

(51) International Patent Classification:
C12Q 1/68 (2006.01)(21) International Application Number:
PCT/EP2015/056028(22) International Filing Date:
22 March 2015 (22.03.2015)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicants: UNIVERSITE DE LIEGE [BE/BE]; place du XX août, B-4000 Liège (BE). CENTRE HOSPITALIER UNIVERSITAIRE DE LIEGE [BE/BE]; avenue de l'hôpital 13, B-4000 Liège (BE).

(72) Inventors: BOURS, Vincent; Université de Liège, Génétique humaine, avenue de l'hôpital 13 (Bat B35), B-4000 Liège (BE). FRERES, Pierre; Université de Liège, Oncologie, avenue de l'hôpital 13 (Bat B35), B-4000 Liège (BE). WENRIC, Stéphane; Université de Liège, Oncologie, avenue de l'hôpital 13 (Bat B35), B-4000 Liège (BE). JOSSE, Claire; Centre Hospitalier Universitaire de Liège, Oncologie médicale, avenue de l'hôpital 13 (Bat B35), B-4000 Liège (BE). JERUSALEM, Guy; Centre Hospitalier Universitaire de Liège, Oncologie médicale, avenue de l'hôpital 13 (Bat B35), B-4000 Liège (BE).

(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, BN, 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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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.

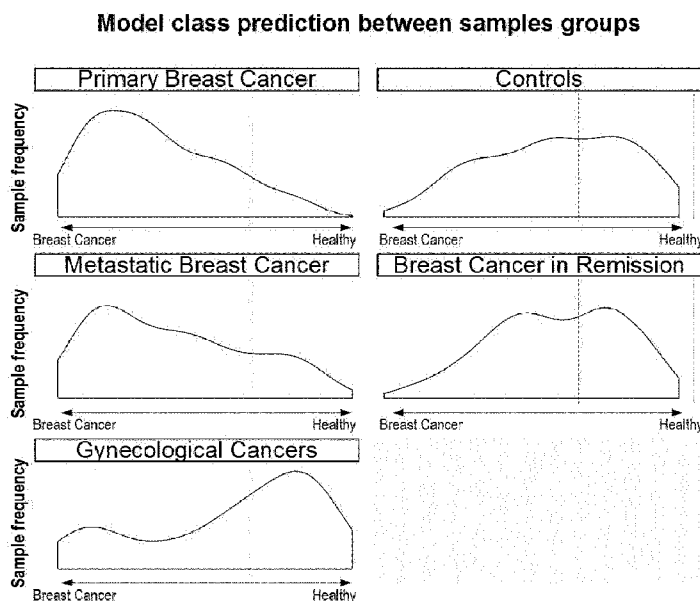
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, 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, KM, 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: CIRCULATING MICRORNAs FOR THE DIAGNOSIS OF BREAST CANCER

Fig. 3B



(57) Abstract: The present invention provides an in vitro method of diagnosing whether a subject has, or is at risk for developing, breast cancer, comprising a) measuring the expression level of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103 in a test biological fluid sample from said subject; b) comparing the expression level obtained in step a) to a reference value of a control biological fluid sample from a healthy subject; wherein an alteration in the expression level of any one or more of said microRNAs is indicative of the subject either having, or being at risk for developing, breast cancer. The invention also provides methods of monitoring the progress of breast cancer and of treatment of breast cancer in a subject. The invention further provides a kit for use in said methods which comprises at least three oligonucleotide probes specific for the detection of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103.

Circulating microRNAs for the diagnosis of breast cancer.

FIELD OF THE INVENTION

The present invention relates to an in vitro method of diagnosing whether a
5 subject has, or is at risk for developing, breast cancer. The invention also
relates to methods of monitoring the progress of breast cancer and of
treatment of breast cancer in a subject. The present invention further relates
to a kit for use in said methods.

BACKGROUND OF THE INVENTION

Breast cancer is the most frequently diagnosed cancer worldwide with 1.38
million cases recorded in 2008 and the leading cause of cancer death in
women.

15 Early diagnosis of breast cancer is currently possible by screening
mammography followed by invasive core needle biopsy in case of suspected
malignancy. Screening mammography is an accessible but unpleasant and
inaccurate test having several disadvantages:

- 20 (i) The risk of false positive. Overdiagnosis rate of screening
mammography is evaluated up to 19%, exposing women to
harmful anti-cancer therapies and affecting their quality of life;
- (ii) The risk of false negative. Mammograms missed a breast cancer
in 17% of cases and up to 30% of cases for dense breasts, such
as in young woman and for women under hormone replacement
25 therapy;
- (iii) The accuracy of screening mammography is sorely affected by the
age. Indeed, young women have dense breast making
interpretation of mammography more difficult.
- 30 (iv) X-ray radiations from mammograms may be one of the factors that
can actually trigger breast cancer in high-risk women, e.g. young
women carrying a mutation in the BRCA genes. Moreover, these
high risk women require early follow-up, beginning at 30 years, an

age where mammography is less effective;

- (v) Mammography performance is operator-dependent.

Several biological markers for breast cancer diagnosis have been identified
5 but CA15.3 is the only validated biomarker. However CA15.3 lacks
sensitivity in case of primary breast tumors and is only useful for the
diagnostic of late stage breast cancer. Hence the accuracy of CA15.3 is
directly influenced by tumor stage.

- 10 MicroRNAs (miRNAs) are noncoding small RNAs that are synthesized inside
the cell, and act as negative regulators of gene expression by binding to the
3'-untranslated region (3'-UTR) within target messenger RNAs (mRNAs).

- Several studies have revealed that microRNAs expression was deregulated
15 in breast cancer tissues, as well as in the circulation of breast cancer
patients. Circulating microRNAs have the advantage of being protected from
degradation by 40 to 100 nm lipoprotein vesicles, called exosomes, and their
presence remains highly stable compared to other RNA molecules.
Circulating microRNAs are therefore easily accessible and can be used as
20 diagnostic markers in multiple cancers.

- In breast cancer, it is well known that circulating microRNAs do not reflect the
abundance of microRNAs in the tumor of origin (Pigati et al, PLoS ONE
2010, 5(10): e13515; Cookson et al, Cell Oncol (Dordr) 2012 35(4): 301-8;
25 Zhu et al. Front Genet. 2013, 5:149-9). Hence, the principle that the
abundance of microRNAs in biological fluids reflects their abundance in the
abnormal cell causing cancer is erroneous. Furthermore blood cells are also
major contributors to circulating microRNA in cancer patients (Pritchard et al,
Cancer Prev Res (Phila) 2012 5(3): 492-7).

- 30 Diagnostic methods measuring circulating microRNAs as biomarkers for
minimally invasive breast cancer detection have been disclosed in the art.

Mixed results in terms of performances were obtained, probably due to variations in study design and analysis, such as choice of proper normalization and careful validation on independent cohort. It is also well established that expression of several microRNAs is deregulated in the same way in benign and malignant breast tumors (Tahiri et al., Carcinogenesis. 2014, 35(1):76-85), which might causes overdiagnosis problems.

As an example, WO2012115885 discloses a big list of circulating microRNAs biomarkers for breast cancer diagnosis, and WO2014202090 discloses circulating microRNAs biomarkers of breast cancer as well as combinations thereof.

Another limitation of diagnostic methods based on circulating microRNAs measurement is normalization of circulating microRNAs expression level, because a universal housekeeping endogenous microRNA has yet to be identified in the circulation. WO2011/110644 discloses the use of one endogenous miRNA, miR-16, for normalization. However it is well-known that miR-16 is predominantly derived from erythrocytes and has been shown to be prone to artificial elevation by hemolysis, as high as 30 folds (Leidner et al, PLoS ONE 2013, 8(3): e57841), which is a problem for accurate normalization of miRNAs expression levels. The use of blood cells-derived housekeeping microRNAs, such as miR-16, for normalization may be moreover problematic in case of anemia, a condition often occurring in breast cancer patients. The use of mir-103 for normalization is also disclosed in the art (Chan et al, 2013, Clin Cancer Res 19:4477–87).

On the other hand, WO2013107459 teaches the use of the mean expression level of the 120 most expressed microRNAs in blood samples for normalization, and Mestdagh et al (Genome Biol 2009, 10:R64) discloses the use of the mean expression value of all commonly expressed microRNAs in a given sample as normalization factor. However, these two approaches are constraining, costly and lead to poor performance to discriminate healthy and

cancer patients (Kodahl et al, Molecular Oncology 2014 Jul;8 (5):874–83).

From the above, it is clear that new methods for breast cancer diagnosis are needed. Ideally they should be minimally invasive, have a higher accuracy to avoid misdiagnosis, and should not be affected by the age of the patient or the tumor stage. MicroRNA-based diagnosis method should moreover further allow proper normalization of microRNA expression levels for the correct interpretation of results.

SUMMARY OF THE INVENTION

The invention provides the following aspects.

1. An in vitro method of diagnosing whether a subject has, or is at risk for developing, breast cancer, comprising the steps:

a) measuring the expression level of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103 in a test biological fluid sample from said subject;

b) comparing the expression level obtained in step a) to a reference value of a control biological fluid sample from a healthy subject;

wherein an alteration in the expression level of any one or more of said microRNAs is indicative of the subject either having, or being at risk for developing, breast cancer.

2. An in vitro method of diagnosing whether a subject has, or is at risk for developing, breast cancer, comprising the steps:

a) measuring the expression level of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103 in a test biological fluid sample from said subject;

b) comparing the expression level obtained in step a) to a reference value of a control biological fluid sample from a subject having a known diagnosis of breast cancer;

wherein a similar expression level of said microRNAs is indicative of the subject either having, or being at risk for developing, breast cancer.

- 5 3. The in vitro method according to aspects 1 or 2, wherein said reference value is the expression level of the same respective microRNAs.
4. The in vitro method according to any one of aspects 1 to 3, wherein breast cancer is primary or metastatic breast cancer.
- 10 5. The in vitro method according to any one of aspects 1, 3 or 4, wherein an up-regulation of the expression level of miR-16 and a down-regulation of the expression level of let-7d and miR-103 as compared to the expression level of said respective microRNAs in a control biological fluid sample from a healthy subject is indicative of the subject either having, or being at risk for developing, primary
15 breast cancer.
6. The in vitro method according to any one of aspects 1, 3 or 4, wherein an up-regulation of the expression level of miR-16, a down-regulation of the expression level of let-7d and no change of the expression level of miR-103 as compared to the expression level of
20 said respective microRNAs in a control biological fluid sample from a healthy subject is indicative of the subject either having, or being at risk for developing, metastatic breast cancer.
7. The in vitro method according to any one of aspects 1 to 6, wherein in step a) the expression level of one or more additional microRNAs is further measured, wherein said additional microRNAs are selected
25 from the group consisting of: miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, miR-20a, let-7i, miR-19a, let-7f-1*, miR-199a-5p, miR-93, miR-451, miR-19b, miR-30b, miR-1, miR-26a, miR-22*, miR-590-5p, miR-101, miR-22, miR-142-5p and miR-32 or combinations
30 thereof.

8. The in vitro method according to aspect 7, wherein said additional microRNAs are the following five microRNAs selected from the group consisting of: miR-107, miR-148a, let-7i, miR-19b and miR-22*.
9. The in vitro method according to aspects 7 or 8, wherein an up-regulation of the expression level of miR-148a and miR-19b, a down-regulation of the expression level of miR-107 and let-7i and no change of the expression level of miR-22* as compared to the expression level of said respective microRNAs in a control biological fluid sample from a healthy subject is indicative of the subject either having, or being at risk for developing, primary breast cancer.
10. The in vitro method according to aspects 7 or 8, wherein an up-regulation of the expression level of miR-148a, let-7i and miR-22* and a down-regulation of the expression level of miR-107 and miR-19b as compared to the expression level of said respective microRNAs in a control biological fluid sample from a healthy subject is indicative of the subject either having, or being at risk for developing, metastatic breast cancer.
11. The in vitro method according to any one of aspects 1 to 10, wherein said expression level is measured by performing a quantitative reverse-transcription real-time polymerase chain reaction (qRT-PCR).
12. The in vitro method according to any one of aspects 1 to 11, wherein said expression level in the test or control biological fluid sample is normalized with a mean expression level of the fifty most expressed microRNAs in said test or control biological fluid sample.
13. The in vitro method according to aspect 12, wherein said fifty most expressed microRNAs, used to normalize the microRNA expression level in both the test and control biological fluid samples, are the same.
14. The in vitro method according to aspects 12 or 13, wherein said most expressed microRNAs are selected from the group comprising miR-484, miR-652, miR-148b, miR-106a, miR-425, let-7g, miR-30b, miR-

- 126, miR-103, miR-146a, miR-93, miR-24, miR-18b, miR-151-5p, miR-423-3p, miR-223, miR-15a, miR-142-3p, miR-26b, miR-15b, miR-191, let-7d*, let-7f, miR-21, miR-126*, miR-125a-5p, miR-181a, miR-23a, miR-222, miR-23b, miR-30c, miR-101, miR-150, miR-320a, miR-26a, miR-145, miR-486-5p, miR-16, miR-199a-5p, miR-107, miR-27a, miR-19b, miR-199a-3p, miR-221, let-7b, miR-92a, miR-27b, miR-451, miR-20a or miR-19a.
- 5
15. The in vitro method according to any one of aspects 1 to 14, wherein the biological fluid sample is selected from the group comprising whole blood, blood serum, blood plasma, blood cells, urine, milk or saliva. Preferably the biological fluid sample is blood plasma.
- 10
16. An in vitro method of monitoring the progress of breast cancer in a subject comprising the steps of:
- 15
- a) measuring the expression level of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103 in two or more test biological fluid samples from said subject, taken at different time intervals;
- 20
- b) determining the progress of breast cancer by comparing the expression level of said microRNAs in the measured biological fluid samples over time.
17. The in vitro method according to aspect 16, wherein in step a) the expression level of the following five microRNAs selected from the group consisting of: miR-107, miR-148a, let-7i, miR-19b and miR-22* is further measured.
- 25
18. A method of treatment of breast cancer in a subject, comprising treating the subject, diagnosed as being in need of breast cancer treatment according to the method of any one of aspects 1 to 15, with a treatment selected from the group consisting of: surgery, radiotherapy, chemotherapy, hormonal therapy or biological treatment.
- 30
19. A method of treatment of breast cancer in a subject, comprising the steps of:

a) determining whether the subject is in need of receiving breast cancer treatment comprising performing the method according to any one of aspects 1 to 15;

5 b) treating the subject diagnosed in step a) as being in need of breast cancer treatment with a treatment selected from the group consisting of: surgery, radiotherapy, chemotherapy, hormonal therapy or biological treatment.

20. A kit for use in an in vitro method of diagnosing whether a subject has, or is at risk for developing, breast cancer, or in a method of
10 treatment of breast cancer in a subject, or in monitoring the progress of breast cancer in a subject, wherein the kit comprises at least or consists of three oligonucleotide probes specific for the detection of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103.

15 21. The kit according to aspect 20, further comprising one or more oligonucleotide probes specific for the detection of one or more additional microRNAs selected from the group consisting of: miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, miR-20a, let-7i, miR-19a, let-7f-1*, miR-199a-5p, miR-93, miR-451, miR-19b, miR-
20 30b, miR-1, miR-26a, miR-22*, miR-590-5p, miR-101, miR-22, miR-142-5p and miR-32 or combinations thereof.

22. The kit according to aspects 20 or 21, wherein said kit consists of eight oligonucleotide probes specific for the detection of the following eight microRNAs selected from the group consisting of: miR-16, let-
25 7d, miR-103, miR-107, miR-148a, let-7i, miR-19b and miR-22*.

