
823	796	788
824	797	791
825	801	796
826	802	797
827	805	799
829	808	801
830	810	804
831	811	805
832	815	806
833	816	807
834	819	808

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
835	821	810
837	823	813
838	827	814
839	828	816
840	830	819
841	834	823
842	837	825
843	841	827
844	842	830
845	844	834
847	845	836
848	847	839
850	850	841
856	851	843
857	853	845
858	854	846
859	856	850
860	857	851
862	859	856
864	860	857
866	861	859
867	862	860
868	863	861
870	864	862
871	865	863
873	866	864
874	868	871
876	869	874
877	873	876
878	874	877
881	875	879
883	877	881
884	878	883
886	880	887
887	883	889
888	888	890
890	889	893
892	891	896
893	893	897
895	895	902
897	896	903

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
899	902	906
902	904	907
903	907	911
904	908	915
905	910	917
907	911	921
908	913	923
911	915	924
915	916	926
917	917	927
920	919	928
921	923	930
923	925	931
925	926	934
928	928	935
931	929	936
933	930	937
935	931	939
936	934	940
938	940	944
941	941	945
946	944	947
948	945	948
950	947	950
955	948	954
957	950	956
958	951	957
959	954	958
960	958	959
961	960	963
962	965	964
963	966	965
964	969	966
967	970	969
968	971	970
969	975	972
973	976	976
976	977	982
978	979	984
980	981	985
982	982	986

SEQ ID Nos for Examples 4-6 of length 500		
Example 4	Example 5	Example 6
983	984	989
985	985	990
991	989	992
992	990	993
995	993	994
996	996	996
999	997	999

Table 3

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
1	1	1
2	2	2
4	4	3
5	5	4
6	6	5
7	8	6
8	9	8
9	10	9
10	11	10
11	13	11
12	14	12
13	15	13
14	16	14
15	17	15
17	18	16
18	19	17
20	20	18
21	21	20
22	22	21
24	23	22
25	24	23
26	25	25
27	26	26
28	27	27
29	28	28
30	29	29
31	30	30
33	31	31
34	32	32
35	33	33
36	34	34
38	35	35
39	36	37
40	38	38
42	39	39
44	41	41
45	43	42
46	44	43
47	45	44
48	46	45
49	47	46

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
50	48	48
51	49	49
52	50	50
53	51	51
54	52	52
55	53	53
57	54	54
58	55	55
59	58	56
62	59	58
63	61	59
65	62	60
66	63	62
67	64	63
71	66	65
72	68	67
73	69	68
74	70	69
76	71	70
78	72	72
79	74	73
80	75	74
81	76	75
85	77	77
87	78	78
88	79	79
89	80	80
90	81	81
91	84	82
93	85	83
94	86	85
95	87	86
96	88	87
97	89	88
98	90	89
99	91	90
100	92	91
101	93	92
102	94	93
103	95	94
104	98	96

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
105	100	97
107	102	98
108	103	99
110	105	100
111	107	101
112	108	102
113	109	104
114	110	105
115	111	107
116	112	111
117	113	112
118	114	113
119	115	116
121	117	117
122	118	118
123	119	119
125	121	121
126	123	122
127	125	126
128	127	127
129	128	128
130	130	130
131	131	131
132	132	132
133	133	133
134	136	134
135	137	136
137	140	137
139	141	138
140	142	140
141	143	141
142	144	142
143	145	143
144	147	144
145	148	145
146	150	146
147	151	147
148	153	148
149	154	149
151	156	151
152	157	152

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
153	158	153
154	159	154
155	160	156
156	161	158
158	162	159
159	163	160
161	164	163
162	165	164
163	166	165
164	167	166
165	168	168
166	169	169
167	170	170
168	171	171
169	172	172
170	173	176
172	174	177
173	175	178
174	176	179
175	177	180
177	178	181
178	179	182
179	180	183
180	181	184
182	182	185
183	183	187
184	184	188
185	185	189
186	188	190
187	189	191
188	190	192
189	192	193
190	193	195
191	194	196
192	195	197
193	196	198
194	197	199
195	198	202
196	201	203
197	202	204
199	203	205

