Abstract:

A METHOD FOR CLASSIFYING A HUMAN CELL SAMPLE AS CANCEROUS

A new method for discriminating between cancer and non-cancer samples is described. The method comprising detecting the level of at least one microRNA (miR) selected from Mir-Group I consisting of: miR-21, miR-34a and miR-145, and detecting the level of at least one miR selected from Mir-Group II consisting of: miR-126, miR-143 and miR-145 in a test cell sample and, comparing the level of expression of said selected miRs in the test cell sample with the level of expression of the same selected miRs in a previously recorded test set.
A METHOD FOR CLASSIFYING A HUMAN CELL SAMPLE AS CANCEROUS

FIELD OF THE INVENTION:
The present invention relates to the field of cancer-diagnostics. In particular the invention relates to a microRNA expression signature that allows discriminating between cancer and non-cancer samples.

BACKGROUND OF THE INVENTION:

Colorectal cancer is the third most frequent cancer disease and the second most frequent cause of cancer related mortality in the Western World. There are approximately 400,000 new cases annually in these countries.

Solid cancers are normally diagnosed based on a histo-pathological evaluation of tissue samples. It is of paramount importance to the correct treatment of patients that correct discrimination between cancerous and non-cancerous cases are made. However the correct determination, especially when only small samples of tissue are at hand, may be problematic. The problem is even more pronounced when the cell sample is a fine needle biopsy as such a biopsy often destroys the tumor histology. Thus, there is a constant need for alternative or improved methods to discriminate between cancerous and non-cancerous specimens.

MicroRNAs (miRs or mir’s) are an abundant class of short endogenous RNAs that act as posttranscriptional regulators of gene expression by base-pairing with their target mRNAs. miRs are 19-25 nucleotide (nt) RNAs that are processed from longer endogenous hairpin transcripts (Ambros et al. 2003, RNA 9: 277-279). To date more than 6000 miRs have been identified in humans, worms, fruit flies and plants according to the miR registry database release 11.0 in April 2008, hosted by Sanger Institute, UK.

miRs represent robust and stable biomarkers in formalin fixed paraffin embedded FFPE material. There are several publications demonstrating that miRs are stable and regulated in cancer. A recent paper by Schetter et al (JAMA, 2008) describes the use of miR as prognostic biomarkers in colon cancer. Of particular interest is that the authors were able to separate non-cancerous specimens from cancerous specimens by the use of one miR (miR-21). This separation was supported by multivariate analyses, showing that miR-21 expression was independent of disease stage.
The traditional need for control samples result in limitations. In most hospitals it is a well
established practice to take out samples of the epithelium from patients that have symptoms
indicating cancer. In the case of colon cancer this is done by endoscopy. During this
procedure the surgeon visually inspects the colon and typically takes out samples from sites
that look "suspicious" for further pathological analysis. It is not common practice that the
surgeon takes out samples from normal epithelium. Thus to be compatible with the current
clinical practice a new diagnostic test should not rely on normal colon samples from the
patient. Furthermore in the more advanced cancers it may not be an easy task to identify and
obtain normal colon tissue.

In the case of other cancers, breast cancer e.g., surgeons often takes a needle aspiration
biopsy, also known as fine needle aspiration biopsy to investigate superficial tumors. A fine
needle aspiration biopsy is safer and less traumatic than an open surgical biopsy, however
the histological structures of the tumor are prone to be perturbed during the sampling making
the histo-pathological evaluation difficult

In order to offer a valuable supplement to present day histo-pathological examinations a new
diagnostic test must prove to be simple, robust and have a high specificity and predictive
accuracy.

The present invention describes a method which do not require normal test material from the
patient is simple, robust and have high specificity and predictive accuracy.

SUMMARY OF THE INVENTION:

Prior to the present invention, the present inventors believed that in order to obtain a
meaningful and robust miR-signature based on quantitative PCR would require that the miR-
level were normalised to a reference gene, typically a house-keeping gene, or an external
standard and the miR-levels compared with the similar levels in a control tissue-sample
(Bustin, 2004).

However, the present inventors, surprisingly, realized that traditional controls, i.e. a control
tissue-sample and a reference gene, were not necessary to obtain a meaningful and robust
miR-signature of cancer provided that the signature is based on the level of a few, but carefully selected miRs.

Thus, in a first aspect, the invention pertains to a method for classifying

a test cell sample as cancerous, the method comprising:

a) detecting the level of at least one miR selected from miR-Group I consisting of: miR-21, miR-34a and miR-141, and detecting the level of at least one miR selected from miR-Group II consisting of: miR-126, miR-143 and miR-145 in a test cell sample and,

b) calculating the level of the miR(s) selected from miR-Group I relative to the level of the miR(s) selected from miR-Group II in the test cell sample, compare it with a predetermined cut-off value and therefrom determine if the test cell sample is cancerous;

said cut-off value being estimated by calculating the level of the same miR(s) that were selected from miR-Group I in step a) relative to the level of the same miR(s) selected from miR-Group II in step a) in a dataset comprising data on miR levels comprising at least one normal cell sample.

In a second aspect, the invention pertains to a method for assessing recurrence of a specific type of cancer in a human subject who previous has been treated for said specific type of cancer, the method comprising:

a) detecting the level of at least one miR selected from miR-Group I consisting of: miR-21, miR-34a and miR-141, and detecting the level of at least one miR selected from miR-Group II consisting of: miR-126, miR-143 and miR-145 in a test cell sample obtained from a site close to and a tissue similar to the site and tissue-type of said specific type of cancer of said human subject and,

b) calculating the level of the miR(s) selected from miR-Group I relative to the level of the miR(s) selected from miR-Group II in the test cell sample, compare it with a predetermined cut-off value and therefrom determine if the human subject is likely to experience recurrence of said specific type of cancer cancer;

said cut-off value being estimated by calculating the level of the same miR(s) that were selected from miR-Group I in step a) relative to the level of the same miR(s) selected from miR-Group II in step a) in a dataset comprising data on miR levels in cell samples from a collection of humans subjects who previously were treated for cancer and did not
experienced recurrence of cancer as well as subjects who previously were treated for cancer and did experienced recurrence of cancer.

It is contemplated that a miR signature comprised of least one miR selected from miR-Group I consisting of: miR-21, miR-34a and miR-141, and at least one miR selected from miR-Group II consisting of: miR-126, miR-143 and miR-145 in will identify stage II colon cancer patients at high risk of recurrence.

A further important aspect of the invention is the provision of a kit of parts for classifying a test cell sample as cancerous or non-cancerous comprising:

a) the primers that are necessary for the detection of said of selected miRs according to any of the preceding claims.

b) with the exception of polymerase all reagents necessary to perform the detection, and

c) instructions on how to perform the detection.

A further advantage of the herein described methods is that they can be run on formalin fixed paraffin embedded (FFPE) tissue sections.

DEFINITIONS

Prior to a discussion of the detailed embodiments of the invention is provided a definition of specific terms related to the main aspects and embodiments of the invention. All terms are defined in accordance with the skilled person's normal understanding of the terms. In addition, all contents of US patent Nos. 6,268,490 and 6,770,748 "Bicyclonucleoside and ligonucleotide analogue" (Imanishi), International patent publication No. WO/1999/014226 "Bi- and tri-cyclic nucleoside, nucleotide and oligonucleotide analogues" (Exiqon a/s) are incorporated herein by reference.

The terms "cancer" and "cancerous " are used interchangeably herein to refer to cells which has acquired one or more mutations allowing them to reproduce in defiance of the normal restraints to the cells of the body and to invade and colonize territories normally reserved for other cells. Cancer cells exhibit relatively autonomous growth, characterized by a significant loss of control of cell proliferation. Cancer cells include pre-malignant (e.g., benign hyperplasic), malignant, metastatic, and non-metastatic cells. Cancer cells often forms tumors
which frequently comprise cells which are not cancer cells. Tumors of the colon are characterized by a number of well described characteristics well-known to the skilled clinician.