23. The kit according to any one of aspects 20 to 22, wherein said oligonucleotide probes are specific for the detection of cDNAs obtained from said microRNAs.

24. The kit according to any one of aspects 20 to 23, adapted for
30 performance of an assay selected from a quantitative reverse-transcription real-time polymerase chain reaction (qRT-PCR), a

locked nucleic acid (LNA) real-time PCR, a northern blotting or a micro-array assay.

25. Use of a kit according to any one of aspects 20 to 24 for diagnosing whether a subject has, or is at risk for developing, breast cancer, preferably by performing the method according to any one of aspects 1 to 15, or for monitoring the progress of breast cancer in a subject, preferably by performing the method according to aspects 16 or 17, or for treating breast cancer in a subject, preferably by performing the method according to aspects 18 or 19.

10

INVENTION DESCRIPTION

We have invented a minimally invasive and accurate in vitro method of diagnosing whether a subject has, or is at risk for developing, breast cancer, or of monitoring the progress of breast cancer based on the measurement of a combination of three circulating microRNAs, namely miR-16, let-7d and miR-103 in a biological fluid sample.

The invention hence provides the following embodiments.

- 20 According to a first aspect, the invention relates to an in vitro method of diagnosing whether a subject has, or is at risk for developing, breast cancer, comprising the steps:

- a) measuring the expression level of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103 in a test biological fluid sample from said subject;
- 25 b) comparing the expression level obtained in step a) to a reference value of a control biological fluid sample from a healthy subject;

wherein an alteration in the expression level of any one or more of said microRNAs is indicative of the subject either having, or being at risk for developing, breast cancer.

30

The diagnostic model based on these three miRNAs was designed in a profiling cohort (41 primary breast cancers, 26 healthy women and 19 benign mammary lesions). The miRNAs-based diagnostic tool was then validated on an independent cohort (108 primary breast cancer, 46 healthy women, 42 benign mammary lesions, 35 breast cancer in complete remission, 31 metastatic breast cancers and 30 gynecologic tumors). Receiver-operating characteristics curve derived from this three miRNAs Random Forests based diagnostic tool exhibited an area under the curve of 0.76 and 0.71, respectively, in the primary breast cancer patients profiling- and validation cohort.

Surprisingly, it was found that miR-16 and miR-103, used in the art as endogenous controls genes, are differentially expressed in plasma from healthy and cancer patients a hence can be used for breast cancer diagnostic.

As used herein, the singular forms “a”, “an”, and “the” include both singular and plural referents unless the context clearly dictates otherwise. By way of example, “a biological fluid sample” refers to one or more than one biological fluid samples.

As used herein, the term “subject” can be any mammal. The term “mammal” refers to all mammals, including, but not limited to, humans, dogs, cats, rabbits, ferrets, guinea pigs, mice, rats, hamsters, gerbils, horses, cows and hedgehogs. Preferred mammals are humans. Most preferred mammals are women. Even most preferred mammals are young women, preferably at high risk for breast cancer.

As used herein, the expression “subject at risk for developing breast cancer” refers to subject exhibiting risk factors for breast cancer. Risk factors for breast cancer include, but are not limited to, age, sex, heredity (such as, but not limited to, BRCA1 and BRCA2 mutations), alcohol, fat intake,

deregulation in some hormone production, environmental factors, etc. The skilled person is aware of the fact that several different risk factors for breast cancer are disclosed in the art.

- 5 The terms microRNA, miRNA, hsa-miR or miR are used herein interchangeably, and refer to 19-25 nucleotides mature non-coding RNAs or precursors thereof, or fragments thereof, derived from endogenous genes of living organisms such as animals. Mature microRNAs are processed from longer hairpin-like precursors termed pre-microRNAs (pre-miRs) having a
- 10 length of approximately 75 nucleotides. MicroRNAs are known from the scientific literature and public databases such as the miRBase database (<http://www.mirbase.org>). A further property of microRNAs is their presence, in a stable, resistant form, in blood (whole blood, blood serum, blood plasma or blood cells) and in various other biological fluids.

15

The microRNAs of interest in the present application are incorporated in table 1 below as an example. The skilled person is well aware that microRNAs may be referred to by different names, or synonyms.

20 **Table 1.** microRNAs of interest in the present application.

SEQ ID NO	microRNAs	Alternative name	Sequence	Accession number in miRBase database
1	let-7b	let-7b-5p	UGAGGUAGUAGGUUGUGUGGUU	MIMAT0000063
2	let-7d*	let-7d-3p	CUAUACGACCUGCUGCCUUUCU	MIMAT0004484
3	let-7d	let-7d-5p	AGAGGUAGUAGGUUGCAUAGUU	MIMAT0000065
4	let-7f-1*	let-7f-1-3p	CUAUACAAUCUAUUGCCUUCU	MIMAT0004486
5	let-7f	let-7f-5p	UGAGGUAGUAGAUUGUAUAGUU	MIMAT0000067
6	let-7g	let-7g-5p	UGAGGUAGUAGUUUGUACAGUU	MIMAT0000414
7	let-7i	let-7i-5p	UGAGGUAGUAGUUUGUGCUGUU	MIMAT0000415
8	miR-1	miR-1-3p	UGGAAUGUAAAGAAGUAUGUAU	MIMAT0000416
9	miR-101	miR-101-	UACAGUACUGUGAUAAACUGAA	MIMAT0000099

		3p		
10	miR-103	miR-103a-3p	AGCAGCAUUGUACAGGGCUAUGA	MIMAT0000101
11	miR-106a	miR-106a-5p	AAAAGUGCUUACAGUGCAGGUAG	MIMAT0000103
12	miR-107	miR-107	AGCAGCAUUGUACAGGGCUAUCA	MIMAT0000104
13	miR-125a-5p	miR-125a-5p	UCCCUGAGACCCUUUAACCUGUGA	MIMAT0000443
14	miR-126	miR-126-3p	UCGUACCGUGAGUAAUAAUGCG	MIMAT0000445
15	miR-126*	miR-126-5p	CAUUAUUACUUUUGGUACGCG	MIMAT0000444
16	miR-142-3p	miR-142-3p	UGUAGUGUUUCCUACUUUAUGGA	MIMAT0000434
17	miR-142-5p	miR-142-5p	CAUAAAGUAGAAAGCACUACU	MIMAT0000433
18	miR-145	miR-145-5p	GUCCAGUUUUCCCAGGAAUCCCU	MIMAT0000437
19	miR-146a	miR-146a-5p	UGAGAACUGAAUCCAUGGGUU	MIMAT0000449
20	miR-148a	miR-148a-3p	UCAGUGCACUACAGAACUUUGU	MIMAT0000243
21	miR-148b	miR-148b-3p	UCAGUGCAUCACAGAACUUUGU	MIMAT0000759
22	miR-150	miR-150-5p	UCUCCCAACCCUUGUACCAGUG	MIMAT0000451
23	miR-151-5p	miR-151a-5p	UCGAGGAGCUCACAGUCUAGU	MIMAT0004697
24	miR-15a	miR-15a-5p	UAGCAGCACAUAAUGGUUUGUG	MIMAT0000068
25	miR-15b	miR-15b-5p	UAGCAGCACAUCAUGGUUUACA	MIMAT0000417
26	miR-16	miR-16-5p	UAGCAGCACGUAAAUUUGGCG	MIMAT0000069
27	miR-181a	miR-181a-5p	AACAUUCAACGCUGUCGGUGAGU	MIMAT0000256
28	miR-18b	miR-18b-5p	UAAGGUGCAUCUAGUGCAGUUAG	MIMAT0001412

29	miR-191	miR-191-5p	CAACGGAAUCCCAAAGCAGCUG	MIMAT0000440
30	miR-199a-3p	miR-199a-3p	ACAGUAGUCUGCACAUUGGUUA	MIMAT0000232
31	miR-199a-5p	miR-199a-5p	CCCAGUGUUCAGACUACCUGUUC	MIMAT0000231
32	miR-19a	miR-19a-3p	UGUGCAAUUCUAUGCAAAACUGA	MIMAT0000073
33	miR-19b	miR-19b-3p	UGUGCAAUCCAUGCAAAACUGA	MIMAT0000074
34	miR-20a	miR-20a-5p	UAAAGUGCUUAUAGUGCAGGUAG	MIMAT0000075
35	miR-21	miR-21-5p	UAGCUUAUCAGACUGAUGUUGA	MIMAT0000076
36	miR-22	miR-22-3p	AAGCUGCCAGUUGAAGAACUGU	MIMAT0000077
37	miR-22*	miR-22-5p	AGUUCUUCAGUGGCAAGCUUUA	MIMAT0004495
38	miR-221	miR-221-3p	AGCUACAUUGUCUGCUGGGUUUC	MIMAT0000278
39	miR-222	miR-222-3p	AGCUACAUCUGGCUACUGGGU	MIMAT0000279
40	miR-223	miR-223-3p	UGUCAGUUUGUCAAAUACCCCA	MIMAT0000280
41	miR-23a	miR-23a-3p	AUCACAUUGCCAGGGAUUUCC	MIMAT0000078
42	miR-23b	miR-23b-3p	AUCACAUUGCCAGGGAUUACC	MIMAT0000418
43	miR-24	miR-24-3p	UGGCUCAGUUCAGCAGGAACAG	MIMAT0000080
44	miR-26a	miR-26a-5p	UUCAAGUAAUCCAGGAUAGGCU	MIMAT0000082
45	miR-26b	miR-26b-5p	UUCAAGUAAUUCAGGAUAGGU	MIMAT0000083
46	miR-27a	miR-27a-3p	UUCACAGUGGCUAAGUUCGCG	MIMAT0000084
47	miR-27b	miR-27b-3p	UUCACAGUGGCUAAGUUCUGC	MIMAT0000419
48	miR-30b	miR-30b-5p	UGUAAACAUCCUACACUCAGCU	MIMAT0000420
49	miR-30c	miR-30c-	UGUAAACAUCCUACACUCUCAGC	MIMAT0000244

		5p		
50	miR-32	miR-32-5p	UAUUGCACAUUACUAAGUUGCA	MIMAT0000090
51	miR-320a	miR-320a	AAAAGCUGGGUUGAGAGGGCGA	MIMAT0000510
52	miR-423-3p	miR-423-3p	AGCUCGGUCUGAGGCCCCUCAGU	MIMAT0001340
53	miR-425	miR-425-5p	AAUGACACGAUCACUCCCGUUGA	MIMAT0003393
54	miR-451	miR-451a	AAACCGUUACCAUACUGAGUU	MIMAT0001631
55	miR-484	miR-484	UCAGGCUCAGUCCCCUCCCGAU	MIMAT0002174
56	miR-486-5p	miR-486-5p	UCCUGUACUGAGCUGCCCCGAG	MIMAT0002177
57	miR-590-5p	miR-590-5p	GAGCUUAUUCAUAAAAGUGCAG	MIMAT0003258
58	miR-652	miR-652-3p	AAUGGCGCCACUAGGGUUGUG	MIMAT0003322
59	miR-92a	miR-92a-3p	UAUUGCACUUGUCCCGGCCUGU	MIMAT0000092
60	miR-93	miR-93-5p	CAAAGUGCUGUUCGUGCAGGUAG	MIMAT0000093

The term “expression level” is any measure for the degree to which the microRNA is produced.

- 5 The “expression level” may be determined by measuring the amount of a microRNA present in the biological fluid sample. The expression level of the microRNA can be determined, for example, with an assay for global gene expression in a biological fluid sample (e.g. using a microarray assay for microRNA expression profiling analysis, or a ready-to-use microRNA qPCR
- 10 plate), or by specific detection assays, for example, but not limited to, quantitative PCR, quantitative reverse-transcription (real-time) PCR (qRT-PCR), locked nucleic acid (LNA) real-time PCR, or northern blotting. All such assays are well known to those skilled in the art. In particular, the measurement of the expression level of a microRNA in a biological fluid
- 15 sample may be carried out with an oligonucleotide probe specific for the detection of said microRNA. Said oligonucleotide probe may bind directly and

specifically to the microRNA, or may specifically reverse transcribe said microRNA. Alternatively, said oligonucleotide probe may bind a cDNA obtained from said microRNA. Said oligonucleotide probe may also amplify a cDNA obtained from said microRNA.

5

The term “biological fluid sample” refers to a whole blood sample, a urine sample, a milk sample or a saliva sample. It can also refer to a sample derived from whole blood, such as, but not limited to, blood serum, blood plasma or blood cells. The term “biological fluid sample” does not refer
10 solely to a liquid sample, but can also refer to a liquid sample that has been dried. Hence, for example, biological fluid sample can refer to dried blood.

Blood samples may be obtained from a subject by various techniques, for example, by using a needle to aspirate a blood sample. Preferably, the
15 sample is obtained in a non-invasive or minimally invasive manner.

The skilled person is well aware as how to isolate blood components, for example blood plasma, blood serum or blood cells.

20 A “control biological fluid sample from a healthy subject” is for example a biological fluid sample from a subject of the same species not affected by breast cancer, and preferably with no reported history of breast cancer.

An alteration in the expression level of a microRNA in a test biological fluid sample generally occurs if a difference of the expression level of the
25 microRNA to the reference value of a control biological fluid sample from a healthy subject is statistically significant. For example, an alteration can refer to a down-regulation of the expression level or an up-regulation of the expression level of a microRNA. If the difference is not considered statistically significant, the expression level is considered unchanged. The

difference may be considered to be statistically significant if its absolute value exceeds a predetermined threshold value. This threshold value can, for example, be the standard deviation of the expression level found in biological fluid samples from a population of healthy subjects as indicated in the table

5 below:

	Expression level in primary breast cancer patients ($2^{-\Delta Ct}$)	Expression level in metastatic breast cancer patients ($2^{-\Delta Ct}$)	Expression level in healthy subjects ($2^{-\Delta Ct}$)
miR-16	20.801 ± 12.688	15.558 ± 7.563	13.677 ± 5.842
let-7d	0.246 ± 0.097	0.267 ± 0.085	0.325 ± 0.122
miR-103	2.753 ± 0.838	3.119 ± 0.646	3.202 ± 0.763
miR-107	0.749 ± 0.226	0.713 ± 0.162	0.863 ± 0.223
miR-148a	0.232 ± 0.118	0.262 ± 0.079	0.173 ± 0.084
let-7i	0.441 ± 0.111	0.56 ± 0.114	0.48 ± 0.119
miR-19b	4.909 ± 1.754	3.523 ± 1.086	4.261 ± 1.299
miR-22*	0.042 ± 0.026	0.047 ± 0.024	0.038 ± 0.018

According to a second aspect, the invention further provides an in vitro method of diagnosing whether a subject has, or is at risk for developing, breast cancer, comprising the steps:

- a) measuring the expression level of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103 in a test biological fluid sample from said subject;
 - b) comparing the expression level obtained in step a) to a reference value of a control biological fluid sample from a subject having a known diagnosis of breast cancer;
- wherein a similar expression level of said microRNAs is indicative of the subject either having, or being at risk for developing, breast cancer.

A “control biological fluid sample from a subject having a known diagnosis of breast cancer” refers to a sample of biological fluid from a subject of the same species that has been diagnosed with breast cancer, such as primary or metastatic breast cancer.

5

As disclosed herein, a “similar” or “unchanged” expression level of a microRNA in a test biological fluid sample generally occurs if a difference of the expression level of the microRNA to the reference value of a control biological fluid sample from a subject having a known diagnosis of breast cancer is not statistically significant.

10

In a preferred embodiment of the methods of the present invention, the reference value is the expression level of the same respective microRNAs of a control biological fluid sample from a healthy subject or of a control biological fluid sample from a subject having a known diagnosis of breast cancer.