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
200	204	206
201	205	208
202	206	209
203	207	210
204	208	211
206	209	214
207	213	215
208	214	216
209	215	217
210	216	218
211	217	219
214	218	220
215	219	221
216	221	223
217	224	224
218	225	225
219	226	227
220	227	228
221	228	230
223	229	231
225	231	232
226	232	233
227	233	235
228	235	236
229	236	237
230	237	238
231	238	239
232	239	240
236	241	241
237	242	246
238	243	248
239	244	249
240	246	250
241	248	252
243	249	253
244	250	254
245	252	255
247	253	257
248	254	258
250	255	259
251	256	261

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
252	257	263
253	258	264
254	259	265
255	261	266
256	263	267
257	264	268
258	265	269
259	266	270
260	267	272
262	268	274
263	269	275
264	270	277
265	271	279
266	272	280
267	274	282
268	275	283
271	276	285
272	277	286
273	279	287
274	280	288
275	281	291
277	282	292
278	283	294
279	284	296
280	285	297
282	286	298
283	287	299
285	288	300
286	290	301
287	291	302
289	292	303
290	293	305
291	295	306
292	297	307
293	298	310
294	299	311
296	300	312
298	302	313
300	304	314
301	305	315
302	309	318

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
303	310	319
304	311	320
306	312	321
307	314	322
308	315	323
309	317	324
310	318	325
311	319	326
312	320	327
314	321	328
315	322	329
316	323	330
317	324	331
319	325	334
320	326	336
321	327	337
323	328	338
325	329	340
326	330	341
327	332	343
328	333	344
329	334	345
330	335	346
331	336	347
332	337	349
335	338	350
338	339	351
339	340	352
340	341	353
341	344	354
342	345	355
343	346	356
348	347	357
350	348	358
351	351	359
352	352	360
353	353	361
354	354	362
355	355	363
356	358	364
357	359	365

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
358	360	366
359	361	367
361	362	368
362	364	369
363	365	370
364	366	371
365	367	372
366	368	373
367	369	374
368	370	375
369	371	376
371	372	379
372	373	380
374	374	382
375	375	383
377	376	384
379	379	385
380	380	386
381	381	387
382	382	388
385	385	389
387	386	390
388	387	391
389	389	392
390	390	393
391	391	394
392	392	396
394	394	399
395	395	400
396	396	401
397	399	402
398	400	403
399	401	404
400	403	405
401	404	406
402	406	407
403	407	408
404	409	409
405	410	410
406	411	411
407	412	413

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
410	413	414
411	414	415
412	415	416
413	416	417
414	418	420
415	419	421
416	420	422
417	421	423
418	422	424
419	424	426
420	425	427
422	426	428
423	427	430
424	428	431
426	429	432
427	430	433
428	431	435
429	432	436
430	433	437
431	435	438
432	436	439
433	437	441
434	438	442
435	439	444
437	441	445
438	442	446
440	443	447
442	445	448
443	446	449
445	448	450
446	449	452
447	450	453
448	451	454
449	452	456
450	455	457
451	456	458
452	457	459
453	458	460
454	461	461
456	462	462
457	464	463

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
458	466	464
459	467	466
461	468	467
462	469	468
463	470	469
465	472	470
466	473	471
468	474	472
469	475	473
470	476	474
471	477	475
472	479	476
473	480	477
475	481	478
476	482	479
477	483	481
478	484	483
479	486	484
480	488	485
481	489	486
482	490	487
483	491	488
484	492	489
485	493	491
488	494	492
489	495	493
490	497	495
492	498	498
493	499	499
494	500	501
495	501	502
496	502	503
499	503	504
500	504	505
501	505	506
502	508	507
504	510	508
505	511	509
507	512	510
508	513	511
509	514	512

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
510	515	513
511	516	515
513	517	516
514	518	517
515	519	518
516	520	519
517	522	520
519	524	521
520	525	522
521	526	525
522	527	526
524	528	527
525	529	528
526	530	529
527	534	530
528	536	531
529	537	533
530	538	536
532	539	537
533	540	538
536	542	539
537	544	540
538	545	541
539	546	542
540	547	543
541	549	544
542	551	545
543	552	546
545	553	547
546	554	549
547	557	550
549	558	551
550	559	552
551	560	553
552	561	554
553	562	555
554	563	557
555	564	558
556	565	559
557	566	560
559	568	561

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
560	569	562
561	570	563
562	571	564
564	572	565
566	574	567
568	575	568
569	576	570
570	577	571
571	578	572
572	579	574
573	580	575
574	581	576
575	582	577
576	583	578
577	584	580
578	585	582
582	586	584
584	588	585
585	590	586
586	591	587
587	592	588
588	593	589
589	594	590
590	596	591
591	597	592
592	598	593
593	599	595
594	601	597
595	602	599
597	604	600
598	606	601
599	607	602
600	608	603
601	610	605
602	611	606
603	612	607
604	613	608
605	614	609
606	617	610
608	618	611
610	620	612