By a "cancerous" (or "cancer") colon test cell sample is referred to a sample which contain cancer cells forming part a tumor or neoplasm or in a number allowing it to be classified as cancer (benign as well as malignant) by a skilled clinician.

The term "non-cancerous" or "non-cancer" which are used interchangeably herein to refer to a test cell sample which not is forming part a tumor or neoplasm or in a number allowing it to be classified as cancer by a skilled clinician.

The terms "miR" and "microRNA" refer to about 18-25 nt non-coding RNAs derived from endogenous genes. They are processed from longer (ca 75 nt) hairpin-like precursors termed pre-miRs. MicroRNAs assemble in complexes termed miRNPs and recognize their targets by antisense complementarity. If the microRNAs match 100% their target, i.e. the complementarity is complete, the target mRNA is cleaved, and the miR acts like a siRNA. If the match is incomplete, i.e. the complementarity is partial, then the translation of the target mRNA is blocked.

As used herein the terms "let-7a", "miR-103", "miR-106a", "miR-125b", "miR-126", "miR-127", "miR-142-3p", "miR-141", "miR-143", "miR-145", "miR-151", "miR-155", "miR-16", "miR-181b", "miR-191", "miR-200a", "miR-200c", "miR-203", "miR-20a", "miR-21", "miR-212", "miR-25", "miR-26a", "miR-27a", "miR-29c", "miR-320", "miR-342", "miR-34a", "miR-492", "miR-498", "miR-708", "miR-9" or "miR-92a" refer to the human miR sequences found in miR registry database release 12.0 or later and hosted by Sanger Institute, UK as well as their animal equivalents.

By the term "miR-signature" is referred to a collection of miRs, the level, expression or amount of which is characteristic of the pathological status of a test sample.

As used herein the term "detecting the level of a miR" refer to the quantification of said miR. One way of quantification is described in the Examples. However, the miR may be quantified in a multitude of other ways e.g. by arrays, northern blots, dot blots, RN'ase protection assays, quantitative mass spectroscopy or various quantitative PCR-based methods such as the TaqMan assay or the UniRT assay used in the examples.
The term "expression", as used herein, refers to the transcription and/or accumulation of RNA-molecules within a call.

In the present context the terms "level of expression of a miR" and "level of a miR" are used synonymously as a measure of the "amount of a specific miR" that is detected in the sample. The "amount of a specific miR" may be expressed in either absolute or relative measures and refers to values obtained by both quantitative, as well as qualitative methods. One particularly preferred measure of the "amount of a specific miR" is the Crossing point (Cp) value obtained by real-time RT-QPCR as described in the examples.

As used herein the term "the level of miR-21, miR-34a, miR-126 and (miR-143 or miR-145)" refer to the level of miR-21, miR-34a, miR-126 and miR-143, or, level of miR-21, miR-34a, miR-126 and miR-145.

The term "level" designates relative as well as absolute amounts of the miRs referred to.

"Q-PCR" quantitative polymerase chain reaction. Q-PCR is highly sensitive method for quantifying the amounts of specific DNA and RNA species in a test sample. A thorough treatise of the Q-PCR technique can be found in Bustin, S.A. (ed.) A-Z of quantitative PCR, IUL Biotechnology Series 5 (2004) 882 pages, which hereby is incorporated herein by reference.

"UniRT" is a novel Q-PCR method. The method is described in Example 1 and in Danish Patent Application PA 2009 00156.

"Sample" refers to a sample of cells, or tissue or fluid isolated from an organism or organisms, including but not limited to solid tissue samples and samples of in vitro cell culture constituents. As a specific cancer e.g. colon cancer is known to metastasise to various organs, for example to liver and lung, the "sample" may be any sample of blood, plasma, serum, spinal fluid, lymph fluid, synovial fluid, urine, tears, blood cells, organs, tumors, that contains cells or lysed cells and also to samples of in vitro cell cultures.

By a "process implying machine learning" is referred to a process which take advantage of computer algorithms that improve automatically through experience, in the art this process of improving the algorithms is often referred to as "training". Machine learning can be used to discover general rules in large data sets, machine learning can e.g. be used to extract clinical
informative data from a dataset comprising miR expression in cancer and non-cancer samples. A general treatise of the concept of machine learning can be found in: Tom Mitchell, Machine Learning, McGraw Hill, 1997, which hereby is incorporated herein by reference.

In the present context the "cut-off value" is a threshold-value above (or below) which, a value calculated to represent the level of a number of miRs indicate that a test cell sample is from a cancer. From the data presented in Figure 3 it is clear that cancerous and non-cancerous samples fall in two separate groups, and accordingly that cut-off values (or scores) that differentiate between cancer and non-cancer samples can be established and used to serve as predetermined cut-off values for new test samples to be evaluated.

By "LNA" or "LNA monomer" (e.g., an LNA nucleoside or LNA nucleotide) or an LNA oligomer (e.g., an oligonucleotide or nucleic acid) is meant a nucleoside or nucleotide analogue that includes at least one LNA monomer. LNA monomers as disclosed in PCT Publication WO 99/14226) are in general particularly desirable modified nucleic acids for incorporation into an oligonucleotide of the invention.

By "oligonucleotide," "oligomer," or "oligo" is meant a successive chain of monomers (e.g., glycosides of heterocyclic bases) connected via internucleoside linkages.

By "LNA modified oligonucleotide" or "LNA substituted oligonucleotide" is meant a oligonucleotide comprising at least one LNA monomer.

In the present context, the terms "nucleobase" covers naturally occurring nucleobases as well as non-naturally occurring nucleobases. Thus, "nucleobase" includes not only the known purine and pyrimidine heterocycles, but also heterocyclic analogues and tautomers thereof. Illustrative examples of nucleobases are adenine, guanine, thymine, cytosine, uracil, purine, xanthine, diaminopurine, and 5-methylcytosine.

In the nucleic acid sequences described herein LNA monomers are depicted in capitals (T, A, G) and DNA monomers in lower case (t, a, c, g). Modified LNA monomers include 5' methyl cytosine shown as "C".

Embodiments of the present invention are described below, by way of examples only.
DETAILED DISCLOSURE OF THE INVENTION:

Central to the present invention is the observation that the level of expression of miR-21 and miR34a are up-regulated in colon as well as other cancer tissues compared to normal tissue, whereas the levels of miR126 and miR143 (or miR145) are down-regulated in colon as well as other cancer tissues compared to normal tissue. I.e. that cancer samples are characterised by an increased level of expression of miR-21 and miR-34a relative to the expression of miR-126 and (miR-143 or miR-145).

A further central element in the invention is the use of a predetermined cut-off value to determine whether a test cell sample is drawn from a cancer or not. Surprisingly, the data presented in example 5 illustrates that a cut-off value established on test cell samples from one specific cancer type and corresponding normal cell test sample(s) obtained from the same type of tissue may be useful also for the assessment of test cell samples obtained from different types of tissues. I.e. a cut-off value established on colon samples appear informative also for the assessment of e.g. bladder and breast cancer samples (see figure 7).

In a preferred embodiment, the cut-off value is determined using a Receiver Operator Characteristic (ROC) curve, according to the method of Sackett et al., Clinical Epidemiology: A Basic Science for Clinical Medicine, Little Brown and Co., 1985, p. 106-7. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (i.e., sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (i.e., the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive.

Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate.

In general, a test cell sample generating a value of the level of the miR(s) selected from miR-Group I relative to the level of the miR(s) selected from miR-Group II that is higher than the cut-off value determined by this method is considered positive for cancer.

One preferred embodiment the invention may be expressed stating that the ratio:
(level of miR21) x (level of miR34a)   
(level of miR126) x (level of miR143) 

or the ratio:

\[
\frac{(level \ of \ miR21) \times (level \ of \ miR34a)}{(level \ of \ miR126) \times (level \ of \ miR145)}
\]

is relatively large in cancer tissue versus in the corresponding normal tissue.