15

Alternatively, the reference value may be a previous value for the expression level of a microRNA obtained from a specific subject. This kind of reference value may be used if the method is to be used to monitor the progress of breast cancer, or to monitor the response of a subject to a particular treatment. Preferably, the reference value is the average expression level of the same microRNA found in biological fluid samples from a population of subjects of the same species not affected by breast cancer or with known diagnosis of breast cancer. Preferably, said average expression level found in biological fluid samples from a population of subjects of the same species is determined once and then stored in a database for reference. Preferably, the reference value is measured in biological fluid samples obtained from one or more subjects of the same species and the same sex and age group as the subject, in which breast cancer is to be diagnosed.

20

25

30

In another preferred embodiment of the method of the present invention, breast cancer is primary or metastatic breast cancer.

5 A primary breast cancer refers to a breast tumor growing at the anatomical site where tumor progression began, namely the breast, and proceeded to yield a cancerous mass.

Metastatic breast cancer may be a complication of primary breast cancer. It refers to a stage of breast cancer where the disease has spread to distant
10 sites. It is also referred to as, but not limited to, metastases, advanced breast cancer, secondary tumors, secondary or stage 4 breast cancer. Metastatic breast cancer may occur several years after the primary breast cancer, or sometimes may be diagnosed at the same time as the primary breast cancer, or before the primary breast cancer has been diagnosed.

15

In a further preferred embodiment of the method of the present invention, an up-regulation of the expression level of miR-16 and a down-regulation of the expression level of let-7d and miR-103 as compared to the expression level of said respective microRNAs in a control biological fluid sample from a
20 healthy subject is indicative of the subject either having, or being at risk for developing, primary breast cancer.

In a yet another preferred embodiment of the method of the present invention, an up-regulation of the expression level of miR-16, a down-
25 regulation of the expression level of let-7d and no change of the expression level of miR-103 as compared to the expression level of said respective microRNAs in a control biological fluid sample from a healthy subject is indicative of the subject either having, or being at risk for developing, metastatic breast cancer.

30

In an even another preferred embodiment of the method of the present invention, in step a) the expression level of one or more additional microRNAs is further measured, wherein said additional microRNAs are selected from the group consisting of: miR-181a, miR-107, miR-142-3p, miR-
5 486-5p, miR-148a, miR-20a, let-7i, miR-19a, let-7f-1*, miR-199a-5p, miR-93, miR-451, miR-19b, miR-30b, miR-1, miR-26a, miR-22*, miR-590-5p, miR-101, miR-22, miR-142-5p and miR-32 or combinations thereof.

Preferably said additional microRNAs are the following five microRNAs
10 selected from the group consisting of: miR-107, miR-148a, let-7i, miR-19b and miR-22*.

In a preferred embodiment, an up-regulation of the expression level of miR-148a and miR-19b, a down-regulation of the expression level of miR-107 and
15 let-7i and no change of the expression level of miR-22* as compared to the expression level of said respective microRNAs in a control biological fluid sample from a healthy subject is indicative of the subject either having, or being at risk for developing, primary breast cancer.

20 In another preferred embodiment, an up-regulation of the expression level of miR-148a, let-7i and miR-22* and a down-regulation of the expression level of miR-107 and miR-19b as compared to the expression level of said respective microRNAs in a control biological fluid sample from a healthy subject is indicative of the subject either having, or being at risk for
25 developing, metastatic breast cancer.

Receiver-operating characteristics curve derived from this 8-miRNAs Random Forests based diagnostic signature (namely miR-16, let-7d, miR-103, miR-107, miR-148a, let-7i, miR-19b and miR-22*) exhibited an area
30 under the curve of 0.84 and 0.81, respectively, in the primary breast cancer patients profiling- and validation cohort. The accuracy of the diagnostic tool remained unchanged in the presence of benign mammary lesion(s) and

according to the age and tumor stage. The 8-miRNA signature correctly identified metastatic breast cancer patients. The use of the classification model on cohorts of gynecologic cancers and breast cancers in complete remission yielded prediction distributions similar to that of the control group.

5 Hence this 8 miRNAs-based diagnostic model shows interesting characteristics for clinical application:

- (i) the model correctly identifies females carrying a benign breast lesion as healthy women;
- (ii) the diagnostic test is not affected by age and could be useful for
10 monitoring young women at high risk for breast cancer, in which mammography is less effective but also harmful because of the exposure to radiations;
- (iii) unlike CA15.3, the diagnostic model is as effective regardless of tumor stage, which allows early stage diagnostic;
- 15 (iv) the model can also detect metastatic breast cancers and classify patients in complete remission as controls, offering potential utility for monitoring patients.

20 In alternative embodiments, said additional microRNAs are selected from the group comprising:

the following eight microRNAs selected from the group consisting of: miR-181a, miR-107, miR-142-3p, miR-148a, let-7-1*, miR-199a-5p, miR-590-5p and miR-32;

25 the following twelve microRNAs selected from the group consisting of: miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, miR-20a, let-7i, miR-19a, let-7-1*, miR-199a-5p, miR-22 and miR-32;

the following twelve microRNAs selected from the group consisting of: miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, let-7i, let-7-1*, miR-
30 199a-5p, miR-30b, miR-22, miR-142-5p and miR-32;

the following thirteen microRNAs selected from the group consisting of: miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, miR-20a, let-7i, let-7-1*, miR-451, miR-1, miR-590-5p, miR-22 and miR-142-5p.

or the following four microRNAs selected from the group consisting of: miR-148a, miR-19a, miR-199a-5p and miR-22.

Preferably, the expression level of said additional microRNAs is compared to a reference expression level of said respective microRNAs, using a method selected from a decision tree-based ensemble classification method such as Random Forest, a bagging method, an extra-trees method, a boosting method, a support vector machine supervised learning model, a logistic regression method or another appropriate systems biology or statistical method. The preferred method is the Random Forest method.

In such a way, the expression level from an unknown test biological fluid sample can be compared to and clustered together with the best fitting reference expression level representative for either healthy subjects or subjects having breast cancer such as primary or metastatic breast cancer, without needing to know the exact up-and down regulation of each individual microRNA.

As the skilled person will understand, "Random Forest method" refers to an ensemble tree-based supervised learning method that operates by building a large number of decision trees on bootstrap samples from the training data where the chosen features are randomly selected. As used herein, the term "logistical regression" refers to a type of regression analysis used for predicting the outcome of a categorical criterion variable (a variable that can take on a limited number of categories) based on one or more predictor variables. The probabilities describing the possible outcome of a single trial are modeled as a function of explanatory variables using a logistic function. Logistic regression measures the relationship between a categorical dependent variable and usually a continuous independent variable (or several), by converting the dependent variable to probability scores.

Preferably the expression level is measured by performing a quantitative reverse-transcription real-time polymerase chain reaction (qRT-PCR).

- 5 In a further preferred embodiment of the method of the present invention, the expression level in the test or control biological fluid sample is normalized with a mean expression level of the fifty most expressed microRNAs in said test or control biological fluid sample.
- 10 The term “normalized or “normalization” refers to the comparison of the expression level to one or several control(s) to remove as much variation as possible between biological fluid samples except for that difference that is a consequence of the breast cancer itself. In the context of the present invention, the expression level is normalized with a mean expression level of
- 15 the fifty most expressed microRNAs in biological fluid samples. Many assays are available in the art to determine the most expressed microRNAs in a biological fluid samples, for example, but non-limited to, using a microarray assay for microRNA expression profiling analysis, or ready-to-use microRNA qPCR plates.
- 20 Preferably the fifty most expressed microRNAs, used to normalize the microRNA expression level in both the test and control biological fluid samples, are the same.
- 25 More preferably the most expressed microRNAs are selected from the group comprising miR-484, miR-652, miR-148b, miR-106a, miR-425, let-7g, miR-30b, miR-126, miR-103, miR-146a, miR-93, miR-24, miR-18b, miR-151-5p, miR-423-3p, miR-223, miR-15a, miR-142-3p, miR-26b, miR-15b, miR-191, let-7d*, let-7f, miR-21, miR-126*, miR-125a-5p, miR-181a, miR-23a, miR-
- 30 222, miR-23b, miR-30c, miR-101, miR-150, miR-320a, miR-26a, miR-145, miR-486-5p, miR-16, miR-199a-5p, miR-107, miR-27a, miR-19b, miR-199a-3p, miR-221, let-7b, miR-92a, miR-27b, miR-451, miR-20a or miR-19a.

In another preferred embodiment of the present invention, the biological fluid sample is selected from the group comprising whole blood, blood serum, blood plasma, blood cells, urine, milk or saliva. Preferred biological fluid sample is blood plasma.

Blood plasma has the advantage of being a cell-free sample, so its microRNA content is not artificially elevated by hemolysis of blood cells, the number of which may be furthermore modified in breast cancer patients.

10

In a third aspect, the invention further relates to an in vitro method of monitoring the progress of breast cancer in a subject comprising the steps of:

- a) measuring the expression level of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103 in two or more test biological fluid samples from said subject, taken at different time intervals;
- b) determining the progress of breast cancer by comparing the expression level of said microRNAs in the measured biological fluid samples over time.

20

Preferably, in step a) the expression level of the following five microRNAs selected from the group consisting of: miR-107, miR-148a, let-7i, miR-19b and miR-22* is further measured.

25 In a preferred embodiment, an alteration in the expression level of any one or more of said microRNAs over time is indicative of a favorable breast cancer progression.

In another preferred embodiment, a similar expression level of said microRNAs over time is indicative of an unfavorable breast cancer

30

progression.

In a fourth aspect, the invention relates to a method of treatment of breast cancer in a subject, comprising treating the subject, diagnosed as being in need of breast cancer treatment according to the method of the first or second aspects of the present invention, with a treatment selected from the group consisting of: surgery, radiotherapy, chemotherapy, hormonal therapy or biological treatment.

10 In a fifth aspect, the invention further relates to a method of treatment of breast cancer in a subject, comprising the steps of:

a) determining whether the subject is in need of receiving breast cancer treatment comprising performing the method of the first or second aspects of the present invention;

15 b) treating the subject diagnosed in step a) as being in need of breast cancer treatment with a treatment selected from the group consisting of: surgery, radiotherapy, chemotherapy, hormonal therapy or biological treatment.

20 The terms "treatment" or "treating" encompasses both the therapeutic treatment of an already developed breast cancer, as well as prophylactic or preventative measures, wherein the aim is to prevent or lessen the chances of incidence of breast cancer. Beneficial or desired clinical results may include, without limitation, alleviation of one or more symptoms or one or more biological markers, such as, but not limited to, the microRNAs according to the present invention, diminishment of extent of disease, stabilized (*i.e.*, not worsening) state of disease, delay or slowing of disease progression, amelioration or palliation of the disease state, and the like. "Treatment" or "treating" can also mean prolonging survival as compared to expected survival if not receiving treatment.

25

30

As used herein, the term "hormonal therapy" refers to the therapeutic use of hormones, for example, but not limited to, the administration of hormones to increase diminished levels in the body, also referred as hormone replacement therapy, or therapy involving the use of drugs or surgical procedures or any other suitable composition or procedure to suppress the production of or inhibit the effects of a hormone. Examples of hormonal therapy include, but are not limited to, the use of selective estrogen receptor modulators (such as tamoxifen), aromatase inhibitors or luteinizing hormone releasing hormone analogues (such as goserelin).

As used herein, the term "biological treatment" refers to the use of living organisms, or substances derived from living organisms, or synthetic versions of such substances to treat breast cancer. Examples of biological treatments include, but are not limited to, antibodies, such as anti-HER2 antibodies, antibody-drug conjugates, antibody fragments, cytokines, therapeutic vaccines, cancer-killing viruses, gene therapy or adoptive T-cell transfer.

In a sixth aspect, the invention relates to a kit for use in an in vitro method of diagnosing whether a subject has, or is at risk for developing, breast cancer, or in a method of treatment of breast cancer in a subject, or in monitoring the progress of breast cancer in a subject, wherein the kit comprises at least or consists of three oligonucleotide probes specific for the detection of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103.

The term "kit" as used herein refers to any combination of reagents or apparatus that can be used to perform a method of the invention.

The kit may also comprise instructions for use in an in vitro method of diagnosing whether a subject has, or is at risk for developing, breast cancer, or in a method of treatment of breast cancer in a subject, or in monitoring the progress of breast cancer in a subject.

5

The term “oligonucleotide probe” refers to a short, non-naturally occurring, synthetically obtained sequence of nucleotides that match a specific region of a microRNA or a cDNA obtained from said microRNA, or fragments thereof, and then used as a molecular probe to detect said microRNA or cDNA sequence.

10

In the context of the present invention, an oligonucleotide probe “specific for the detection of a microRNA” for example refers to an oligonucleotide probe that bind directly and specifically to a microRNA or a fragment thereof, or specifically reverse transcribe said microRNA. Alternatively, said oligonucleotide probe may bind specifically to a cDNA obtained from said microRNA. Said oligonucleotide probe may also specifically amplify a cDNA obtained from said microRNA.

15

In a preferred embodiment, the kit further comprises one or more oligonucleotide probes specific for the detection of one or more additional microRNAs selected from the group consisting of: miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, miR-20a, let-7i, miR-19a, let-7f-1*, miR-199a-5p, miR-93, miR-451, miR-19b, miR-30b, miR-1, miR-26a, miR-22*, miR-590-5p, miR-101, miR-22, miR-142-5p and miR-32 or combinations thereof.

20

25

In another preferred embodiment, the kit consists of eight oligonucleotide probes specific for the detection of the following eight microRNAs selected from the group consisting of: miR-16, let-7d, miR-103, miR-107, miR-148a, let-7i, miR-19b and miR-22*.

30

In alternative embodiments, the kit consists of the following oligonucleotide probes selected from the group comprising:

eleven oligonucleotide probes specific for the detection of the following eleven microRNAs selected from the group consisting of: miR-16, let-7d, miR-103, miR-181a, miR-107, miR-142-3p, miR-148a, let-7-1*, miR-199a-5p, miR-590-5p and miR-32;

fifteen oligonucleotide probes specific for the detection of the following fifteen microRNAs selected from the group consisting of: miR-16, let-7d, miR-103, miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, miR-20a, let-7i, miR-19a, let-7-1*, miR-199a-5p, miR-22 and miR-32;

fifteen oligonucleotide probes specific for the detection of the following fifteen microRNAs selected from the group consisting of: miR-16, let-7d, miR-103, miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, let-7i, let-7-1*, miR-199a-5p, miR-30b, miR-22, miR-142-5p and miR-32;

sixteen oligonucleotide probes specific for the detection of the following sixteen microRNAs selected from the group consisting of: miR-16, let-7d, miR-103, miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, miR-20a, let-7i, let-7-1*, miR-451, miR-1, miR-590-5p, miR-22 and miR-142-5p;

or seven oligonucleotide probes specific for the detection the following seven microRNAs selected from the group consisting of: miR-16, let-7d, miR-103, miR-148a, miR-19a, miR-199a-5p and miR-22.

In a further preferred embodiment, said oligonucleotide probes are specific for the detection of cDNAs obtained from said microRNAs.

A "cDNA" or "complement DNA" refers to a complementary DNA produced by reverse transcription of an RNA template, such as a microRNA, using a reverse transcriptase enzyme. Examples of reverse transcriptase are reverse transcriptases derived from moloney murine leukemia virus (M-MuLV) reverse transcriptase, avian myeloblastosis virus (AMV) reverse transcriptase, bovine leukemia virus (BLV) reverse transcriptase, Rous

sarcoma virus (RSV) reverse transcriptase or human immunodeficiency virus (HIV) reverse transcriptase.