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
611	623	613
612	624	614
613	625	615
614	627	617
616	628	619
617	629	620
618	630	621
619	631	623
620	632	625
621	633	627
623	634	628
624	635	630
626	636	631
627	637	632
628	638	633
629	641	634
630	642	635
631	644	637
633	646	638
634	647	639
636	648	640
637	649	641
638	650	643
641	652	644
643	653	645
644	654	647
645	655	649
647	656	650
648	658	652
649	659	653
650	660	654
652	661	655
656	663	656
657	664	657
658	665	659
660	666	660
661	667	661
663	668	662
664	671	663
665	672	664
666	673	665

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
667	674	666
668	675	667
669	676	668
670	677	670
671	678	671
672	679	672
673	682	673
675	683	674
676	684	675
677	685	676
678	686	677
679	688	678
680	690	679
681	691	680
683	693	681
684	694	682
687	696	683
688	698	684
689	699	685
690	700	686
693	701	688
695	703	689
697	705	690
698	706	691
699	707	692
701	708	694
702	709	695
704	710	696
705	711	699
706	712	700
707	713	703
708	714	704
709	715	706
710	717	707
711	718	708
712	719	709
713	720	710
714	721	711
715	722	712
716	724	714
717	727	715

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
718	728	717
719	729	719
720	730	720
721	731	721
722	732	722
724	733	723
725	734	724
726	735	725
727	737	726
728	738	729
729	740	730
730	742	732
731	744	733
732	745	734
733	748	735
735	749	736
737	750	737
738	751	738
739	752	739
740	753	740
741	754	741
742	755	742
743	756	743
744	757	744
745	759	745
746	760	746
747	761	747
748	762	748
749	763	749
750	764	750
751	765	751
752	769	752
756	770	753
757	772	754
758	773	756
759	774	757
763	775	760
764	776	761
765	777	762
767	779	763
768	781	764

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
769	782	765
770	783	767
771	784	769
772	785	770
773	786	771
774	787	772
775	788	773
776	789	774
777	790	775
778	791	776
779	792	777
781	793	778
782	794	779
783	795	780
784	796	781
785	797	782
787	798	783
788	799	785
789	801	786
790	802	787
791	803	788
792	804	791
793	806	792
795	807	796
796	808	797
797	809	798
798	810	799
799	811	801
801	812	804
802	813	805
803	814	806
804	815	807
805	816	808
806	818	809
807	821	810
808	823	811
810	824	812
812	825	813
813	826	814
815	827	815
817	828	816

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
818	829	817
819	830	819
820	832	820
821	833	821
823	834	823
824	835	824
825	836	825
827	837	826
828	838	827
829	840	829
831	841	830
832	842	831
833	843	833
834	844	834
836	846	835
837	847	836
838	848	837
839	849	838
840	850	839
842	851	840
843	852	841
846	853	842
847	855	843
848	856	844
849	857	845
850	859	846
853	860	847
855	862	848
856	863	849
857	864	850
859	865	851
860	867	853
861	868	856
862	869	857
863	870	859
864	871	860
865	872	861
866	874	862
867	875	863
868	876	864
869	877	867

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
870	880	869
871	881	871
872	883	873
873	885	874
874	886	875
875	887	876
876	888	877
877	889	879
879	890	880
880	891	881
881	892	883
882	893	885
883	895	886
884	896	887
886	897	888
887	898	889
888	899	890
889	900	891
890	901	892
894	902	893
896	904	894
897	906	895
899	907	896
900	908	897
901	909	899
902	910	902
903	911	903
904	912	906
906	913	907
907	914	908
908	915	909
909	916	910
910	917	911
911	918	913
912	919	914
915	920	915
916	921	916
917	922	917
918	923	918
919	924	919
920	925	920

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
923	927	921
924	928	923
925	929	924
926	930	925
927	931	926
928	934	927
929	935	928
930	936	930
932	937	931
933	939	933
934	940	934
935	941	935
936	942	936
937	943	937
939	944	938
940	945	939
942	946	940
943	947	941
945	948	942
946	950	944
947	951	945
948	952	946
949	954	947
950	956	948
951	958	949
953	959	950
954	960	951
955	961	952
956	962	954
957	965	956
959	966	957
960	967	958
961	968	959
962	969	961
964	970	963
965	971	964
966	972	965
968	973	966
969	974	967
971	976	969
972	977	970