The level of miRs may conveniently be quantified by Quantitative real-time Reverse Transcriptase mediated Polymerase Chain Reaction method, RT-QPCR (Bustin, S.A. (ed.) A-Z of quantitative PCR, IUL Biotechnology Series 5 (2004) 882 pages), see also Example 1 and 2. The read-out from a typical real-time QPCR instrument is often the so-called Cp (crossing point)-value. The Cp-value is related to the level of e.g. a specific miR, by the relation:

\[
(linar) \ expression \ level \ of \ miRx \sim 2^{-C_p(miRx)}
\]

Wherein \(C_p(miRx)\) designates the Cp-readout from real-time QPCR instrument specifically detecting one specific miR called \(miRx\). Example 2 describes such an assay in details.

Accordingly, when the Cp-values are used as quantifiers of miR-levels, the expression:

\[
\frac{(level \ of \ miR21) \times (level \ of \ miR34a)}{(level \ of \ miR126) \times (level \ of \ miR145)}
\]

is equivalent to:

\[+C_p(miR126) + C_p(miR143) - C_p(miR21) - C_p(miR34a)\]

and likewise the expression:

\[
\frac{(level \ of \ miR21) \times (level \ of \ miR34a)}{(level \ of \ miR126) \times (level \ of \ miR145)}
\]

is equivalent to:

\[+C_p(miR126) + C_p(miR145) - C_p(miR21) - C_p(miR34a)\]
Realizing that both colon as well as other cancer samples are characterised by an increased level of expression of miR-21 and miR-34a relative to the expression of miR-126 and (miR-143 or miR-145) make it possible to formulate a wide range of estimators with accompanying cut-off values that can be used to characterise a test sample as cancerous or non-cancerous.

One example of such an estimator is the simple expression:

\[ +C_p(miR126) + C_p(miR143) - C_p(miR21) - C_p(miR34a) \]

As can be seen in Example 2-5, using this simple estimator and a cut-off value of approximately -1.3 to -2 allows discriminating between normal and cancer samples in particularly between normal colon and colon cancer samples. As indicated by the Area Under receiver operating characteristic Curves (AUC)-values of 0.998 and 0.996 the accuracy of the test is almost perfect (for further explanation see example 2). We are not aware of any other miR-based diagnostic test for any cancer that approaches such high accuracy.

Another example of a useable estimator is a linear regression model, such as:

\[ X \times C_p(miR126) + Y \times C_p(miR143) + Z \times C_p(miR21) + W \times C_p(miR34a) \]

where the coefficients X, Y, Z and W are determined by the regression-analysis according to the particular set-up.

Logistic regression is a widely used method for generating linear models to classify data. Performing a logistic regression analysis on the data of Example 2 results in a best fit regression model:

\[ -109.3 \times C_p(miR126) - 98.1378 \times C_p(miR143) + 295.0965 \times C_p(miR21) - 75.747 \times C_p(miR34a) \]

wherein, a value over -817.297 is indicative of a non-cancer colon test sample, and a value less that -817.297 is indicative of a colon cancer test sample.

It is an important feature of the present invention that once the test has been implemented in the actual test laboratory that are to perform the tests the need for additional control colon cell samples diminishes as illustrated by the data in example 5.
To validate the particular experimental set up it is an advantage that the previously obtained dataset comprise data from at least one cancer and as least one non-cancer cell sample. However, a previously obtained dataset comprising data from at least 10 cancer cell samples and at least 10 corresponding non-cancer cell samples is preferred. A dataset of that size furthermore posses the advantage that a cut-off value that characterize non-cancer samples from cancer samples can be calculated by methods well know in the art.

As can be seen from the examples comparing the levels of miR-21 and miR-34a, to the levels of miR-126 and miR-143 are particularly informative (i.e. stable and accurate), indicated by the AUC values of 0.998 and 0.996, and an accuracy of 97.3 % (see table 2), see also example 4.

The data depicted in Figure 4 shows that there is an almost perfect 1:1 correlation between the level of miR-143 and miR-145. This is indeed to be expected as the genes for miR-143 and miR-145 are localized very close to each other in the genome, and are probably expressed from the same transcript.

Accordingly, comparing the levels of miR-21 and miR-34a, to the levels of miR-126 and miR-145 are similar informative to comparing the levels of miR-21 and miR-34a, to the levels of miR-126 and miR-143. Our data indicate that miR-143 and miR-145 can be used interchangeably.

The data in figure 3 also show that the level of miR-21 comparing the levels of miR-143 (or miR-145) is remarkably informative (i.e. both stable and accurate), indicated by the AUC values of 0.991 and 0.992 and an accuracy of 94.6%. Comparing the level of miR-21 to the levels of miR-126 is almost as good having AUC values of 1,000 and 0,970 and an accuracy of 91.9%.

Further comparing the levels of miR-34a and miR-141 to the levels of miR-126 and miR-143 (miR-145) are also both stable and accurate as indicated by the AUC values of 0.977 and 0.968.

Also comparing the levels of miR-21 and miR-141 to the levels of miR-126 and miR-143 (or miR-145) are surprisingly accurate AUC = 0.987, see Example 3.
It is particularly noteworthy that the preferred signature: miR-21 and miR-34a, relative to miR-126 and miR-143 not is improved by the addition yet a miR, in casu miR-141.

In a further embodiment of the invention the method for detecting cancer in a tissue sample according further comprises:

a) obtaining a biological sample from a patient;

b) determining the probability of the test cancer cell sample being cancerous by a process implying machine learning (such as decision tree, principal component analysis, neural networks, or Bayesian networks) analysis wherein the applied program has been trained on a dataset comprised of expression data of the miRs selected from Mir-Group I and the miRs selected from Mir-Group II in cancer as well as in corresponding normal cell samples.

In addition to logistic regression, it is contemplated that various techniques in the art referred to as: principal component analysis, nearest neighbour analysis, neural network and Gaussian mixture model (GMM), Genetic algorithms (1950’s), Decision trees(1960’s) and Support vector machines can be used to produce estimators useable to separate cancer test samples from non-cancerous samples based on the level of miR-21, miR-34a, miR-126 and (miR-143 or miR-145).

As illustrated in Example 2-5 the calculation of even a simple score: \( \text{Cp(miR-126)} + \text{Cp(miR-143)} - \text{Cp(miR-21)} - \text{Cp(miR-34a)} \) allow a cut-off value to be estimated and used to classify a test cell sample as cancer or non-cancer.

Cut-off values from -1.3 to -2 have been obtained, however other values at such as in the range from -0.5 to -4, such as within the range of -1 to -3, e.g. within the range of -1.1 to -2.5, such as within the range of -1.2 to -2.3, including the range of -1.5 to -2.2 and the range of -1.7 to -2.1 may be obtained depending on the precise experimental setup and the type of tissues analysed.

Using the technique of linear regression \( X' \text{Cp(miR126)} + Y' \text{Cp(miR-143)} + Z' \text{Cp(miR-21)} + W' \text{Cp(miR-34a)} \) wherein \( X, Y, Z \) and \( W \) are coefficients determined by linear regression.
Whereas the miRs may be quantified in a multitude of ways a preferred method of quantification is one wherein the the levels of the miRs are determined by Q-PCR. A quantification based on the UniRT-method described in Example 1 is particularly preferred.

Today the standard method for histological tissue preparation in pathology is to prepare Formalin-Fixed Paraffin-Embedded (FFPE) samples. Every year millions of FFPE samples are classified and archived by pathologists. Consequently, it a significant advantage of the herein described methods that they apply to cell samples which are formalin fixed paraffin embedded (FFPE) tissue sections.

A further aspect of the invention is a kit of parts for classifying a test cell sample as cancerous or non-cancerous comprising:

a) the primers that are necessary for the detection of said of selected miRs according to any of the preceding claims.

b) with the exception of polymerase all reagents necessary to perform the detection, and
c) instructions on how to perform the detection.

It is preferable if the reagents kit contains additional means to perform the quantification by the UniRT method of example 1 such a kit would conveniently comprise a selection of UniRT-primers, according to Table 1, that are required to detect the necessary miRs, e.g. the primers required to detect miR-21, miR-34a, miR-126 and miR-143 by the UniRT method.