5 In a yet another preferred embodiment, the kit is adapted for performance of an assay selected from a quantitative reverse-transcription real-time polymerase chain reaction (qRT-PCR), a locked nucleic acid (LNA) real-time PCR, a northern blotting or a micro-array assay.

All such assays are well known to those skilled in the art.

10

In a seventh aspect, the invention relates to the use of a kit according to the sixth aspect of the invention for diagnosing whether a subject has, or is at risk for developing, breast cancer, preferably by performing the method according to the first or second aspect of the invention, or for monitoring the progress of breast cancer in a subject, preferably by performing the method according to the third aspect of the invention, or for treating breast cancer in a subject, preferably by performing the method according to the fourth or fifth aspect of the invention.

20 BRIEF DESCRIPTION OF THE FIGURES

The present invention is illustrated by the following figures which are to be considered for illustrative purposes only and in no way limit the invention to the embodiments disclosed therein:

25 **Figure 1:** represents the Random forests based methodology. The profiling cohort ($n = 86$) contains 41 primary breast cancer patients, 26 healthy women and 19 patients with benign mammary lesions. The validation cohort ($n = 198$) contains 108 primary breast cancer patients, 46 healthy women and 42 patients with benign mammary lesions. The other cancers cohort ($n =$
30 96) contains 35 breast cancer patients in remission, 31 metastatic breast cancer patients and 30 gynecologic cancer patients.

Figure 2: represents 8 miRNAs present in the best-performing signature for breast cancer diagnosis. Relative expression (mean fold change) of the 8 diagnostic miRNAs in primary breast cancer patients (PBC), breast cancer patients in remission (BCR), metastatic breast cancer (MBC) and gynecologic cancer (Gyn) patients compared to controls is shown.

Figure 3: represents the best performing 8 miRNAs based diagnostic tool performance in the validating cohort. (A) Receiver-operator characteristics (ROC) curve of the diagnostic miRNAs model applied to the validating cohort. The area under the curve (AUC) obtained equals to 0.81. (B) Model outcome distributions for the primary breast cancers, controls, metastatic breast cancers, breast cancers in complete remission, and gynecologic cancers. The x-axis corresponds to the model predictions. The dashed line represents the chosen threshold used to compute finite values for sensitivity and specificity for each cohort. The table below reports the AUC, sensitivity and specificity on the independent cohort, the sensitivity and specificity on the other cancers cohort. The true positive count for the metastatic breast cancers amounts to 25. The true negative count amounts to 14 for breast cancers in remission and gynecologic cancers.

AUC on independent cohort	0.81 ± 0.013
Sensitivity on independent cohort	0.91 ± 0.006
Specificity on independent cohort	0.49 ± 0.027
Specificity on breast cancer in remission	0.40 ± 0.076
Sensitivity on metastatic cancers	0.80 ± 0.045
Specificity on gynecologic cancers	0.41 ± 0.063

20

Figure 4: represents a comparison of the accuracy between the best performing diagnostic 8 miRNA signature, the mammography and CA15.3. (A) While the diagnostic performance of screening mammography (white histograms) is weaker in women under 50 years, the area under the curve (AUC) of the 8 miRNAs based-diagnostic model (black histograms) was as stable for women under and over 50 years. (B) The CA15.3 is not useful for the diagnosis of early breast cancer. While the AUC of CA15.3 (white

25

histograms) increases proportionally to the tumor stage, our 8 miRNAs based-diagnostic model performance (black histograms) was stable regardless of the tumor stage.

- 5 The present invention is further illustrated by the following examples, which do not limit the scope of the invention in any way.

EXAMPLES

MATERIALS AND METHODS

- 10 For general methods relating to the invention, reference is made to well-known textbooks, including, e.g. "Current Protocols in Molecular Biology and Short Protocols in Molecular Biology, 3rd Ed." (F. M. Ausubel et al., eds., 1987 & 1995); "MicroRNA Protocols", series: Methods in Molecular Biology, Vol. 936 Ying, Shao-Yao (Ed.) 2nd ed. 2013, XI; "MicroRNA and Cancer, Methods and Protocols", series: Methods in Molecular Biology, Vol. 676 Wu, Wei (Ed.), incorporated by reference herein.

- For further elaboration of general techniques useful in the practice of this invention, the practitioner can refer to standard textbooks and reviews in microRNA detection and blood fractionation. Included are "MicroRNA Expression Detection Methods", Wang Zhiguo, Yang Baofeng, 2010, XX; "Circulating MicroRNAs Methods and Protocols", series: Methods in Molecular Biology, Vol. 1024 Ochiya, Takahiro (Ed.) 2013; "Blood separation and plasma fractionation", James R. Harris Wiley-Liss ed., 1991, incorporated by reference herein.

25 **Patients, controls and plasma collection**

Ethics approval was obtained from the Institutional Review Board (Ethical Committee of the Faculty of Medicine of the University of Liège) in compliance with the Declaration of Helsinki. Patients with treatment-naïve primary breast cancer ($n = 149$, median age = 55 yr, range = 26 - 87 yr),

breast cancer in remission ($n = 35$, median age = 49 yr, range = 28 - 79 yr), metastatic breast cancer ($n = 31$, median age = 59 yr, range = 35 - 79 yr), gynecologic cancer ($n = 30$, median age = 62 yr, range = 38 - 83 yr) and benign mammary lesion(s) ($n = 61$, median age = 55 yr, range = 40 - 74 yr) were recruited prospectively at CHU of Liège and Clinic Saint-Vincent (Liège, Belgium) from 7/2011 to 9/2014. Gynecologic tumors consist in endometrial ($n = 16$), ovarian ($n = 10$) and cervical ($n = 4$) cancers. Benign mammary lesions consist in fibrocystic breast diseases ($n = 31$) and benign breast calcifications ($n = 30$). Controls were obtained from 72 healthy females of similar age (median age = 51 yr, range = 40 - 71 yr) with normal mammography. Healthy women and women with benign mammary lesions had no history of cancer in the last 5 years. All patients signed a written informed consent form. This work consisted of a prospective study and did not lead to any change in the treatment of enrolled patients; 378 patients were included in this study.

Fifty-four patients with primary locally advanced breast cancer received anthracycline-based neoadjuvant chemotherapy. Only patients achieving a pathological complete response (ypT0N0 following the AJCC-UICC classification) were considered responders.

All primary breast cancer patients and tumor characteristics are summarized in table 2 below.

Table 2: represents patients and primary breast tumors characteristics. NA = not accessed, ER = estrogen receptor, PR = progesterone receptor, HER2 = human epidermal growth factor 2, IDC = invasive ductal carcinoma, ILC = invasive lobular carcinoma.

Characteristics	Primary breast cancer patients (<i>n</i> = 149)
Median age (range) (y)	55 (26-87)
Estrogen receptor [n (%)]	117 (79)
Progesterone receptor [n (%)]	109 (73)
HER2 [n (%)]	30 (20)
Ki67 (median $\pm$ SD) (%)	20 $\pm$ 24
Initial T staging [n (%)]	
NA	1
1	62 (42)
2	58 (39)
3	15 (10)
4	13 (9)
Lymph node involvement [n (%)]	70 (47)
Tumor node metastasis (TNM) stage [n (%)]	
NA	1
1	45 (30)
2	73 (49)
3	31 (21)
Scarff-Bloom-Richardson grading system [n (%)]	
NA	1
1	7 (5)
2	84 (57)
3	57 (38)
Histologic subtype [n (%)]	
NA	0
IDC	125 (84)
ILC	19 (13)
Others	5 (3)
Lymphovascular invasion [n (%)]	27 (21)

Specimen characteristics

Blood samples were withdrawn in 9 ml EDTA tubes. Plasma was prepared within 1 hour by retaining supernatant after double centrifugation at 4°C (10 min at 815 g and 10 min at 2500 g) and was stored at -80°C. Absorbance at 414 nm (ABS₄₁₄) was measured with NanoDrop in all samples to evaluate the degree of hemolysis.

RNA extraction and qRT-PCR of miRNAs

Essential MIQE guidelines were followed during specimen preparation. Circulating miRNAs were purified from 100 μ l of plasma with the miRNeasy mini kit (Qiagen, Germany) according to the manufacturer's instructions. The

standard protocol was adapted on the basis of Kroh's recommendations (Kroh et al, Methods. Elsevier Inc; 2010 Apr 1;50(4): 298-301). MS2 (Roche, Belgium) was added to the samples as a carrier, cel-miR-39 and cel-miR-238 were added as spike-ins. RNA was eluted in 50 µl of RNase-free water at the end of the procedure.

Reverse transcription was performed using the miRCURY LNA™ Universal RT microRNA PCR, polyadenylation and cDNA synthesis kit (Exiqon, Denmark). Quantitative PCR was performed according to the manufacturer's instructions on custom panels of 188 selected miRNAs (Pick-&-Mix microRNA PCR Panels, Exiqon). Controls included the reference genes described in the text, inter-plate calibrators in triplicate (Sp3) and negative controls.

All PCR reactions were performed on an Applied Biosystems 7900HT Real-Time PCR System (Applied Biosystems, USA). miRNAs were considered for analysis with a quantification cycle (Cq) value < 36.

Data analysis

Analyses were conducted using the $2^{-\Delta Cq}$ method ($\Delta Cq = Cq_{\text{sample}} - Cq_{\text{reference gene}}$) for each sample to obtain a normalized expression value.

Data were normalized using the ΔCq method as recommended by Mestdagh *et al.* (Genome Biol. 2009;10(6):R64). The mean Cq of the 50 most highly expressed miRNAs was used for normalization, as it was the most stable reference gene according to the GeNorm software. The analysis with the GeNorm software reveals the best reference gene for accurate normalization in an experimental system by ranking the candidate reference genes according to their expression stability. A gene expression normalization factor (M value) can be calculated based on the geometric mean of a user-defined number of reference genes. The M-value threshold for stability of a gene according to GeNorm is 1.5. As reference gene set, we defined the following combination and compared their M values : all microRNA tested, each taken alone; the global mean of expression values obtained for each microRNA ; the mean of the expression values of the 50 most expressed genes obtained for each microRNA. The last combination has given the best

(meaning smaller) M value (table 3). Once the combination of reference genes has been defined, the mean of expression value of those 50 microRNAs is calculated and this mean is used as reference to apply the ΔCt method.

5 Results of GeNorm analysis are available in table 3 below.

Table 3: Results of GeNorm analysis. GeNorm is an algorithm that determines the most stable reference genes from a set of tested candidate reference genes in a given sample panel. The mean Ct of the 50 most highly
10 expressed miRNAs was used for normalization because it was the most stable reference gene according to the GeNorm software.

Reference genes	M < 1.5
Global mean	0,846
miR-93	0,936
miR-223	0,951
miR-425	0,960
miR-103	0,981
miR-423-3p	0,986
let-7g	0,996
let-7d*	0,999
miR-484	1,003
miR-18b	1,003
miR-142-3p	1,005
miR-126*	1,012
miR-26b	1,017
miR-15b	1,020
miR-222	1,020
miR-30b	1,036
miR-101	1,037
miR-652	1,039
miR-146a	1,041

miR-191	1,072
miR-15a	1,099
miR-125a-5p	1,113
let-7f	1,127
miR-26a	1,128
miR-30c	1,130
miR-21	1,133
miR-148b	1,144
miR-145	1,172
miR-23b	1,180
miR-27a	1,215
miR-16	1,229
miR-486-5p	1,232
miR-181a	1,258
miR-221	1,258
miR-92a	1,265
let-7b	1,272
miR-20a	1,302
miR-19b	1,316
miR-23a	1,365
miR-151-5p	1,370
miR-126	1,383
miR-320a	1,385
miR-199a-5p	1,422
miR-24	1,460
miR-27b	1,539
miR-451	1,626
miR-150	1,754
miR-107	1,892
miR-199a-3p	1,909
miR-19a	2,006
miR-106a	2,090

Futhermore, the delta Cq (miR-23a – miR-451) was determined for each sample to evaluate the risk of hemolysis.

Finally, data homogeneity was tested to detect outliers. Patients presenting extreme values (mean $\pm$ 3 sigma) were discarded. This operation leads to the elimination of one patient.

Statistical analyses were performed with *R version 3.0.1* (R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>). To compare miRNA expression levels, two-sided Mann-Whitney U tests and Kruskal-Wallis one-way tests were used. To correlate the expression of the 8 diagnostic miRNAs and the main histo-prognostic markers in primary breast cancer patients, Spearman tests were used for continuous variables. Statistical significance was established as $P < .05$ (*), $P < .01$ (**), $P < .001$ (***) or $P < .0001$ (****). All represented values were adjusted for multiple testing using the Benjamini-Hochberg procedure.

Study design

The REMARK guidelines were followed.

The analysis and computational methods rely on several steps which make use of the Random forests algorithm. Random forests are an ensemble tree-based supervised learning method that operates by building a large number of decision trees on bootstrap samples from the training data where the chosen features are randomly selected (hence the name, Random forests).

For all steps of the method, an *R* implementation of Breiman's original Random forests algorithm has been used, provided in the *R* package *randomForest* (Liaw et al, R news. 2002;2(3):18-22.). A methodology somewhat similar to the algorithmic solution proposed by Geurts *et al.* has been used (Geurts et al, Bioinformatics. 2005 Jul 14;21(14):3138-45.) and is represented in Fig. 1. The different steps are described in detail below.

1. Model building with all miRNAs

A Random forests model was built on the profiling cohort (86 samples: 41 individuals with primary breast cancers, 26 healthy women and 19 individuals

with benign mammary lesions) with the normalized expression values all 188 miRNAs as features.

A conservative value of 3000 has been selected for the number of trees used in the model, as the rankings of the miRNAs in terms of both importance metrics, the Mean Decrease in Accuracy (MDA) and the Mean Decrease in Gini (MDG), were stable at such value.

An incremental approach was tried to select an appropriate value for m_{try} (the number of variables assessed at each splitting step of each internal decision tree used by the Random forests method), but no significant performance change was seen for values of m_{try} going from 1 to n , where n is the total number of miRNAs. Performance was measured here by either the out-of-bag error rate, or the area under the curve (AUC) obtained by plotting a receiver-operating characteristics (ROC) curve when doing a 10-fold cross-validation on the model.

A ranking for all 188 miRNAs based on the model importance metrics, MDA and MDG, was obtained.

2. miRNA signature identification

A subset of m miRNAs was determined based on both importance metrics (the m top ranked miRNAs based on the mean of their rankings in terms of MDA and MDG were selected).

All c possible combinations from 2 to m miRNAs out of this subset were generated, where $c = 2^{m-1}$ combinations.

For each combination of miRNAs, a Random forest model was built and cross-validated on the profiling cohort. The same number of trees as in step 1 was chosen. A default value of $m_{try} = \sqrt{\text{number of miRNAs in the combination}}$ was chosen. Based on the performance of each model measured through the AUC obtained through ten-fold cross-validation, the best performing miRNA combination was selected.

3. Building of the final model

A Random forests model was built on the profiling cohort with the best performing miRNA subset. This classification tool constituted the final diagnostic model. The number of trees chosen to build each model was determined as in step 1 and a default value of $m_{try} = \sqrt{\text{number of miRNAs in the combination}}$ was chosen.

A threshold related to the numerical prediction of the algorithm was chosen to be able to compute finite values for sensitivity and specificity. The Random forests algorithm's output is a numerical value representing the probability for a sample to be part of a specific class (case or control). To obtain finite values, a specific threshold had to be picked to separate the 2 classes.