SEQ ID Nos for Examples 7-9 of length 800		
Example 7	Example 8	Example 9
973	978	971
974	979	972
975	980	974
976	981	976
979	982	978
982	983	981
983	984	982
984	985	984
985	987	985
986	988	986
987	989	989
991	991	990
992	992	991
994	993	992
995	994	993
996	995	994
997	996	995
998	997	996
999	999	998
1000	1000	999

Table 4

Claims

1. A method for the detection of colorectal cancer in a human subject based on RNA from a blood sample obtained from said subject, comprising:
- measuring the abundance of at least 250 RNAs in the sample, that are chosen from the RNAs listed in table 1, and
- concluding based on the measured abundance whether the subject has colorectal cancer.

2. The method of claim 1, wherein the abundance of at least 500 RNAs, of at least 800 RNAs, of at least 1000 RNAs that are chosen from the RNAs listed in table 1 is measured.

3. The method of claim 1, wherein the abundance of 250 RNAs as listed in table 2 is measured.

4. The method of claim 2, wherein the abundance of 500 RNAs as listed in table 3 is measured.

5. The method of claim 2, wherein the abundance of 800 RNAs as listed in table 4 is measured.

6. The method of claim 2, wherein the abundance of 1000 RNAs as listed in table 1 is measured.

7. The method of claims 1 to 6, wherein the measuring of RNA abundance is performed using a microarray, a real-time polymerase chain reaction or sequencing.

8. The method of claims 1 to 7, wherein the decision whether the subject has colorectal cancer comprises the step of training a classification algorithm on a training set of cases and controls, and applying it to measured RNA abundance.

9. The method of claims 1 to 8, wherein the classification method is a random forest method, a support vector machine (SVM), or a K-nearest neighbor method (K-NN), such as a 3-nearest neighbor method (3-NN).

10. The method of claims 1 to 9, wherein the RNA is mRNA, cDNA, micro RNA, small nuclear RNA, unspliced RNA, or its fragments.

11. Use of a method of claims 1 to 10 for detection of colorectal cancer in a human subject based on RNA from a blood sample.

12. A microarray, comprising a solid support and a set of oligonucleotide probes, the set containing from 250 to about 100,000 probes, and including at least 250 probes selected from table 1.

13. Use of a microarray for detection of colorectal cancer in a human subject based on RNA from a blood sample, comprising measuring the abundance of at least 250 RNAs listed in table 1, wherein the microarray comprises at least 1 probe for measuring the abundance of each of at least 250 RNAs.

14. A kit for the detection of colorectal cancer in a human subject based on RNA obtained from a blood sample, comprising means for measuring the abundance of at least 250 RNAs that are chosen from the RNAs listed in table 1, preferably comprising means for exclusively measuring the abundance of RNAs that are chosen from table 1.

15. Use of a kit of claim 14 for the detection of colorectal cancer in a human subject based on RNA from a blood sample, comprising means for measuring the abundance of at least 250 RNAs that are chosen from the RNAs listed in table 1, comprising
- measuring the abundance of at least 250 RNAs in a blood sample from a human subject, wherein the at least 250 RNAs are chosen from the RNAs listed in table 1, and
- concluding based on the measured abundance whether the subject has colorectal cancer.

16. A method for preparing an RNA expression profile that is indicative of the presence or absence of colorectal cancer in a subject, comprising:
- isolating RNA from a blood sample obtained from the subject, and
- determining the abundance of from 250 to about 1000 RNAs, including at least 250 RNAs selected from table 1.