The invention is further illustrated in the following non-limiting examples and the figures wherein

**LEGENDS**

FIGURE 1: Show the steps involved in microRNA specific qRT-PCR by the UniRT-protocol. STEP 1 is a one-tube-reaction for all microRNAs present in a sample. STEP 2 is a microRNA specific qPCR using forward and reverse primer pairs for a specific microRNA. An oval indicate insertion of LNAs in forward and reverse primers. NV designates a 3'-end VN-degenerate sequence motif comprising the two 3'-end terminal nucleotide residues, wherein V is either an adenine residue, a guanine residue, or a cytosine residue and N is either an adenine residue, a guanine residue, a cytosine residue or a thymine residue.
FIGURE 2: Show the relative expression of 5 selected miRs in Colon cancer and normal (NAT) FFPE tissue. Shown are ΔCp values of all 5 miR relative to miR-103 (reference). miR-141, miR-21 and miR-34a are upregulated in cancer samples compared to normal, whereas miR-126 and miR-143 are downregulated in cancer relative to normal tissue. Boxes indicate the distribution of expression values from the 25th to the 75th percentile for each miR. Vertical lines indicate values from the 5th to the 95th percentile. The value on Y-axis is the miR score.

FIGURE 3: Show scatterplots of the distribution of 3 different the miR based scores obtained. Data from two separate rounds of experiments performed on the same clinical samples are shown. On each panel the "spots" to the left are the scores obtained in normal colon tissue (normal adjacent tissue, NAT), the "spots" to the left are the scores obtained in colon cancer tissue. Panel A, "miR-126 vs miR-21" show the score calculated as Cp(miR-126)-Cp(miR-21), B: "miR-21 vs miR-143" show the score calculated as Cp(miR-143)-Cp(miR-21), and C "miR-21, miR-34, miR-143 and miR-126" refer to the score calculated as Cp(miR-126)+Cp(miR-143)-(miR-21)-Cp(miR-34a). The numbers under the abscissa (x-axis) are the Area Under receiver operating characteristic Curves (AUC) for each set of data. Values on ordinates are the scores referred to above. Each "spot" in the panels refer to the score of one tissue sample. The value on Y-axis is the miR score.

FIGURE 4: Shows the relative level of miR-143 plotted against the relative level of miR-143 is an almost perfect 1:1 correlation. The relative level of miR-143, ΔCp(miR-143), is calculated based on miR-103. Also the relative level of miR-145, ΔCp(miR-145), is calculated based on miR-103.

FIGURE 5: Shows receiver operator curves (ROC) of specificity versus sensitivity of miRs indicated. Two separate rounds of measuring miR’s and evaluating data were performed on the clinical material consisting of an equal numbers of cancer and non-cancer samples. The curves to the left are from round 1, those to the right are from round 2.

FIGURE 6: Show scatterplots of the distribution of the miR based score +Cp(miR126)+Cp(miR143)-Cp(miR21)-Cp(miR34a) described in example 4. Data from 1) reevaluation of the 74 Proteogenex colon cancer and normal samples used in example 2 (discovery set), 2) an independent validation set of 40 Proteogenex colon cancer and normal samples and 3) reevaluation of the RanxO5 samples used in example 1 now with percentage tumor information. The value on Y-axis is the miR based score.
FIGURE 7: miR based score +Cp(miR126)+Cp(miR143)-Cp(miR21)-Cp(miR34a) obtained on cancer specimens of different origin and compared with the score obtained in normal adjacent tissue (NAT) of the same origin. The tissues are from left to right: bladder tumor, bladder NAT, breast tumor, breast NAT, gallbladder tumor, gallbladder NAT, kidney tumor, kidney NAT, lung tumor, lung NAT, pancreas tumor, pancreas NAT, small intestine tumor, small intestine NAT and uterus tumor, uterus NAT. The value on Y-axis is the miR based score.

Example 1: The UniRT method

In this example the UniRT method for amplification and quantification of small non-coding RNA molecules by use of quantitative reverse transcription polymerase chain reaction (qRT-PCR) technology is described in brief.

In brief, see figure 1, the UniRT protocol is a two step protocol. In STEP 1 the miRs present in a sample are firstly poly-A-tailed using a poly(A) polymerase, which adds adenine residues to the 3'-end of RNA molecules. Secondly, an extension primer, which has a poly-T-core nucleotide sequence, a 3'-end VN-degenerate motif and a 5'-end tail, is annealed to the poly-A-tailed miRs through hybridization with the VN-poly-T-sequence of the extension primer. Subsequently, the extension primer is extended in a reverse transcription reaction using the miR as template. The resulting primary extension product is composed of the extension primer and the newly synthesized cDNA, which complementary to the miRs in the sample.

In the next step, STEP 2, a miR-specific PCR is carried out. A miR-specific forward primer is annealed to 3'-end of the newly synthesized cDNA and the upper-strand synthesis is carried out by extending the forward primer in a DNA-polymerization reaction using the primary extension product as template. A miR-specific reverse primer composed of a miR-specific 3'-end sequence, a poly-T-stretch and a 5'-end tail is then hybridized to the upper-strand and the lower-strand is synthesized by extension of the reverse primer.

In both STEP 1 and STEP 2 the LNA's help to ensure a specific and efficient annealing of the primers to their respective targets.
Example 2: Identification of microRNA signatures.

The identification of the miRs that constitute a meaningful and robust miR-signature that differentiate cancerous from non-cancerous Formalin Fixed Paraffin Embedded (FFPE) Colon Cancers tissues were carried out in two separate stages.

Stage I. A purely exploratory stage.

In this stage Formalin Fixed Paraffin Embedded (FFPE) Colon Cancers tissues (133 stage II colon cancer specimens) from the RANX05 set of the The Danish RANX05 Colorectal Cancer Study Group (Werther K, Christensen IJ, Brünner N and Nielsen HJ (2000) Eur J Surg Oncol. 26:657-62.) and FFPE-samples from Colon normal tissues from ProteoGenex Inc., Culver City, CA, USA were evaluated.

The level of miR-498, miR-191, miR-21, miR-103, miR-320, miR-106a, let-7a, miR-181b, miR-20a, miR-26a, miR-151, miR-127, miR-141, miR-143, miR-145, miR-342, miR-34a, miR-200c, miR-125b, miR-126, miR-212, miR-492, miR-142-5p, miR-92a, miR-9, miR-15181b, miR-20a, miR-26a, miR-151, miR-127, miR-141, miR-143, miR-145, miR-342, miR-34a, miR-200c, miR-125b, miR-126, miR-212, miR-492, miR-142-5p, miR-92a, miR-9, miR-151 were quantified using the UniRT qPCR method (see, example 1). In brief,

RNA Purification:
RNA from 1x 10 µm sections was purified with the Qiagen FFPE miRNeasy kit (Qiagen Denmark, cat. No. 217404) according to manufacturer’s protocol. RNA concentration was determined by use of a Nanodrop 1000 instrument (Thermo Scientific, Wilmington, DE, USA) and following the instructions of the manufacturer.

UniRT
In STEP 1 of the UniRT protocol 10 ng of total RNA was used in a 10 µl RT reaction having the composition:

1x Poly(A) polymerase buffer (E Coli Poly A Polymerase, NEB M0276L, New England Biolabs, Ipswich MA).
0.1 mM ATP
0.1 µM RT-primer (L2TA: 5’-ggtactagttttttttttttttvnn (SEQ ID NO. I)), or (v designates cytosine, guanine and adenine residues, n designates cytosine, guanine, adenine and thymine residues).
0.1 mM dATP, 0.1 mM dCTP, 0.1 mM dGTP, 0.1 mM dTTP
0.25 unit of Transcriptor (Roche Diagnostics GmbH, Mannheim, Germany)
2 unit of Poly(A) polymerase (E Coli Poly A Polymerase, NEB M0276L, New England Biolabs, Ipswich MA)

The RT reaction was incubated at 42°C for 1 hour, 95°C for 5 minutes.

Then the RT reaction was diluted 50x in water prior to qPCR analysis - STEP 2.