4. Model validation

The classification tool was then validated on an independent cohort with a similar ratio of primary breast cancers, benign mammary lesions, and healthy women as the profiling cohort and a total number of samples 2.3 times greater (198 samples: 108 individuals with primary breast cancers, 46 healthy women and 42 individuals with benign mammary lesions).

An AUC was obtained through this validation. Finite values for sensitivity and specificity were computed thanks to the threshold defined on the profiling cohort.

The classification tool was also tested on a separate cohort consisting of 35 individuals in breast cancer complete remission, 31 metastatic breast cancer patients and 30 gynecologic cancers.

Establishment of the miRNA signature target genes prediction and validated genes lists

For each of the miRNAs composing the signatures, genes predicted by at least 4 algorithms among 5 used (miRanda, miRDB, miRWalk, TargetScan and RNA22) were retained. As miR-22* was not present in TargetScan, DIANAmT was used. The miRWalk 2.0 database was employed for this purpose.

For each of the miRNA composing the signatures, their experimentally validated target genes curated in miRTarBase (release 4.5) and DIANA-TarBase v7.0 were retained.

RESULTS

5 Pilot study

A pilot study was first conducted and consisted of measuring the expression of 742 plasma miRNAs in 18 primary breast cancer patients. On the basis of their expression levels (mean Cq value < 36) in the pilot experiment, 188 miRNAs were chosen (not shown).

10 Evaluation of hemolysis

We first evaluated the quality of our samples collection and preparation. Hemolysis conducts to contamination of plasma with RNA from red blood cells. Absorbance at 414 nm (ABS_{414}), the maximum absorbance of hemoglobin, correlated with the degree of hemolysis. ABS_{414} was measured with NanoDrop for all samples. The median ABS_{414} level was 0.19 ± 0.1 , with a hemolysis cut-off fixed at 0.2. Furthermore, the level of a miRNA highly expressed in red blood cells (miR-451) was compared with the level of a miRNA unaffected by hemolysis (miR-23a), considering that a ΔCq (miR-23a - miR-451) of more than 5 is an indicator of possible erythrocyte miRNA contamination. The median ΔCq (miR-23a - miR-451) was 2.6 ± 1.5 in our cohort (primary breast cancers group = 3 ± 1.5 , healthy women group = 2.3 ± 1.2 , benign mammary lesion(s) group = 1.9 ± 1.1 , breast cancer patients in remission = 2.5 ± 1.5 , metastatic breast cancers group = 2.8 ± 1.2 , gynecologic cancers group = 2.3 ± 1.8).

25 Design of the control group

It is well established that, compared to normal breast tissue, the miRNA expression is modified in benign breast lesions. To determine whether the circulating miRNA profile was affected by benign breast diseases, the expression of the circulating miRNAs was compared between women with a normal mammogram and women carrying benign mammary lesion(s). We did not find any miRNA significantly deregulated after multiple comparisons testing correction. As benign breast lesions did not interfere with our

circulating miRNAs profile, these individuals were included in the control group with healthy women.

miRNA deregulation in breast cancer patients

When comparing miRNA profile of newly diagnosed primary breast cancers to controls, 112 miRNAs were founded significantly deregulated, 107 after adjusted p-value for multiple testing (table 4). miR-16 and let-7d were respectively the most up- and down-regulated miRNA. A global upregulation of miRNAs was observed in primary breast cancer patients compared to controls (1.7 fold change).

In a second analysis, miRNA profiles from plasma of patients with metastatic breast cancer were compared to those of controls. 84 miRNAs were found significantly deregulated, 53 after adjusted p-value for multiple testing (table 4). The most significantly upregulated miRNA was miR-148a and the most significantly downregulated miRNA was miR-15b. As seen in primary breast cancer samples, a global upregulation of miRNAs was observed when compared to healthy subjects (1.1 fold-change).

Statistical analyzes were also performed to compare both primary and metastatic breast cancer patients to controls, using a Kruskal-Wallis test (table 4). 56 miRNAs were significantly modified in the same way among primary and metastatic breast cancer patients. miR-16 and let-7d were the most co-deregulated miRNAs.

Table 4. Results of statistical analyses. To compare miRNA expression levels, two-sided Mann-Whitney U tests and Kruskal-Wallis one-way tests were used. All represented values were corrected for multiple testing using the Benjamini-Hochberg procedure PBC = primary breast cancer, Ctrl = healthy women + benign mammary lesion, MBC = metastatic breast cancer.

	PBC vs. Ctrl			MBC vs. Ctrl			Kruskal-Wallis (PBC vs. MBC vs. Ctrl)	
	pval	cor. pval	Fold change	pval	cor. pval	Fold change	pval	cor. pval
let-7a	4.45E-04	1.27E-03	0.85	1.68E-04	1.43E-03	1.27	3.16E-02	4.44E-02
let-7a*	5.89E-06	3.36E-05	2.26	9.05E-02	1.63E-01	1.25	1.97E-01	2.22E-01
let-7b	6.43E-01	7.03E-01	1.04	4.03E-01	4.77E-01	0.96	3.00E-08	3.90E-07
let-7c	1.69E-04	5.79E-04	0.81	7.88E-01	8.14E-01	0.98	3.63E-02	4.98E-02
let-7d	4.01E-13	3.77E-11	0.72	2.50E-04	2.04E-03	0.78	1.15E-03	2.66E-03
let-7d*	1.75E-02	3.20E-02	1.09	1.10E-01	1.81E-01	1.11	7.37E-03	1.32E-02
let-7f	2.54E-01	3.41E-01	0.98	3.42E-03	1.89E-02	1.25	2.08E-02	3.08E-02
let-7f-1*	1.11E-07	1.89E-06	2.09	3.69E-01	4.47E-01	1.18	3.65E-02	4.98E-02
let-7g	3.39E-04	1.05E-03	0.92	1.47E-04	1.38E-03	0.85	6.75E-06	4.53E-05
let-7i	4.29E-05	1.75E-04	0.88	1.07E-02	4.18E-02	1.12	1.23E-02	2.01E-02
let-7i*	7.77E-02	1.23E-01	1.71	2.53E-01	3.40E-01	1.12	1.40E-02	2.22E-02
miR-1	2.37E-07	3.18E-06	1.93	1.26E-01	2.00E-01	1.44	7.76E-05	2.56E-04
miR-101	1.20E-06	1.28E-05	1.30	1.40E-02	4.97E-02	0.87	1.77E-05	9.79E-05
miR-103	2.29E-06	2.15E-05	0.84	5.82E-01	6.51E-01	0.95	2.41E-02	3.51E-02
miR-103-2*	5.24E-06	3.20E-05	2.26	8.74E-02	1.62E-01	1.25	3.58E-04	9.48E-04
miR-106a	1.07E-03	2.69E-03	1.08	2.41E-02	7.08E-02	1.08	1.15E-04	3.61E-04
miR-106b	2.01E-01	2.79E-01	1.09	4.36E-07	1.37E-05	1.48	3.95E-02	5.35E-02
miR-106b*	3.69E-02	6.30E-02	1.45	2.09E-02	6.66E-02	1.27	1.95E-02	2.94E-02
miR-107	3.27E-08	6.15E-07	0.83	5.30E-06	8.30E-05	0.79	1.51E-06	1.14E-05
miR-10a	1.22E-06	1.28E-05	2.23	1.70E-01	2.55E-01	1.21	6.07E-03	1.12E-02
miR-10b	8.94E-04	2.36E-03	1.92	9.77E-04	6.81E-03	1.73	4.01E-02	5.38E-02
miR-122	1.34E-02	2.48E-02	1.45	1.11E-01	1.82E-01	2.72	1.80E-01	2.05E-01
miR-125a-5p	4.01E-03	8.65E-03	0.84	1.83E-03	1.15E-02	0.70	8.13E-02	1.01E-01
miR-125b	4.41E-01	5.31E-01	1.17	7.59E-01	7.89E-01	1.08	1.15E-04	3.61E-04
miR-126	6.39E-01	7.02E-01	0.98	2.16E-06	5.08E-05	0.83	7.47E-07	5.85E-06
miR-126*	3.45E-03	7.77E-03	1.10	3.69E-02	9.36E-02	0.90	1.21E-07	1.19E-06
miR-127-3p	1.30E-02	2.45E-02	0.84	2.03E-01	2.89E-01	0.89	1.61E-02	2.51E-02
miR-1296	4.31E-06	2.87E-05	2.26	9.63E-02	1.66E-01	1.25	5.92E-01	6.05E-01

miR-130a	5.46E-01	6.22E-01	0.97	2.26E-02	6.96E-02	1.23	1.44E-04	4.31E-04
miR-130b	2.88E-01	3.81E-01	1.25	1.89E-02	6.23E-02	1.27	9.58E-15	1.80E-12
miR-130b*	3.40E-01	4.32E-01	1.53	1.13E-02	4.34E-02	1.28	1.18E-03	2.66E-03
miR-132	5.68E-04	1.59E-03	1.76	2.68E-01	3.57E-01	1.15	1.70E-04	5.01E-04
miR-133a	9.27E-04	2.39E-03	1.47	7.50E-02	1.47E-01	1.30	1.83E-03	3.95E-03
miR-134	1.08E-01	1.62E-01	1.39	8.21E-02	1.56E-01	0.76	1.43E-08	2.38E-07
miR-139-5p	4.09E-01	5.03E-01	0.95	5.48E-01	6.21E-01	1.08	1.71E-03	3.74E-03
miR-140-3p	7.15E-06	3.84E-05	1.51	3.20E-07	1.20E-05	1.69	9.93E-13	9.34E-11
miR-140-5p	5.21E-01	6.08E-01	0.96	1.13E-03	7.61E-03	1.31	2.24E-05	1.08E-04
miR-141	1.44E-04	5.29E-04	1.95	4.12E-04	3.23E-03	2.27	6.56E-01	6.63E-01
miR-142-3p	9.97E-09	2.75E-07	0.78	1.05E-05	1.32E-04	0.72	2.68E-05	1.17E-04
miR-142-5p	1.26E-03	3.01E-03	0.90	6.19E-04	4.48E-03	0.82	3.76E-05	1.44E-04
miR-143	1.77E-01	2.50E-01	0.86	7.24E-01	7.69E-01	0.91	9.35E-03	1.58E-02
miR-145	5.28E-03	1.10E-02	0.83	4.56E-02	1.07E-01	1.16	1.45E-10	6.81E-09
miR-146a	3.63E-01	4.55E-01	0.97	8.37E-01	8.55E-01	1.00	2.57E-08	3.71E-07
miR-146a*	3.61E-06	2.84E-05	2.27	9.21E-02	1.63E-01	1.25	3.28E-01	3.55E-01
miR-146b-5p	2.80E-01	3.73E-01	1.39	2.84E-01	3.68E-01	0.92	2.87E-02	4.11E-02
miR-148a	3.93E-06	2.84E-05	1.36	1.18E-07	5.56E-06	1.54	3.30E-04	8.99E-04
miR-148b	8.14E-03	1.65E-02	1.10	2.99E-02	8.03E-02	1.12	3.35E-04	9.00E-04
miR-150	4.23E-04	1.24E-03	1.74	1.82E-01	2.67E-01	0.90	4.80E-05	1.67E-04
miR-151-5p	3.41E-04	1.05E-03	0.88	5.88E-01	6.54E-01	0.98	2.94E-03	5.87E-03
miR-152	3.26E-01	4.23E-01	1.07	1.66E-03	1.07E-02	1.23	3.42E-02	4.77E-02
miR-153	6.20E-05	2.41E-04	2.11	3.62E-01	4.47E-01	1.16	5.24E-03	9.85E-03
miR-1537	3.51E-06	2.84E-05	2.24	9.30E-02	1.63E-01	1.23	2.30E-03	4.78E-03
miR-155	2.48E-01	3.35E-01	1.13	8.08E-01	8.30E-01	0.97	3.16E-06	2.20E-05
miR-15a	8.71E-06	4.31E-05	1.27	6.20E-01	6.77E-01	1.02	8.61E-02	1.06E-01
miR-15b	1.61E-01	2.31E-01	0.92	1.72E-09	3.22E-07	0.51	4.77E-08	5.61E-07
miR-15b*	2.37E-04	7.68E-04	1.27	7.16E-02	1.42E-01	1.15	3.79E-01	4.07E-01
miR-16	1.44E-15	2.70E-13	1.68	9.92E-03	4.18E-02	1.25	1.93E-05	1.01E-04
miR-17	7.12E-01	7.70E-01	1.07	3.96E-01	4.71E-01	1.05	3.05E-01	3.32E-01
miR-17*	8.49E-02	1.32E-01	1.50	7.47E-03	3.38E-02	1.38	4.06E-01	4.31E-01
miR-181a	1.49E-08	3.51E-07	0.77	2.43E-05	2.54E-04	0.71	2.92E-02	4.16E-02

miR-181a-2*	1.70E-05	7.63E-05	2.07	1.37E-02	4.97E-02	1.26	3.12E-04	8.63E-04
miR-181a*	3.04E-05	1.27E-04	2.19	1.17E-01	1.90E-01	1.22	2.43E-07	1.99E-06
miR-181c	1.55E-04	5.50E-04	0.87	3.13E-01	3.98E-01	0.45	2.62E-04	7.45E-04
miR-181d	2.47E-05	1.06E-04	2.04	3.30E-01	4.13E-01	1.13	5.40E-01	5.55E-01
miR-186	8.22E-01	8.49E-01	1.03	3.54E-06	6.65E-05	1.35	7.15E-08	7.91E-07
miR-18a	6.00E-02	9.97E-02	1.05	7.56E-03	3.38E-02	1.10	2.51E-04	7.27E-04
miR-18a*	3.85E-03	8.41E-03	1.75	1.79E-01	2.65E-01	1.18	2.01E-05	1.02E-04
miR-18b	1.77E-04	5.96E-04	1.14	4.38E-02	1.04E-01	1.14	5.64E-09	1.06E-07
miR-1908	6.60E-01	7.17E-01	1.33	6.10E-02	1.27E-01	1.18	3.02E-01	3.30E-01
miR-191	2.21E-02	3.95E-02	0.92	2.31E-02	6.96E-02	1.09	4.73E-05	1.67E-04
miR-195	7.65E-06	4.00E-05	1.99	7.63E-02	1.48E-01	1.25	1.51E-02	2.38E-02
miR-196b	4.69E-03	9.91E-03	1.70	1.99E-01	2.85E-01	1.10	4.54E-02	5.89E-02
miR-196b*	5.85E-06	3.36E-05	2.26	8.97E-02	1.63E-01	1.25	8.81E-08	9.20E-07
miR-197	7.84E-01	8.19E-01	0.97	1.05E-07	5.56E-06	1.98	5.10E-01	5.30E-01
miR-199a-3p	2.23E-02	3.96E-02	0.90	1.94E-01	2.81E-01	1.07	1.04E-02	1.72E-02
miR-199a-5p	3.46E-04	1.05E-03	0.81	2.73E-01	3.61E-01	1.06	1.32E-01	1.58E-01
miR-19a	3.24E-08	6.15E-07	1.30	1.03E-02	4.18E-02	1.14	2.05E-01	2.30E-01
miR-19b	4.04E-07	5.06E-06	1.23	2.46E-02	7.13E-02	0.89	1.19E-05	6.96E-05
miR-200b	5.27E-06	3.20E-05	2.26	8.13E-02	1.56E-01	1.25	8.50E-01	8.50E-01
miR-20a	4.93E-10	2.32E-08	1.19	5.76E-01	6.49E-01	0.97	1.27E-04	3.92E-04
miR-20a*	8.81E-02	1.35E-01	1.13	2.75E-03	1.57E-02	0.71	3.00E-03	5.94E-03
miR-20b	3.87E-06	2.84E-05	2.15	1.49E-01	2.27E-01	1.23	3.96E-01	4.23E-01
miR-21	1.34E-02	2.48E-02	1.06	5.43E-01	6.18E-01	1.05	7.47E-03	1.33E-02
miR-21*	4.05E-03	8.65E-03	1.85	6.77E-02	1.38E-01	1.14	1.24E-02	2.01E-02
miR-210	3.70E-04	1.10E-03	1.38	3.07E-06	6.41E-05	1.64	3.43E-05	1.34E-04
miR-2110	6.00E-04	1.66E-03	1.84	5.09E-03	2.45E-02	1.28	2.16E-09	6.77E-08
miR-215	1.23E-03	2.96E-03	1.59	3.71E-01	4.47E-01	1.20	1.85E-01	2.09E-01
miR-219-5p	1.46E-04	5.29E-04	2.04	2.79E-01	3.66E-01	1.16	2.75E-02	3.98E-02
miR-22	9.85E-01	9.85E-01	1.06	5.87E-03	2.76E-02	1.31	4.00E-04	1.04E-03
miR-22*	3.34E-01	4.28E-01	1.02	3.73E-01	4.47E-01	1.14	1.59E-01	1.86E-01
miR-221	2.04E-04	6.73E-04	0.88	3.65E-02	9.36E-02	0.91	2.99E-05	1.23E-04