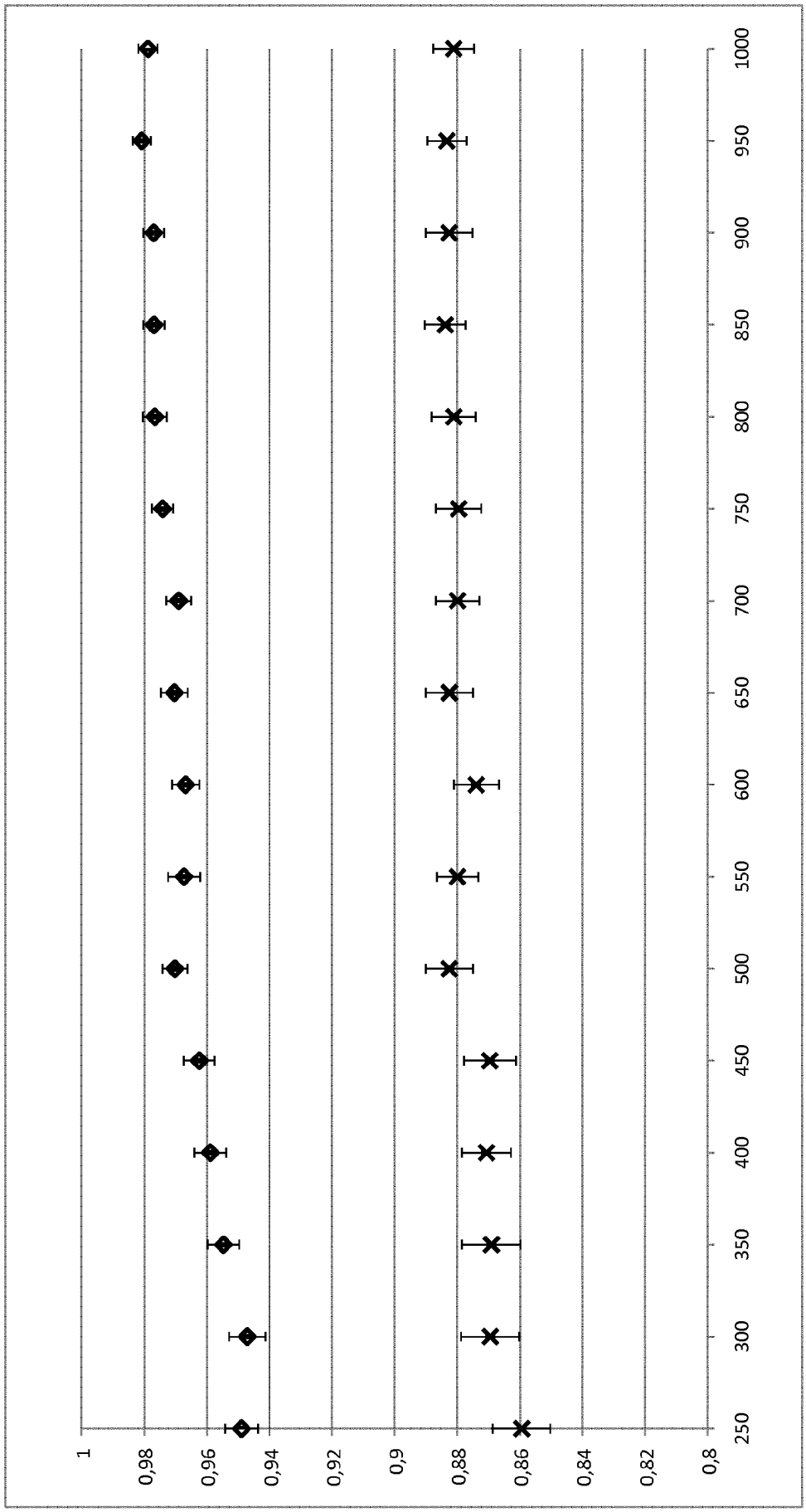

1
0,98
0,96
0,94
0,92
0,9
0,88
0,86
0,84
0,82
0,8
250
300
350
400
450
500
550
600
650
700
750
800
850
900
950
1000

Figure 1

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2012/059339

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2009/038754 A2 (UNIV NEW YORK STATE RES FOUND [US]; CHEN WEN-TIEN [US]) 26 March 2009 (2009-03-26) page 56 - page 102; claims 12,13,20; figures 2,4,6,11,16; examples 1,3,6; tables 1,2,3,5,11 the whole document	1-16

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 August 2012	13/08/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Schmitt, Anja

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	O. GALAMB ET AL: "Diagnostic mRNA Expression Patterns of Inflamed, Benign, and Malignant Colorectal Biopsy Specimen and their Correlation with Peripheral Blood Results", CANCER EPIDEMIOLOGY, BIOMARKERS & PREVENTION, vol. 17, no. 10, 1 October 2008 (2008-10-01), pages 2835-2845, XP055008197, ISSN: 1055-9965, DOI: 10.1158/1055-9965.EPI-08-0231 the whole document	1-16