In STEP 1 of the UniRT protocol 1 µl of the diluted RT reaction was mixed with the PCR primer sets of table 1 (final concentration of each primer is 0.125 µM) for each miR and Fast start SYBR Green Master mix (Roche Diagnostics GmbH, Mannheim, Germany) according to protocol by the providers.

All qPCRs were run in duplicate and in 10 µl volume. The qPCR reactions were run in 384 well plates in a Roche Lightcycler 480 II (Roche Diagnostics GmbH, Mannheim, Germany).

All real-time PCR data were analyzed using the Cp(Crossing point) method calculating the relative expression ratios of the specific target miRs as the crossing point difference (ΔCp) of the specific miRs relative to a reference gene as described by Bustin (2004).

The Cp-values were calculated using the Lightcycler 480 software release 1.5.0, version 1.5.0.39 accompanying the Lightcycler 480 II instrument.
<table>
<thead>
<tr>
<th>miR Base 12</th>
<th>miR sequence 5'→3'</th>
<th>F primers_Primer sequence 5'→3'</th>
<th>R primers_Primer sequence 5'→3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>let-7a</td>
<td>tggagttagttgtaaatgtg (SEQ ID NO. 2)</td>
<td>tGgGTagttagttgg (SEQ ID NO. 34)</td>
<td>cggaggtactagtttttttttttttAactat (SEQ ID NO. 66)</td>
</tr>
<tr>
<td>miR-103</td>
<td>agagcattgtacagggtgtgta (SEQ ID NO. 3)</td>
<td>aGgacGtagttgtacagg (SEQ ID NO. 35)</td>
<td>gtagtttttttttttttttCatagc (SEQ ID NO. 67)</td>
</tr>
<tr>
<td>miR-106a</td>
<td>aaaaagtctactgtgaggttag (SEQ ID NO. 4)</td>
<td>aaaaGttCtactagctgc (SEQ ID NO. 36)</td>
<td>gtagtttttttttttttttCtacctg (SEQ ID NO. 68)</td>
</tr>
<tr>
<td>miR-125b</td>
<td>tctctgagacctcaacttagta (SEQ ID NO. 5)</td>
<td>tccccTgAgacctta (SEQ ID NO. 37)</td>
<td>tgacacggaggtactagttttttttttTcaca (SEQ ID NO. 69)</td>
</tr>
<tr>
<td>miR-126</td>
<td>tccgtcaggtgatataatgcg (SEQ ID NO. 6)</td>
<td>ctGgcTacgctgatgta (SEQ ID NO. 38)</td>
<td>gtagtttttttttttttttTcgcaat (SEQ ID NO. 70)</td>
</tr>
<tr>
<td>miR-127-5p</td>
<td>ctgaagcctcagggcggtctagt (SEQ ID NO. 7)</td>
<td>ctaagcctcagggcg (SEQ ID NO. 39)</td>
<td>ggtacagttttttttttttCaga (SEQ ID NO. 71)</td>
</tr>
<tr>
<td>miR-142-5p</td>
<td>ctaaagatgaagacactact (SEQ ID NO. 8)</td>
<td>cgacAtAaaGtgaagag (SEQ ID NO. 40)</td>
<td>gtagtttttttttttttttTgaagtagt (SEQ ID NO. 72)</td>
</tr>
<tr>
<td>miR-143</td>
<td>tgaggagtgaagcctagttgct (SEQ ID NO. 9)</td>
<td>tGaGTagtgaagcactg (SEQ ID NO. 41)</td>
<td>tgacacggaggtactagtttttttttttGagc (SEQ ID NO. 73)</td>
</tr>
<tr>
<td>miR-145</td>
<td>gtccagttttccccaggatcctc (SEQ ID NO. 10)</td>
<td>gtccagttttccccagga (SEQ ID NO. 42)</td>
<td>gtagtttttttttttttttAagggtat (SEQ ID NO. 74)</td>
</tr>
<tr>
<td>miR-151-5p</td>
<td>tcgagagctcactagttctg (SEQ ID NO. 11)</td>
<td>tcgagagctcactagctg (SEQ ID NO. 43)</td>
<td>gtagtttttttttttttttActa (SEQ ID NO. 75)</td>
</tr>
<tr>
<td>miR-155</td>
<td>ttagtgtaaactctgtaggggt (SEQ ID NO. 12)</td>
<td>ttaTgCtaatctgtagtg (SEQ ID NO. 44)</td>
<td>tgacacggaggtactagtttttttttttttAccctta (SEQ ID NO. 76)</td>
</tr>
<tr>
<td>miR-16</td>
<td>tagcgcagcttaaatcttggc (SEQ ID NO. 13)</td>
<td>ctagCGGagctcactg (SEQ ID NO. 45)</td>
<td>cagcttagtttttttttttttTcgcaat (SEQ ID NO. 77)</td>
</tr>
<tr>
<td>miR-181b</td>
<td>aacacatgtctgtgctgtgggtg (SEQ ID NO. 14)</td>
<td>aaCtCgCtactgctgc (SEQ ID NO. 46)</td>
<td>tgacacggaggtactagtttttttttttttAccca (SEQ ID NO. 78)</td>
</tr>
<tr>
<td>miR-191</td>
<td>caacggaaatccacaaaaacacgtc (SEQ ID NO. 15)</td>
<td>caacgcgaatccacaaaaacacgt (SEQ ID NO. 47)</td>
<td>gtagtttttttttttttttCagc (SEQ ID NO. 79)</td>
</tr>
<tr>
<td>miR-200a</td>
<td>taacactctttgaatcagctgt (SEQ ID NO. 16)</td>
<td>taacCgtacatgtgtaac (SEQ ID NO. 48)</td>
<td>gtagtttttttttttttttTaaCatcg (SEQ ID NO. 80)</td>
</tr>
<tr>
<td>miR-200c</td>
<td>taacactctttgaatcagctgt (SEQ ID NO. 17)</td>
<td>taacCgtacatgtgtaac (SEQ ID NO. 48)</td>
<td>gtagtttttttttttttttCac (SEQ ID NO. 81)</td>
</tr>
<tr>
<td>miR-203</td>
<td>gtaaatgttttagtacagcatcagc (SEQ ID NO. 18)</td>
<td>gttGaattttggtagacca (SEQ ID NO. 49)</td>
<td>tgacacggaggtactagtttttttttttttCtag (SEQ ID NO. 82)</td>
</tr>
<tr>
<td>miR-20a</td>
<td>taaggctagttgctaggttag (SEQ ID NO. 19)</td>
<td>taaggtgcTTataaggtac (SEQ ID NO. 50)</td>
<td>tgacacggaggtactagtttttttttttttTcactc (SEQ ID NO. 83)</td>
</tr>
<tr>
<td>miR-21</td>
<td>tagcttatcagctatgtgta (SEQ ID NO. 20)</td>
<td>tAgCttatatcagctatgtg (SEQ ID NO. 51)</td>
<td>gtagtttttttttttttttCagc (SEQ ID NO. 84)</td>
</tr>
<tr>
<td>miR-212</td>
<td>taacagctcctcagcc (SEQ ID NO. 21)</td>
<td>taacGtcctcagcc (SEQ ID NO. 52)</td>
<td>tttttttttttttttttttTgtcgttag (SEQ ID NO. 85)</td>
</tr>
<tr>
<td>miR-25</td>
<td>cgtgagcactgtctccgggtc (SEQ ID NO. 22)</td>
<td>cgtgacggtccggctt (SEQ ID NO. 53)</td>
<td>gtagtttttttttttttttCagc (SEQ ID NO. 86)</td>
</tr>
<tr>
<td>miR-26a</td>
<td>ttcagactatcagctaggtagct (SEQ ID NO. 23)</td>
<td>ttcGaaCttatcag (SEQ ID NO. 54)</td>
<td>tgacacggaggtactagtttttttttttttAgtcct (SEQ ID NO. 87)</td>
</tr>
</tbody>
</table>
Lower case designates natural occurring nucleotides, upper case designates LNA. Upper case C designates 5-methylcytosine LNA. All miRs are human miR sequences found in miR registry database release 12.0 hosted by Sanger Institute, UK (miR 12.0).