miR-222	7.50E-03	1.53E-02	1.22	7.18E-01	7.67E-01	0.96	4.10E-02	5.43E-02
miR-223	5.28E-01	6.09E-01	1.01	1.27E-01	2.00E-01	0.93	1.43E-01	1.70E-01
miR-223*	9.03E-04	2.36E-03	1.59	1.01E-01	1.69E-01	1.13	2.34E-01	2.60E-01
miR-23a	3.79E-01	4.72E-01	0.98	1.91E-01	2.79E-01	0.94	1.56E-07	1.40E-06
miR-23b	1.86E-02	3.36E-02	0.92	6.90E-02	1.38E-01	1.12	3.22E-03	6.30E-03
miR-24	8.81E-02	1.35E-01	1.06	9.21E-02	1.63E-01	0.93	1.70E-01	1.98E-01
miR-24-2*	7.30E-01	7.81E-01	1.15	5.31E-01	6.12E-01	0.91	2.17E-05	1.07E-04
miR-26a	1.83E-07	2.64E-06	0.80	4.25E-06	7.26E-05	0.73	1.74E-01	2.00E-01
miR-26a-1*	1.68E-06	1.66E-05	2.24	1.22E-01	1.96E-01	1.23	1.76E-07	1.51E-06
miR-26b	1.51E-01	2.20E-01	0.95	5.53E-05	5.47E-04	0.78	4.00E-09	8.80E-08
miR-26b*	6.57E-02	1.07E-01	1.52	6.03E-01	6.66E-01	0.97	1.20E-03	2.68E-03
miR-27a	2.34E-01	3.18E-01	0.93	1.33E-08	1.25E-06	0.50	3.02E-05	1.23E-04
miR-27b	7.97E-02	1.25E-01	1.06	8.60E-01	8.74E-01	1.00	1.99E-02	2.97E-02
miR-28-3p	1.55E-01	2.24E-01	0.97	9.80E-01	9.80E-01	1.01	1.33E-01	1.58E-01
miR-28-5p	6.66E-02	1.07E-01	0.91	8.90E-01	8.99E-01	0.96	4.85E-01	5.10E-01
miR-299-5p	6.36E-04	1.73E-03	1.95	5.15E-01	6.01E-01	1.10	3.39E-11	2.12E-09
miR-29a	4.83E-01	5.73E-01	1.31	3.02E-01	3.88E-01	1.09	9.05E-03	1.56E-02
miR-29a*	1.64E-01	2.33E-01	1.61	6.20E-01	6.77E-01	0.99	4.69E-05	1.67E-04
miR-29b	4.32E-04	1.25E-03	1.29	3.76E-03	2.02E-02	1.18	5.06E-01	5.28E-01
miR-29b-2*	8.30E-04	2.23E-03	1.90	8.43E-02	1.59E-01	1.26	9.63E-06	5.84E-05
miR-29c	1.42E-01	2.08E-01	1.23	7.50E-01	7.87E-01	0.95	1.96E-02	2.94E-02
miR-29c*	1.78E-03	4.19E-03	2.00	3.76E-02	9.43E-02	1.28	6.70E-03	1.21E-02
miR-301a	3.60E-01	4.54E-01	0.97	1.03E-02	4.18E-02	1.58	1.52E-08	2.38E-07
miR-301b	6.28E-02	1.04E-01	1.44	5.23E-01	6.07E-01	1.07	2.85E-01	3.14E-01
miR-30a	4.43E-06	2.87E-05	2.01	1.01E-05	1.32E-04	1.51	8.12E-06	5.09E-05
miR-30b	1.61E-07	2.52E-06	0.82	5.06E-04	3.81E-03	0.80	1.51E-05	8.59E-05
miR-30c	1.47E-05	6.75E-05	0.85	2.48E-01	3.36E-01	0.92	8.23E-02	1.02E-01
miR-30e*	7.48E-01	7.85E-01	1.29	2.96E-02	8.03E-02	1.19	9.28E-02	1.13E-01
miR-32	6.60E-02	1.07E-01	1.21	2.05E-01	2.89E-01	1.11	3.74E-03	7.18E-03
miR-320a	3.34E-01	4.28E-01	0.98	2.23E-03	1.31E-02	0.81	2.56E-03	5.23E-03
miR-324-3p	2.11E-01	2.91E-01	1.25	1.62E-04	1.43E-03	1.46	2.66E-04	7.45E-04
miR-324-5p	9.84E-01	9.85E-01	0.99	1.78E-01	2.65E-01	1.06	6.85E-02	8.65E-02
miR-326	8.64E-01	8.83E-01	1.04	2.17E-03	1.31E-02	1.51	4.10E-02	5.43E-02

miR-328	2.76E-03	6.42E-03	0.90	9.60E-01	9.65E-01	1.01	6.29E-04	1.54E-03
miR-329	7.40E-01	7.85E-01	0.94	2.33E-02	6.96E-02	0.79	7.58E-04	1.80E-03
miR-330-5p	4.37E-06	2.87E-05	2.20	5.03E-02	1.11E-01	1.27	1.72E-06	1.24E-05
miR-331-3p	9.79E-01	9.85E-01	1.03	2.39E-05	2.54E-04	0.69	8.09E-06	5.09E-05
miR-335	1.05E-01	1.58E-01	1.15	7.85E-07	2.11E-05	0.57	3.54E-03	6.87E-03
miR-338-3p	1.11E-01	1.65E-01	0.95	3.25E-02	8.50E-02	1.21	1.61E-02	2.51E-02
miR-339-3p	3.54E-02	6.11E-02	1.38	4.38E-02	1.04E-01	1.21	2.65E-05	1.17E-04
miR-339-5p	2.84E-02	5.00E-02	0.93	2.80E-01	3.66E-01	1.06	1.91E-02	2.93E-02
miR-33a	1.02E-03	2.59E-03	0.83	3.28E-01	4.13E-01	0.92	7.77E-03	1.35E-02
miR-33b	1.12E-05	5.26E-05	2.10	1.45E-01	2.23E-01	1.16	1.03E-02	1.72E-02
miR-340*	4.95E-02	8.30E-02	1.85	6.23E-01	6.77E-01	1.03	4.17E-05	1.57E-04
miR-34a	6.28E-05	2.41E-04	1.98	4.02E-03	2.04E-02	1.63	1.50E-03	3.31E-03
miR-361-3p	1.13E-03	2.79E-03	1.81	2.72E-02	7.51E-02	1.29	6.14E-04	1.54E-03
miR-370	5.90E-01	6.57E-01	1.26	2.28E-01	3.15E-01	0.89	7.48E-05	2.51E-04
miR-374a	5.94E-01	6.57E-01	1.03	2.63E-02	7.49E-02	1.17	1.92E-02	2.93E-02
miR-374b	2.96E-01	3.89E-01	0.95	2.00E-02	6.48E-02	1.13	7.00E-01	7.04E-01
miR-376a	4.33E-01	5.25E-01	0.86	2.22E-05	2.54E-04	0.43	2.06E-03	4.36E-03
miR-376b	8.31E-01	8.54E-01	1.18	5.23E-02	1.14E-01	0.83	5.51E-04	1.40E-03
miR-376c	2.19E-01	3.00E-01	1.12	4.89E-02	1.09E-01	0.80	1.00E-02	1.69E-02
miR-377	5.40E-01	6.19E-01	1.11	3.96E-03	2.04E-02	0.62	3.11E-08	3.90E-07
miR-378	8.72E-03	1.74E-02	1.39	4.22E-01	4.96E-01	1.03	4.30E-02	5.62E-02
miR-382	1.05E-01	1.58E-01	1.19	4.25E-02	1.04E-01	0.71	1.28E-07	1.20E-06
miR-409-3p	4.84E-01	5.73E-01	1.05	5.28E-02	1.14E-01	0.81	5.37E-03	9.99E-03
miR-409-5p	7.98E-05	3.00E-04	1.91	5.34E-01	6.12E-01	1.05	1.45E-01	1.70E-01
miR-410	6.70E-02	1.07E-01	1.45	4.65E-02	1.07E-01	0.80	9.27E-03	1.58E-02
miR-411	4.03E-01	4.98E-01	1.46	6.87E-01	7.38E-01	1.00	1.12E-01	1.36E-01
miR-411*	6.58E-06	3.64E-05	2.26	9.63E-02	1.66E-01	1.25	2.46E-01	2.72E-01
miR-423-3p	1.86E-01	2.60E-01	0.95	1.40E-02	4.97E-02	1.13	5.16E-01	5.33E-01
miR-423-5p	5.28E-01	6.09E-01	0.99	2.69E-02	7.51E-02	1.17	1.34E-02	2.16E-02
miR-425	9.70E-03	1.90E-02	1.07	3.71E-01	4.47E-01	1.06	3.11E-02	4.39E-02
miR-425*	5.21E-01	6.08E-01	0.98	5.99E-02	1.27E-01	0.88	3.52E-02	4.86E-02
miR-431	9.61E-01	9.77E-01	1.18	1.23E-02	4.63E-02	0.71	7.72E-03	1.35E-02
miR-432	7.31E-01	7.81E-01	1.14	4.08E-02	1.01E-01	0.88	9.06E-05	2.94E-04

miR-433	5.79E-01	6.56E-01	1.51	2.23E-01	3.13E-01	0.92	5.36E-05	1.83E-04
miR-451	5.47E-09	2.06E-07	1.96	1.50E-02	5.14E-02	1.53	1.96E-09	6.77E-08
miR-484	4.14E-01	5.05E-01	1.03	8.77E-03	3.84E-02	1.14	6.32E-01	6.42E-01
miR-486-5p	8.90E-11	5.58E-09	1.83	9.57E-06	1.32E-04	1.80	3.79E-09	8.80E-08
miR-491-5p	5.91E-01	6.57E-01	1.32	9.88E-02	1.67E-01	1.16	2.80E-03	5.66E-03
miR-496	3.93E-06	2.84E-05	2.11	3.36E-01	4.19E-01	1.18	3.40E-05	1.34E-04
miR-497	7.45E-01	7.85E-01	1.45	4.74E-02	1.07E-01	1.35	4.97E-02	6.40E-02
miR-503	3.47E-03	7.77E-03	1.59	4.70E-02	1.07E-01	1.44	2.64E-05	1.17E-04
miR-505	1.14E-03	2.79E-03	1.53	1.45E-02	5.05E-02	1.32	1.83E-05	9.84E-05
miR-505*	4.95E-05	1.98E-04	2.10	1.67E-02	5.60E-02	1.24	8.75E-04	2.06E-03
miR-532-3p	8.80E-03	1.74E-02	1.62	1.27E-01	2.00E-01	1.13	5.11E-02	6.54E-02
miR-532-5p	3.00E-03	6.87E-03	1.44	1.07E-02	4.18E-02	1.32	1.35E-04	4.08E-04
miR-548c-5p	2.84E-04	9.06E-04	1.83	1.08E-01	1.79E-01	1.15	3.89E-03	7.38E-03
miR-551b	5.83E-01	6.57E-01	1.20	3.12E-02	8.26E-02	0.79	4.45E-05	1.64E-04
miR-589	9.78E-06	4.71E-05	2.24	6.16E-02	1.27E-01	1.28	6.27E-04	1.54E-03
miR-589*	8.20E-06	4.16E-05	2.25	9.80E-02	1.67E-01	1.25	4.40E-04	1.13E-03
miR-590-5p	3.77E-06	2.84E-05	1.32	1.53E-01	2.32E-01	1.13	6.65E-03	1.21E-02
miR-625*	4.60E-02	7.79E-02	1.35	2.25E-01	3.13E-01	1.21	7.84E-02	9.83E-02
miR-628-3p	1.31E-02	2.45E-02	1.79	2.46E-01	3.36E-01	1.18	4.67E-01	4.93E-01
miR-629	3.69E-03	8.16E-03	1.66	6.90E-02	1.38E-01	1.20	1.80E-01	2.05E-01
miR-652	1.15E-02	2.21E-02	0.92	7.59E-01	7.89E-01	1.00	2.00E-03	4.27E-03
miR-744	3.31E-02	5.76E-02	0.94	3.67E-01	4.47E-01	1.08	4.28E-02	5.62E-02
miR-766	3.02E-01	3.95E-01	0.94	4.58E-03	2.27E-02	1.20	2.76E-05	1.18E-04
miR-889	2.26E-05	9.87E-05	2.16	1.45E-01	2.23E-01	1.21	1.16E-03	2.66E-03
miR-92a	1.00E-02	1.94E-02	1.31	3.03E-01	3.88E-01	1.02	5.28E-02	6.70E-02
miR-92b	7.42E-07	8.72E-06	2.23	2.16E-02	6.77E-02	1.28	4.22E-09	8.80E-08
miR-93	1.02E-08	2.75E-07	1.15	7.31E-01	7.72E-01	1.01	2.35E-02	3.45E-02
miR-93*	5.75E-03	1.19E-02	1.33	6.71E-01	7.25E-01	1.01	2.34E-05	1.10E-04
miR-940	4.78E-01	5.72E-01	1.05	2.35E-01	3.22E-01	0.71	2.31E-03	4.78E-03
miR-98	1.67E-04	5.79E-04	0.85	8.82E-02	1.62E-01	0.91	9.02E-02	1.10E-01
miR-99b	8.02E-01	8.34E-01	0.97	5.71E-02	1.22E-01	0.82	7.42E-04	1.79E-03

Design and validation of a diagnostic miRNA signature based model

1. Model construction and miRNA signature identification

The model importance metrics, Mean Decrease in Accuracy (MDA) and Mean Decrease in Gini (MDG), were obtained through the construction of a Random forests model with all 188 miRNAs on the profiling cohort. A conservative value of 3000 for n_{tree} was chosen for all steps of the construction of Random forests models of our methodology. As no significant performance change was observed for incremental values of m_{try} , a default value of $m_{try} = \sqrt{\text{number of miRNAs}}$ has been chosen for all steps of the construction of Random forests models of our methodology. A value of $m = 25$ was chosen for the number of top ranked miRNAs to be used to generate the combinations tested. It corresponds to conservative values of MDG greater than or equal to 1 and MDA greater than or equal to 0.001. The list of the 25 top ranked miRNAs is available in table 5 below.