X	Agilent Technologies: "Agilent SurePrint G3 Human Catalog CGH Microarrays", , 8 January 2009 (2009-01-08), pages 1-8, XP002660065, USA Retrieved from the Internet: URL:http://www.chem.agilent.com/Library/brochures/5990-3368en_lo.pdf [retrieved on 2011-09-26]	12,14
Y	the whole document	13,15,16

X	"Affymetrix Human Genome U133 Plus 2.0 Array", GENE EXPRESSION OMNIBUS, 2003, XP002627319, the whole document	12-16

X	US 2009/208942 A1 (LIEW CHOONG-CHIN [CA] ET AL) 20 August 2009 (2009-08-20)	12-16
Y	examples 1-12; tables A-E	1,2,7-11
A		3-6

X	KENNETH WAYNE MARSHALL ET AL: "A blood-based biomarker panel for stratifying current risk for colorectal cancer", INTERNATIONAL JOURNAL OF CANCER, JOHN WILEY & SONS, INC, UNITED STATES, SWITZERLAND, GERMANY, vol. 126, no. 5, 1 March 2010 (2010-03-01) , pages 1177-1186, XP002647810, ISSN: 0020-7136, DOI: 10.1002/IJC.24910 [retrieved on 2009-09-30]	12-16
Y	the whole document	1,2,7-11
A		3-6

	-/--	

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HAN, M ET AL: "Novel blood-based, five-gene biomarker set for the detection of colorectal cancer", CLINICAL CANCER RESEARCH, THE AMERICAN ASSOCIATION FOR CANCER RESEARCH, US, vol. 14, no. 2, 15 January 2008 (2008-01-15), pages 455-460, XP002642643, ISSN: 1078-0432, DOI: 10.1158/1078-0432.CCR-07-1801 [retrieved on 2008-01-18]	12-16
Y	the whole document -----	1,2,7-11
Y	WO 2009/037572 A2 (DIAGNOPLEX [CH]; RUEGG CURZIO [CH]; MONNIER-BENOIT SYLVAIN [FR]; CIARL) 26 March 2009 (2009-03-26)	1,2,7-16
A	abstract; figures 13,15; example 1,; table 1 page 46 - page 47 the whole document -----	3-6
Y	YIP KOK-THYE ET AL: "A case-controlled validation study of a blood-based seven-gene biomarker panel for colorectal cancer in Malaysia", JOURNAL OF EXPERIMENTAL & CLINICAL CANCER RESEARCH, BIOMED CENTRAL LTD, LONDON UK, vol. 29, no. 1, 16 September 2010 (2010-09-16), page 128, XP021083412, ISSN: 1756-9966, DOI: 10.1186/1756-9966-29-128 the whole document -----	1,2,7-16
Y	MOLNAR B ET AL: "347 Highly Sensitive and Specific Detection of Colorectal Cancer and Adenoma in Peripheral Blood By mRNA Expression Array Technology-Relation to Local Expression Patterns and Identification of Migrating Cell Population", GASTROENTEROLOGY, ELSEVIER, PHILADELPHIA, PA, vol. 136, no. 5, 1 May 2009 (2009-05-01), pages A-55, XP026110797, ISSN: 0016-5085, DOI: 10.1016/S0016-5085(09)60250-X [retrieved on 2009-05-01] abstract -----	1,2,7-16

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2012/059339

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2009038754 A2	26-03-2009	EP 2201134 A2	30-06-2010
		US 2010297634 A1	25-11-2010
		WO 2009038754 A2	26-03-2009
US 2009208942 A1	20-08-2009	AT 518007 T	15-08-2011
		AU 2006304883 A1	26-04-2007
		CA 2626604 A1	26-04-2007
		CN 101370946 A	18-02-2009
		CN 102260742 A	30-11-2011
		EP 1945816 A1	23-07-2008
		EP 2319941 A2	11-05-2011
		JP 2009512440 A	26-03-2009
		US 2007213939 A1	13-09-2007
		US 2009208942 A1	20-08-2009
		WO 2007048074 A1	26-04-2007
WO 2009037572 A2	26-03-2009	AU 2008300338 A1	26-03-2009
		CA 2688558 A1	26-03-2009
		CN 101971033 A	09-02-2011
		EP 2171470 A2	07-04-2010
		JP 2011517765 A	16-06-2011
		US 2010330079 A1	30-12-2010
		WO 2009037572 A2	26-03-2009