### 5 RESULT

A short list of 6 miRs was identified (miR-21, miR-34a, miR-126, miR-143, miR-141, miR-103).

These were 5 miRs that showed differential expression between colon cancer and normal colon tissue. miR-103 was selected as reference gene as it showed uniform expression levels across the colon cancer and normal tissue and appeared to be useful as a control reference gene for qPCR quantification.

Relative expression ratios of the 5 remaining miRs (miR-21, miR-34a, miR-126, miR-143, miR-141) expressed at the crossing point difference (ΔCp) of the specific miRs relative to miR-103 (-ACp(miRx) = Cp(miRx) - Cp(miR103)) are shown in Figure 2.

### Stage II. Discovery of RUO signature.

A completely independent set of Formalin Fixed Paraffin Embedded (FFPE) Colon Cancers tissues samples was obtained.
The sample set was a staging set purchased from Proteogenex, Ca. USA. (74 samples in total including all stage I-IV and matched NATs (normal adjacent tissue).

The six miRs described above were run on the staging set using the method described above. Based on differential expression in colon cancer and normal colon tissue of the new sample-set miRs were further selected.

In order to find the miRs with a potential to distinguish between normal and cancer tissue, a simplified version of linear regression was used.

While performing these analysis we surprisingly realized that a better differentiation between cancerous and non-cancerous samples were obtained when the level of different miRs were normalized to other miR's instead to a traditional control gene. To our surprise we also observed that it was not necessary to relate the miR-level to the level in normal control tissue. Accordingly, this possibility was explored.

For each pair of miRs the difference in Cp values were calculated for all samples and a t-test between the groups and used to determine the miR pair's having the most promising potential as a biomarker.

Two separate rounds of measuring miR's and evaluating data were performed on the clinical material consisting of an equal numbers of cancer and non-cancer samples (approximately 40 of each, depending on the round).

RESULTS

The results are presented in figure 3.

Receiver Operating Characteristic curves, ROC-curves, From 1st and 2nd round are presented in figure 5.

From the data presented in Figure 3 it is clear that scores of cancerous and non-cancerous samples fall in two separate groups. Using the calculated scores as a test score cut-off values can be determined on the scatter plots shown.
In case of the data presented in Figure 3, C and 5, C the cut-off value were calculated to -1.3 and used as the predetermined cut-off value to determine if the test colon cell sample is cancerous or not; i.e. a value of:

\[ +Cp(miR126) + Cp(miR143) - Cp(miR21) - Cp(miR34a) < -1.3 \] is indicative of a non-cancerous sample, whereas a value of:

\[ +Cp(miR126) + Cp(miR143) - Cp(miR21) - Cp(miR34a) \geq -1.3 \] is indicative of a cancerous sample.

The accuracy of the test depends on how well the test separates the group being tested into those with and without cancer. Accuracy is measured by the area under an Receiver Operating Characteristic curve, ROC-curve. An Area Under receiver operating characteristic Curves (AUC) of 1 represents a perfect test; an area of .5 represents a worthless test. A rough guide for classifying the accuracy of a diagnostic test is the traditional academic point system:
As indicated by the Area Under receiver operating characteristic Curves, the AUC-values, excellent separation could be obtained with miR-21 and miR-126. However almost perfect separation (AUC=0.998 and 0.996) was found when the four miR signature was made.

**In conclusion**

The unexpected result of this experiment is that direct, non-normalized values of the miR expression levels of only few, but carefully selected miRs even in a single FFPE tumor section of 10um without the use of normal tissue or fresh frozen tissue as controls can be used to produce an almost perfect distinction between cancer and non-cancer colon tissue.

**Example 3: Identification of further microRNA signatures.**

Closely following the experimental setup described in Example 1 and 2 a number of values were calculated for two separated runs on rounds of miR quantification performed on the same set of test samples.

Receiver Operating Characteristic curves, ROC-curves. From 1st and 2nd round are presented in figure 5.

Detailed data are presented in table 2 below.
Table 2

<table>
<thead>
<tr>
<th>miR signature</th>
<th>AUC run 2</th>
<th>AUC run 1</th>
<th>TPR (%)</th>
<th>SPC (%)</th>
<th>FPR (%)</th>
<th>ACC (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>FDR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>21, 34a, 126, 143</td>
<td>0.998</td>
<td>0.996</td>
<td>97.3</td>
<td>97.3</td>
<td>2.7</td>
<td>97.3</td>
<td>97.3</td>
<td>97.3</td>
<td>2.7</td>
</tr>
<tr>
<td>21, 143</td>
<td>0.991</td>
<td>0.992</td>
<td>94.6</td>
<td>94.6</td>
<td>5.4</td>
<td>94.6</td>
<td>94.6</td>
<td>94.6</td>
<td>5.4</td>
</tr>
<tr>
<td>21, 126</td>
<td>1.000</td>
<td>0.970</td>
<td>91.9</td>
<td>91.9</td>
<td>8.1</td>
<td>91.9</td>
<td>91.9</td>
<td>91.9</td>
<td>8.1</td>
</tr>
<tr>
<td>34a, 126, 141, 143</td>
<td>0.977</td>
<td>0.968</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21, 34a, 126, 141, 143</td>
<td>0.996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21, 126, 141, 143</td>
<td>0.987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21, 34a, 141, 143</td>
<td>0.985</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TPR=true positive rate, FPR=false positive rate, SPC=specificity, ACC=accuracy, PPV=positive predictive value, NPV=negative predictive value, FDR=False discovery rate

5 Example 4: Robustness of microRNA signature.

To determine the cutoff value with new reagents, a new set of experiments were performed. Following the new experimental setup the level of miR21, miR34a, miR126 and miR143 were determined as described in example 1 and 2, and the score:

\[ +Cp(miR126) + Cp(miR143) - Cp(miR21) - Cp(miR34a) \]

for each sample calculated.

The levels of miR-21, miR-34a, miR-126 and miR-143 was assessed using the following protocol (finalized protocol, now called miRCURY LIMA™ Universal RT microRNA PCR from Exiqon, Denmark:

**miRCURY LNA™ Universal RT microRNA PCR**

The assessment of miRs was basically performed as described in Example 2 but using the commercial Universal cDNA synthesis kit (Exiqon Vedbaek, cat no 203300). STEP 1 of the assay was performed using the Universal cDNA synthesis kit protocol.
10 ng of total RNA was used in a 10 µl RT reaction having the composition:

- 1x Reaction buffer, Exiqon cat no 203300-04.
- 1 µl of RT enzymes, Exiqon cat no 203300-03
- Water
- RNA (10 ng)

The RT reaction was incubated at 42°C for 1 hour, 95°C for 5 minutes.

Then the RT reaction was diluted 50x with water prior to qPCR analysis - STEP 2.

In STEP 1 of the UniRT protocol 1 µl of the diluted RT reaction was mixed with the PCR primer sets of table 3 (final concentration of each primer is 0.3 µM) for each miR and Fast start SYBR Green Master mix (Roche Diagnostics GmbH, Mannheim, Germany) according to protocol by the providers.

All qPCRs were run in 10 µl volume. The qPCR reactions were run in 384 well plates in a Roche Lightcycler 480 II (Roche Diagnostics GmbH, Mannheim, Germany).

All real-time PCR data were analyzed using the Cp(Crossing point, 2nd derivative method) method.