15

Table 5: list of the 25 top ranked miRNAs. The model importance metrics, Mean Decrease in Accuracy (MDA) and Mean Decrease in Gini (MDG), were obtained through the construction of a Random forests model with all 188 miRNAs on the profiling cohort. The first top ranked 25 miRNAs were chosen to be used to generate the combinations tested. It corresponds to conservative values of MDG greater than or equal to 1 and MDA greater than or equal to 0.001.

20

Top ranked miRNA	Name	Sequence	mean MDA- MDG
miR-16	miR-16-5p	UAGCAGCACGUAAAUAUUGGCG	1
let-7d	let-7d-5p	AGAGGUAGUAGGUUGCAUAGUU	2

miR-103	miR-103a-3p	AGCAGCAUUGUACAGGGCUAUGA	3.5
miR-181a	miR-181a-5p	AACAUUCAACGCUGUCGGUGAGU	4
miR-107	miR-107	AGCAGCAUUGUACAGGGCUAUGA	4.5
miR-142-3p	miR-142-3p	UGUAGUGUUUCCUACUUUAUGGA	6
miR-486-5p	miR-486-5p	UCCUGUACUGAGCUGCCCCGAG	7
miR-148a	miR-148a-3p	UCAGUGCACUACAGAACUUUGU	9
miR-20a	miR-20a-5p	UAAAGUGCUUAUAGUGCAGGUAG	9.5
let-7i	let-7i-5p	UGAGGUAGUAGUUUGUGCUGUU	10
miR-19a	miR-19a-3p	UGUGCAAAUCUAUGCAAAACUGA	10.5
let-7f-1*	let-7f-1-3p	CUAUACAAUCUAUUGCCUUC	12.5
miR-199a-5p	miR-199a-5p	CCCAGUGUUCAGACUACCUGUUC	13
miR-93	miR-93-5p	CAAAGUGCUGUUCGUGCAGGUAG	14.5
miR-451	miR-451a	AAACCGUUACCAUUACUGAGUU	16.5
miR-19b	miR-19b-3p	UGUGCAAAUCCAUGCAAAACUGA	17
miR-30b	miR-30b-5p	UGUAAACAUCCUACACUCAGCU	18
miR-1	miR-1-3p	UGGAAUGUAAAGAAGUAUGUAU	18
miR-26a	miR-26a-5p	UUCAAGUAAUCCAGGAUAGGCU	18.5
miR-590-5p	miR-590-5p	GAGCUUAUUCAUAAAAGUGCAG	20
miR-22*	miR-22-3p	AAGCUGCCAGUUGAAGAACUGU	20
miR-101	miR-101-3p	UACAGUACUGUGAUAAACUGAA	23
miR-22	miR-22-3p	AAGCUGCCAGUUGAAGAACUGU	23.5

miR-142-5p	miR-142-5p	CAUAAAGUAGAAAGCACUACU	23.5
miR-32	miR-32-5p	UAUUGCACAUUACUAAGUUGCA	25

5

The total number of miRNAs combinations tested amounts to 33554431.

The best performing model, based on a ten-fold cross-validation on the profiling cohort, makes use of the following miRNAs : miR-16, let-7d, miR-103, miR-107, miR-148a, let-7i, miR-19b, miR-22*. Table 6 summarizes the

10 Mann-Whitney U p-values, MDG and MDA values. Fig. 2 summarizes relative expression changes for the best-performing 8 miRNAs included in the

signature.

Table 6: Results of statistical analyses comparing the 8 miRNAs expression present in the diagnostic signature between different groups. No significant difference was observed between healthy women and benign mammary lesions patients. Healthy women and benign mammary lesions patients were thus determined as controls. The 8 diagnostic miRNAs were then compared between primary breast cancer patients, breast cancer patients in complete remission, metastatic breast cancer patients, gynecologic cancer patients and controls. P-values and Benjamini-Hochberg adjusted p-values were obtained through a Mann-Whitney U test. The Mean Decrease in Accuracy (MDA) and the Mean Decrease in Gini (MDG) were obtained through the Random Forests model construction step on the profiling cohort.

miRNAs	Healthy women Vs Benign mammary lesions		Controls Vs Primary Breast cancers		Controls Vs Metastatic breast cancers		Controls Vs Gynecologic cancers		Controls Vs Breast cancers in remission		M D A	MDG
	P	Fold Change	P	Fold Change	P	Fold Change	P	Fold Change	P	Fold Change		
miR-16	=0.05	1.2	<0.001	1.7	<0.05	1.2	>0.05	1.5	<0.05	1.4	0.0122	4.7131
Let-7d	>0.05	0.9	<0.001	0.7	<0.01	0.8	>0.05	0.9	<0.01	1.2	0.0116	4.1416
miR-103	>0.05	1	<0.001	0.8	>0.05	1	>0.05	1	>0.05	1	0.0037	1.9599
miR-107	>0.05	0.9	<0.001	0.8	<0.001	0.8	>0.05	0.9	>0.05	1	0.0035	1.8914
miR-148a	>0.05	1	<0.001	1.4	<0.001	1.6	<0.0001	1.9	<0.001	0.5	0.0025	1.6358
Let-7i	>0.05	0.9	<0.01	0.9	<0.05	1.1	>0.05	1.1	>0.05	0.9	0.0024	1.6012
miR-19b	>0.05	1.2	<0.001	1.2	>0.05	0.9	>0.05	1.1	<0.001	0.8	0.0019	1.1567
miR-22*	>0.05	0.9	>0.05	1	>0.05	1.1	<0.01	1.5	<0.01	1.4	0.0017	1.1286

An AUC of 0.84 ± 0.02 was obtained when doing the ten-fold cross-validation on the profiling cohort.

A threshold value of 0.68 was chosen to allow the computation of finite sensitivity and specificity values. The value of 0.68 corresponded to an acceptable trade-off between a high sensitivity (>0.9) and a satisfactory specificity (>0.5).

2. Model validation

The validation of our models on the independent cohort including 198 samples yielded an AUC of 0.81 ± 0.01 . Fig. 3A represents the ROC curve obtained by validating the model on the independent cohort.

- 5 With a threshold value of 0.68, a sensitivity of 0.91 ± 0.01 and a specificity of 0.49 ± 0.03 were obtained.

The test of the classification model on the other cancer groups yielded slightly lower values for sensitivity (0.80 ± 0.05 for metastatic breast cancer patients) and specificity (0.40 ± 0.08 for breast cancer patients in remission and 0.41 ± 0.06 for
10 gynecologic cancer patients) (Fig. 3B). As shown on Fig. 3B, the breast cancer patients in complete remission show the same classification outcome distribution as the control group.

3. Alternative signatures

Six different other miRNAs combinations (from 3 to 15 miRNAs) that also allows
15 satisfactory results on the profiling cohort were tested on the independent cohorts, and the results can be found in table 7 below. Interestingly, all of them shared 3 miRNAs : miR-16, let-7d and miR-103. The performances of these 3 microRNAs was evaluated: an AUC of 0.76 ± 0.010 was obtained on the profiling cohort, and $AUC=0.71 \pm 0.016$ on the validation cohort (see table 7 for details). Hence the
20 combination of miR-16, let-7d and miR-103 also allows discrimination of breast cancer patients from healthy woman.

Table 7: performances of 6 alternative miRNAs combinations. The performances of 6 different other miRNAs combinations (from 3 to 15 miRNAs) were also tested on the validation cohort. AUC for each of these combinations are calculated as well as their
25 sensitivity and specificity at the specified threshold. Specificities on independent cohorts of gynecological cancers, benign mammary lesions, healthy women and breast cancer in remission were also tested, as well as sensitivity on metastatic breast cancers.

					Performances on other cohorts				
	Performances on independent cohort (Primary Breast Cancer + Healthy women + Benign Mammary Lesions)				<u>Sensitivity</u>	<u>Specificity</u>			
	<u>AUC</u>	<u>Thresh</u> <u>hold</u>	<u>Sensitivity</u>	<u>Specificity</u>	<u>Metastatic</u>	<u>Gynecologic</u>	<u>Benign mammary lesions</u>	<u>Healthy women</u>	<u>Breast cancer remission</u>
miR-16, let-7d, miR-103	0.74	0.83	0.90	0.32	0.79	0.41	0.40	0.22	0.32
miR-16, let-7d, miR-103, miR-181a, miR-107, miR-142-3p, miR-148a, let-7-1*, miR-199a-5p, miR-590-5p, miR-32	0.80	0.66	0.90	0.46	0.79	0.40	0.57	0.35	0.45
miR-16, let-7d, miR-103, miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, miR-20a, let-7i, miR-19a, let-7-1*, miR-199a-5p, miR-22, miR-32	0.80	0.68	0.90	0.46	0.72	0.56	0.57	0.36	0.42
miR-16, let-7d, miR-103, miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, let-7i, let-7-1*, miR-199a-5p, miR-30b, miR-22, miR-142-5p, miR-32	0.81	0.69	0.91	0.44	0.77	0.51	0.54	0.34	0.43
miR-16, let-7d, miR-103, miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, miR-20a, let-7i, let-7-1*, miR-451, miR-1, miR-590-5p, miR-22, miR-142-5p	0.81	0.68	0.90	0.46	0.81	0.48	0.58	0.34	0.42
miR-16, let-7d, miR-103, miR-148a, miR-19a, miR-199a-5p, miR-22	0.81	0.69	0.91	0.50	0.81	0.46	0.58	0.42	0.43

Comparison between the 8-miRNA signature and the established diagnostic methods

Next, we sought to compare the performance of the best performing 8 miRNA signature, namely miR-16, let-7d, miR-103, miR-107, miR-148a, let-7i, miR-19b and miR-22*, to mammography and CA15.3 dosage.

The accuracy of screening mammography is sorely affected by the age. Indeed, young women have dense breast making interpretation of mammography more difficult ($AUC = 0.69 \pm 0.05$ for women under the age of 50 yr). As shown in the Fig. 4A, the diagnostic accuracy of the miRNA signature does not appear to be affected by the age, as the AUC remaining stable at 0.81 in patients younger than 50 yr.

CA15.3 is the only validated biomarker in breast cancer and its accuracy is directly influenced by tumor stage, with AUC ranging from 0.56 in stage I to 0.80 in stage III breast cancers. Thereby, CA15.3 is only useful for the diagnostic of late stage and metastatic breast cancers. Interestingly, tumor stage does not seem to affect the signature miRNAs performance, remaining stable at 0.81 through breast cancer stage I to III (Fig. 4B).

The 8 miRNAs signature association with cancer related functions and pathways

The expression of miRNAs can be related to genomic rearrangement occurring in the tumor. We checked for the genomic localization of the 8 miRNAs composing the best performing signature to determine if the co-up- (or down-) regulation of some of the 8 miRNAs could be related to their genomic localization. As represented in Table 8 below, the 8 miRNAs were widely dispersed on different chromosomes, thereby eliminating this hypothesis. Common localization of miRNAs of the signature and the frequently amplified ERBB2 gene (17q11.2-q12) was also excluded.

Table 8: the best performing 8 miRNA signature: miRNAs names, accession number, and cluster as curated in the miRBase release 21. Genomic position as curated in HNGC. miRNAs families as curated in TargetScan 5.2.

miRNA	Position	Families/ Cluster
let-7i-5p (MIMAT0000415)	12q14.1	Let-7/ 98/ 4458/ 4500
let-7d-5p (MIMAT0000065)	9q22.32	Let-7/ 98/ 4458/ 4500
miR-103a-3p (MIMAT0000101)	103a-1 : 5q34 / 103a-2 : 20p13	miR-103a/ 107/ 107ab
has-miR-107 (MIMAT0000104)	10q23.31	miR-103a/ 107/ 107ab
miR-148a-3p (MIMAT0000243)	7p15.2	miR-148ab-3p/ 152
miR-19b-3p (MIMAT0000074)	19b-1 :13q31.3 / 19b-2 : Xq26.2	miR-19ab in miR17-92 cluster
miR-22-5p (MIMAT0004495)	17p13.3	miR-22-5p/ 3568
miR-16-5p (MIMAT0000069)	miR-16-1 : 13q14.2/ miR-16-2 : 3q25.33	miR-15/16

5

List of abbreviations

3'-UTR = 3'-untranslated region

AJCC-UICC = American Joint Committee on Cancer and the International Union for Cancer Control

10 AUC = area under the curve

Cq = quantification cycle

DNA = deoxyribonucleic acid

HER2 = human epidermal growth factor 2

LNA = locked nucleic acid

15 MDA = mean decrease accuracy

MDG = mean decrease gini

miRNAs = microRNAs

mRNAs = messenger RNAs

NA = not accessed

20 Ns = non significant

RNA = ribonucleic acid

ROC = receiver-operating characteristic

SEM = standard error of the mean.

CLAIMS

1. An in vitro method of diagnosing whether a subject has, or is at risk for developing, breast cancer, comprising the steps:
- a) measuring the expression level of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103 in a test biological fluid sample from said subject;
- b) comparing the expression level obtained in step a) to a reference value of a control biological fluid sample from a healthy subject;
- wherein an alteration in the expression level of any one or more of said microRNAs is indicative of the subject either having, or being at risk for developing, breast cancer.
2. An in vitro method of diagnosing whether a subject has, or is at risk for developing, breast cancer, comprising the steps:
- a) measuring the expression level of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103 in a test biological fluid sample from said subject;
- b) comparing the expression level obtained in step a) to a reference value of a control biological fluid sample from a subject having a known diagnosis of breast cancer;
- wherein a similar expression level of said microRNAs is indicative of the subject either having, or being at risk for developing, breast cancer.
3. The in vitro method according to claims 1 or 2, wherein said reference value is the expression level of the same respective microRNAs.
4. The in vitro method according to any one of claims 1 to 3, wherein breast cancer is primary or metastatic breast cancer.
5. The in vitro method according to any one of claims 1, 3 or 4, wherein an up-regulation of the expression level of miR-16 and a down-regulation of the expression level of let-7d and miR-103 as compared to the expression level

of said respective microRNAs in a control biological fluid sample from a healthy subject is indicative of the subject either having, or being at risk for developing, primary breast cancer.

- 5 6. The in vitro method according to any one of claims 1, 3 or 4, wherein an up-regulation of the expression level of miR-16, a down-regulation of the expression level of let-7d and no change of the expression level of miR-103 as compared to the expression level of said respective microRNAs in a control biological fluid sample from a healthy subject is indicative of the
- 10 subject either having, or being at risk for developing, metastatic breast cancer.
7. The in vitro method according to any one of claims 1 to 6, wherein in step a) the expression level of one or more additional microRNAs is further
- 15 measured, wherein said additional microRNAs are selected from the group consisting of: miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, miR-20a, let-7i, miR-19a, let-7f-1*, miR-199a-5p, miR-93, miR-451, miR-19b, miR-30b, miR-1, miR-26a, miR-22*, miR-590-5p, miR-101, miR-22, miR-142-5p and miR-32 or combinations thereof.
- 20 8. The in vitro method according to claim 7, wherein said additional microRNAs are the following five microRNAs selected from the group consisting of: miR-107, miR-148a, let-7i, miR-19b and miR-22*.
- 25 9. The in vitro method according to claims 7 or 8, wherein an up-regulation of the expression level of miR-148a and miR-19b, a down-regulation of the expression level of miR-107 and let-7i and no change of the expression level of miR-22* as compared to the expression level of said respective microRNAs in a control biological fluid sample from a healthy subject is
- 30 indicative of the subject either having, or being at risk for developing, primary breast cancer.
10. The in vitro method according to claims 7 or 8, wherein an up-regulation of the expression level of miR-148a, let-7i and miR-22* and a down-regulation

of the expression level of miR-107 and miR-19b as compared to the expression level of said respective microRNAs in a control biological fluid sample from a healthy subject is indicative of the subject either having, or being at risk for developing, metastatic breast cancer.