The Cp-values were calculated using the Lightcycler 480 software release 1.5.0, version 1.5.0.39 accompanying the Lightcycler 480 II instrument.
TABLE 3

<table>
<thead>
<tr>
<th>miR 12_0</th>
<th>miR sequence 5'→3'</th>
<th>F primers_Prim sequence 5'→3'</th>
<th>R primers_Prim sequence 5'→3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-21</td>
<td>tagttatcagactgatgtagta (SEQ ID NO. 20)</td>
<td>tAgCttacagactgtagta (SEQ ID NO. 52)</td>
<td>ggtactagtttttttttttttCaac (SEQ ID NO. 84)</td>
</tr>
<tr>
<td>miR-143</td>
<td>tgtatgcagcactgtacgttc (SEQ ID NO. 9)</td>
<td>tGaGtcgagacactgc (SEQ ID NO. 41)</td>
<td>tgcacacgaggtactagttttttttttGagcta (SEQ ID NO. 73)</td>
</tr>
<tr>
<td>miR-34a</td>
<td>tgtcgttttagctgtcgtgt</td>
<td>tGgcAggttgcttg (SEQ ID NO. 28)</td>
<td>tgcacacgaggtactagtttttttttttaCaacca (SEQ ID NO. 92)</td>
</tr>
<tr>
<td>miR-126</td>
<td>tgcGgagctaatagcgtgt (SEQ ID NO. 6)</td>
<td>ctgCgtacccgataga (SEQ ID NO. 38)</td>
<td>gtactagtttttttttttttTgcgtc (SEQ ID NO. 70)</td>
</tr>
</tbody>
</table>

Lower case designates natural occurring nucleotides, upper case designates LNA. Upper case C designates 5-methylcytosine LNA. All miRs are human miR sequences found in miR registry database release 12.0 hosted by Sanger Institute, UK (miR 12_0).

Cutoff determination

The 74 Proteogenex colon normal and colon cancer samples used in example 2 were evaluated and the best cutoff determined to be -2 (see figure 6) using the score determined as in example 1 and 2;

In case of the data from the training set presented in Figure 6, left 2 scatter-plot panels, the cut-off value were calculated to -2. i.e a value of:

\[ +\text{Cp(miR126)} + \text{Cp(miR143)} - \text{Cp(miR21)} - \text{Cp(miR34a)} < -2 \] is indicative of a non-cancereous sample, whereas a value of:

\[ +\text{Cp(miR126)} + \text{Cp(miR143)} - \text{Cp(miR21)} - \text{Cp(miR34a)} \geq -2 \] is indicative of a cancerous sample.

Validation set

An additional independent set of 40 colon normal and colon (purchased from Proteogenex, Ca. USA.) were also analysed and showed perfect separation using the determined cutoff of -2 (see figure 6) on the score determined in example 1 and 2.

\[ +\text{Cp(miR126)} + \text{Cp(miR143)} - \text{Cp(miR21)} - \text{Cp(miR34a)} \]
Reexamination of the RANX05 sample set of samples incorporating percentage tumor information

Furthermore, the RANX05 samples colon samples were reanalyzed with respect to the 4 miRs: miR-21, miR-34a, miR-126 and miR-143.

Sections of the RANX05 FFPE blocks were stained with H&E stain (hematoxylin and eosin stain) which is often used by pathologists to examine histology. The sections were carefully assessed for percent tumor area, that is the area of tumor tissue as a percentage of the area of the tissue. The percent tumor area was binned into 5 categories; 100-75%, 75-50%, 50-25%, 25-10% and 10-0%.

By showing the score:

$$+Cp(miR126) + Cp(miR143) - Cp(miR21) - Cp(miR34a)$$

for each of the percentage tumor groups (see figure 6), it is clear that the score and cutoff is able to separate colon tumor and colon normal with high accuracy with more than 50% tumor, whereas the accuracy is only mildly decreased on >25% tumor. Even most of the <10% tumor samples are classified as tumor.

Result

This experiment show that the score: $$+Cp(miR126) + Cp(miR143) - Cp(miR21) - Cp(miR34a)$$ can be used to identify a sample as cancerous or not, provided that sample was characterized by an area of tumor tissue that is at least 50 percentage.

Surprisingly, the majority of samples were correctly identified as cancerous when the score were calculated and the predetermined cut-off value used to determine their cancer-status, even when the percentage of tumor cells in the sample was as low as 25%.

This demonstrate that score based on the 4 miR-signature is unexpectedly robust.

Example 5: The microRNA signature can discriminate normal from cancer in general.

To investigate if the 4-miR signature developed in example 2 could differentiate cancer from non-cancer in general samples from various solid cancer-types were obtained the level of
miR-21, miR-34a, miR-126 and miR-143 determined and the miR based score
+Cp(miR126)+Cp(miR143)-Cp(miR21)-Cp(miR34a) calculated

RNA was purified from 16 samples were Formalin Fixed Paraffin Embedded (FFPE) blocks from different human tissues using the Roche High Pure miR Isolation Kit. All the FFPE blocks were from NDRI (National Development and Research Institutes, Inc. New York). The specimens were: Bladder and Bladder cancer, Breast and breast cancer, gallbladder cancer (50%) and gallbladder cancer (100%), kidney and kidney cancer, lung and lung cancer, pancreas and pancreas cancer, small intestine and small intestine cancer, stomach and stomach cancer and uterus and uterus cancer (Patient IDS: 28478, 25017, 19560, 20876, 26169, 26545, 23220, 21521).

The levels of miR-21, miR-143, miR-126 and miR-34a was measured using the miRCURY LNA™ Universal RT microRNA PCR system as described in example 4. The score was calculated as previously described, as:
\[ +Cp(miR126) + Cp(miR143) - Cp(miR21) - Cp(miR34a) \]

Result
Surprisingly, the miR-21 score is higher in all cancer specimens tested - compared to the normal tissue of the same origin. Even more surprising, the cutoff determined in the colon experiments (-2), seems to be able to differentiate most of the cancers from their normals.

When taking into consideration that the best cutoff has to be determined for each tissue from a panel of normal samples from that particular tissue these results indicate a very wide use of the 4-miR-score.

REFERENCES:


6. Sackett et al., Clinical Epidemiology: A Basic Science for Clinical Medicine, Little Brown and Co., 1985,

CLAIMS

1. A method for classifying a test cell sample as cancerous, the method comprising:
   a) detecting the level of at least one miR selected from miR-Group I consisting of: miR-21, miR-34a and miR-141, and detecting the level of at least one miR selected from miR-Group II consisting of: miR-126, miR-143 and miR-145 in a test cell sample and,
   b) calculating the level of the miR(s) selected from miR-Group I relative to the level of the miR(s) selected from miR-Group II in the test cell sample, compare it with a predetermined cut-off value and therefrom determine if the test cell sample is cancerous;
   said cut-off value being estimated by calculating the level of the same miR(s) that were selected from miR-Group I in step a) relative to the level of the same miR(s) selected from miR-Group II in step a) in a dataset comprising data on miR levels comprising at least one normal cell sample.

2. The method according to claim 1, wherein the test cell sample is drawn from the group of tissues consisting of tumors of epithelial origin,

3. The method according to claim 1, wherein the test cell sample is drawn from the group of tissues consisting of solid tumors.

4. The method according to claim 1, wherein the test cell sample is drawn from the group of tissues consisting of bladder, breast, gallbladder, kidney, lung, pancreas, colon, small intestine and uterus.

5. The method according to any of claims 1 - 4, wherein the previously obtained dataset comprise data from at least one cancer and as least one non-cancer cell sample.

6. The method according to any of claims 1 - 5, wherein the test cell sample and the normal cell sample are obtained from the same type of tissue.

7. The method according to any of claims 1 - 6, wherein the previously obtained dataset comprise data from at least 10 cancer cell samples and at least 10 non-cancer cell samples.

8. The method according to any of claims 1 - 7, wherein the miRs selected from Mir-Group I are miR-21 and miR-34a, and the miRs selected from Mir-Group II are miR-126 and miR-143 or miR-126 and miR-145.
9. The method according to any of claims 1-7, wherein the miR selected from Mir-Group I is miR-21, and the miR selected from Mir-Group II is miR-143 or miR-145.

10. The method according to claims 1-7, wherein the miRs selected from Mir-Group I are miR-21 and miR-141, and the miRs selected from Mir-Group II are miR-126 and miR-143 or miR-126 and miR-145.

11. The method according to claims 1-7, wherein the miR selected from Mir-Group I is miR-21, and the miR selected from Mir-Group II is miR-126.

12. The method according to claims 1-7, wherein the miRs selected from Mir-Group I are miR-34a and miR-141, and the miRs selected from Mir-Group II are miR-126 and miR-143 or miR-126 and miR-145.

13. The method for detecting cancer in a tissue sample according to claims 1-12 which further comprises:

   a) obtaining a biological sample from a patient;

   b) determining the probability of the test cancer cell sample being cancerous by a process implying machine learning (such as decision tree, principal component analysis, neural networks, or Bayesian networks) analysis wherein the applied program has been trained on a dataset comprised of expression data of the miRs selected from Mir-Group I and the miRs selected from Mir-Group II in cancer as well as in normal cell samples.

14. The method of claims 1-7, wherein the score: Cp(miR-126) + Cp(miR-143) - Cp(miR-21) - Cp(miR-34a) is calculated, a cut-off value estimated and used to classify a test cell sample as cancer or non-cancer.

15. The method of claims 1-7, wherein the score: (miR-126) + Cp(miR-145) - Cp(miR-21) - Cp(miR-34a) is calculated, a cut-off value estimated and used to classify a test cell sample as cancer or non-cancer.
16. The method of claims 1 - 7, which imply the calculation of the score: \( X \cdot \text{Cp(miR126)} + Y \cdot \text{Cp(miR-143)} + Z \cdot \text{Cp(miR-21)} + W \cdot \text{Cp(miR-34a)} \) wherein \( X, Y, Z \) and \( W \) are coefficients determined by linear regression.

17. The method of claims 1 - 7, which imply the calculation of the score: \( X \cdot \text{Cp(miR126)} + Y \cdot \text{Cp(miR-145)} + Z \cdot \text{Cp(miR-21)} + W \cdot \text{Cp(miR-34a)} \) wherein \( X, Y, Z \) and \( W \) are coefficients determined by linear regression.

18. The method of claim 14 or 15, wherein the cut-off value is from -1.3 to -2.

19. The method of any of the preceding claims, wherein the levels of the miRs are determined by Q-PCR.

20. The method of claim 19, wherein the Q-PCR method applied is the UniRT-method according to Example 1.

21. A method for assessing recurrence of a specific type of cancer in a human subject who previously has been treated for said specific type of cancer, the method comprising:

20 a) detecting the level of at least one miR selected from miR-Group I consisting of: miR-21, miR-34a and miR-141, and detecting the level of at least one miR selected from miR-Group II consisting of: miR-126, miR-143 and miR-145 in a test cell sample obtained from a site close to and a tissue similar to the site and tissue-type of said specific type of cancer of said human subject and,

25 b) calculating the level of the miR(s) selected from miR-Group I relative to the level of the miR(s) selected from miR-Group II in the test cell sample, compare it with a predetermined cut-off value and there from determine if the human subject is likely to experience recurrence of said specific type of cancer;

said cut-off value being estimated by calculating the level of the same miR(s) that were selected from miR-Group I in step a) relative to the level of the same miR(s) selected from miR-Group II in step a) in a dataset comprising data on miR levels in cell samples from a collection of humans subjects who previously were treated for cancer and did not experienced recurrence of cancer as well as subjects who previously were treated for cancer and did experienced recurrence of cancer.
22. The method of claim 21, wherein said cut-off value is estimated by calculating the level of the same miR(s) that were selected from miR-Group I in step a) relative to the level of the same miR(s) selected from miR-Group II in step a) in a dataset comprising data on miR levels in cell samples obtained from a site close to and a tissue similar to the site and tissue-type of said specific type of cancer from a collection of humans subjects who previously were treated for said specific type of cancer and did not experienced recurrence of said specific type of cancer as well as subjects who previously were treated for said specific type of cancer and did experienced recurrence of said specific type of cancer.

23. The method of any of the preceding claims, wherein the cell sample is formalin fixed paraffin embedded (FFPE) tissue sections.

24. The method of any of the preceding claims, wherein the cell sample is a fine needle biopsy.

25. A kit of parts for classifying a test cell sample as cancerous or non-cancerous comprising:
   a) the primers that are necessary for the detection of said of selected miRs according to any of the preceding claims.
   b) with the exception of polymerase all reagents necessary to perform the detection, and
   c) instructions on how to perform the detection.

26. The kit of parts according to claim 25, wherein the the primers that are necessary for the detection of said of selected miRs are selected form the group of primers in Table 1.

27. A kit according to claims 25 or 26, wherein the detection of miRs are performed by the the UniRT method of example 1, and the kit comprise a selection of UniRT-primers according to Table 1.
Fig. 2
A

miR-21 vs miR-126

---

B

miR-21 vs miR-143

---

C

miR-21, miR-34a, miR-143 and miR-126

---

Fig. 3
Fig. 4
Fig. 5
Fig. 6


## A. CLASSIFICATION OF SUBJECT MATTER

### INV. C12Q1/68

According to International Patent Classification (IPC) or 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 database consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, With indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>SLABY 0 ET AL: &quot;Altered expression of miR-21, miR-31, miR-143 and miR-145 is related to clinicopathologic features of colorectal cancer&quot; ONCOLOGY, vol. 72, no. 5-6, 1 December 2007 (2007-12-01), pages 397-402, XP009111358 s. KARGER, BASEL, CH ISSN: 0030-2414 DOI: 10.1159/000113489 [retrieved on 2008-01-15] table 1</td>
<td>1-7,13, 19-24</td>
</tr>
<tr>
<td>Y</td>
<td></td>
<td>8-12, 14-18</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C

See patent family annex

** X 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 document but published on or after the international filing date

'L1' 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 refer refing 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

*A' document member of the same patent family

Date of the actual completion of the international search

11 November 2010

Date of mailing of the international search report

08/12/2010

Name and mailing address of the ISA

European Patent Office, P B 5818 Patentlaan 2 NL - 2280 HV RUISWAIK Tej (+31-70) 340-3040, Fax (+31-70) 340-3016

Authorized officer

Santagati, Fabio

Form p c rISA/210 (second sheet) (April 2005)
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>CLAUDIUS FABER ET AL: &quot;The impact of microRNAs on colorectal cancer&quot;</td>
<td>1-7, 13, 21-24</td>
</tr>
<tr>
<td></td>
<td>VIRCHOWS ARCHIV, vol. 454, no. 4, 14 March 2009 (2009-03-14), pages 359-367,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SPRINGER, BERLIN, DE</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>page 360, right-hand col umn, last paragraph - page 361, left-hand column;</td>
<td>8-12, 14-18</td>
</tr>
<tr>
<td></td>
<td>figures 2, 3</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>BANDRES E ET AL: &quot;Identification by Real-time PCR of 13 mature microRNAs</td>
<td>1-7, 13, 21-24</td>
</tr>
<tr>
<td></td>
<td>differentially expressed in colorectal cancer and non-tumoral tissues&quot;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MOLECULAR CANCER, vol. 5, July 2006 (2006-07), XP002607207</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ISSN: 1476-4598</td>
<td></td>
</tr>
<tr>
<td></td>
<td>tables 1, 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13; tables 6A, 6B, 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CHENG ANGIE; JARVI) 28 December 2006 (2006-12-28) figures 11, 12, 14; example</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7 May 2009 (2009-05-07) tables 1, 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>tables 1, 6; sequences 74, 114, 156,202, 204</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>TAZAWA H ET AL: &quot;Tumor-suppressive miR-34a induces senescence-like growth arrest</td>
<td>8-12, 14-18</td>
</tr>
<tr>
<td></td>
<td>through modulation of the E2F pathway in human colon cancer cells&quot;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES (PNAS) ,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>US ISSN: 0027-8424 DOI: 10.1073/PNAS. 0707351104 figure 4</td>
<td></td>
</tr>
</tbody>
</table>
### DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2008026951 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2008171667 A1</td>
</tr>
<tr>
<td>WO 2006137941 A2</td>
<td>28-12-2006</td>
<td>AU 2005333165 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2587189 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1838852 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2008519606 T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2009176723 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2008176766 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2008171715 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2008050744 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CN 101878314 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 2222873 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 20100093539 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2010075304 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2693974 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 2173907 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2010533501 T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2009012433 A1</td>
</tr>
</tbody>
</table>