5

11. The in vitro method according to any one of claims 1 to 10, wherein said expression level is measured by performing a quantitative reverse-transcription real-time polymerase chain reaction (qRT-PCR).

10 12. The in vitro method according to any one of claims 1 to 11, wherein said expression level in the test or control biological fluid sample is normalized with a mean expression level of the fifty most expressed microRNAs in said test or control biological fluid sample.

15 13. The in vitro method according to claim 12, wherein said fifty most expressed microRNAs, used to normalize the microRNA expression level in both the test and control biological fluid samples, are the same.

20 14. The in vitro method according to claims 12 or 13, wherein said most expressed microRNAs are selected from the group comprising miR-484, miR-652, miR-148b, miR-106a, miR-425, let-7g, miR-30b, miR-126, miR-103, miR-146a, miR-93, miR-24, miR-18b, miR-151-5p, miR-423-3p, miR-223, miR-15a, miR-142-3p, miR-26b, miR-15b, miR-191, let-7d*, let-7f, miR-21, miR-126*, miR-125a-5p, miR-181a, miR-23a, miR-222, miR-23b, miR-30c, miR-101, miR-150, miR-320a, miR-26a, miR-145, miR-486-5p, miR-16, miR-199a-5p, miR-107, miR-27a, miR-19b, miR-199a-3p, miR-221, let-7b, miR-92a, miR-27b, miR-451, miR-20a or miR-19a.

25

30 15. The in vitro method according to any one of claims 1 to 14, wherein the biological fluid sample is selected from the group comprising whole blood, blood serum, blood plasma, blood cells, urine, milk or saliva.

16. An in vitro method of monitoring the progress of breast cancer in a subject comprising the steps of:

a) measuring the expression level of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103 in two or more test biological fluid samples from said subject, taken at different time intervals;

b) determining the progress of breast cancer by comparing the expression level of said microRNAs in the measured biological fluid samples over time.

17. The in vitro method according to claim 16, wherein in step a) the expression level of the following five microRNAs selected from the group consisting of: miR-107, miR-148a, let-7i, miR-19b and miR-22* is further measured.

18. A method of treatment of breast cancer in a subject, comprising treating the subject, diagnosed as being in need of breast cancer treatment according to the method of any one of claims 1 to 15, with a treatment selected from the group consisting of: surgery, radiotherapy, chemotherapy, hormonal therapy or biological treatment.

19. A method of treatment of breast cancer in a subject, comprising the steps of:

a) determining whether the subject is in need of receiving breast cancer treatment comprising performing the method according to any one of claims 1 to 15;

b) treating the subject diagnosed in step a) as being in need of breast cancer treatment with a treatment selected from the group consisting of: surgery, radiotherapy, chemotherapy, hormonal therapy or biological treatment.

20. A kit for use in an in vitro method of diagnosing whether a subject has, or is at risk for developing, breast cancer, or in a method of treatment of breast cancer in a subject, or in monitoring the progress of breast cancer in a subject, wherein the kit comprises at least three oligonucleotide probes

specific for the detection of the following three microRNAs selected from the group consisting of: miR-16, let-7d and miR-103.

21. The kit according to claim 20, further comprising one or more oligonucleotide probes specific for the detection of one or more additional microRNAs selected from the group consisting of: miR-181a, miR-107, miR-142-3p, miR-486-5p, miR-148a, miR-20a, let-7i, miR-19a, let-7f-1*, miR-199a-5p, miR-93, miR-451, miR-19b, miR-30b, miR-1, miR-26a, miR-22*, miR-590-5p, miR-101, miR-22, miR-142-5p and miR-32 or combinations thereof.

22. The kit according to claims 20 or 21, wherein said kit consists of eight oligonucleotide probes specific for the detection of the following eight microRNAs selected from the group consisting of: miR-16, let-7d, miR-103, miR-107, miR-148a, let-7i, miR-19b and miR-22*.

23. The kit according to any one of claims 20 to 22, wherein said oligonucleotide probes are specific for the detection of cDNAs obtained from said microRNAs.

24. The kit according to any one of claims 20 to 23, adapted for performance of an assay selected from a quantitative reverse-transcription real-time polymerase chain reaction (qRT-PCR), a locked nucleic acid (LNA) real-time PCR, a northern blotting or a micro-array assay.

25. Use of a kit according to any one of claims 20 to 24 for diagnosing whether a subject has, or is at risk for developing, breast cancer, preferably by performing the method according to any one of claims 1 to 15, or for monitoring the progress of breast cancer in a subject, preferably by performing the method according to claims 16 or 17, or for treating breast cancer in a subject, preferably by performing the method according to claims 18 or 19.

Fig. 1

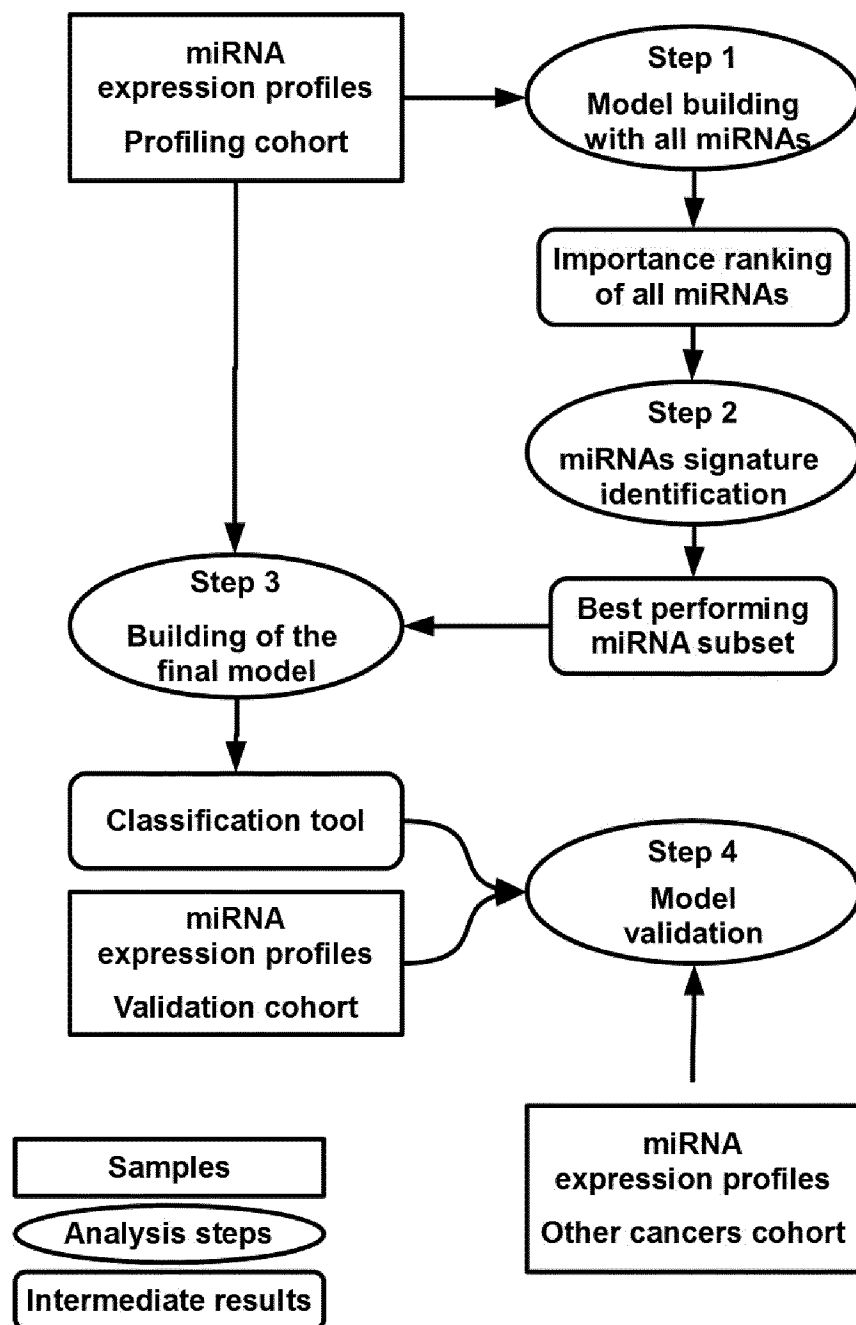


Fig. 2

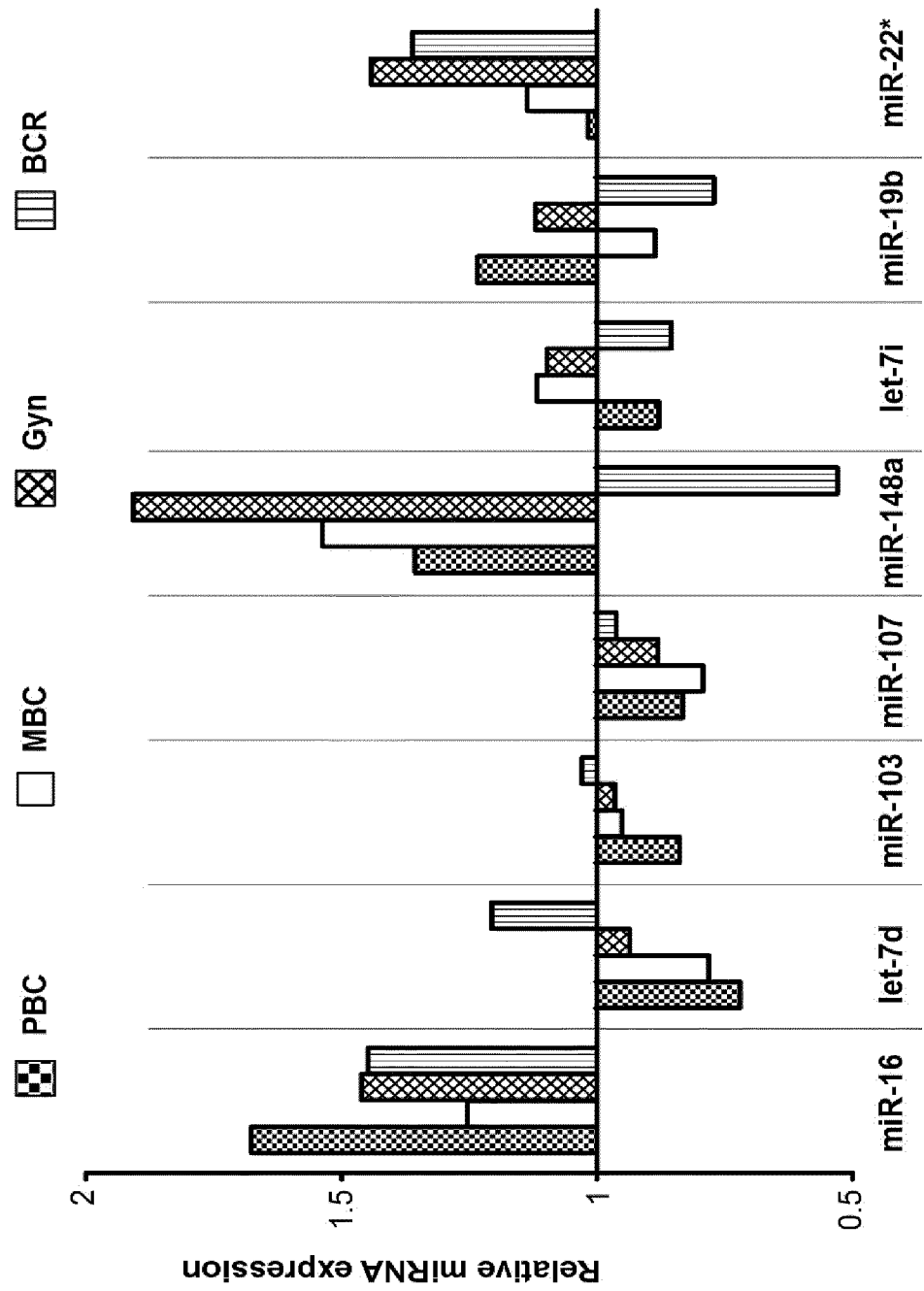


Fig. 3A

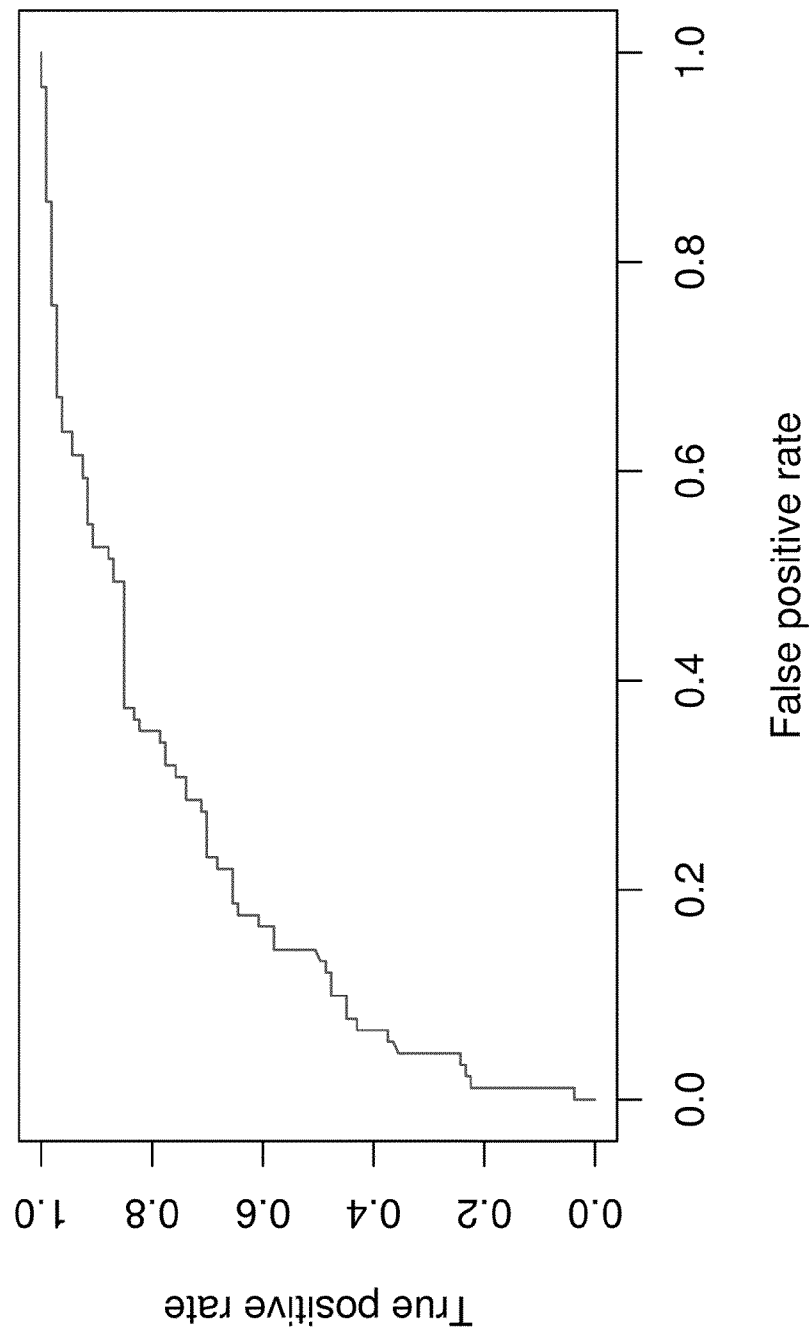


Fig. 3B

Model class prediction between samples groups

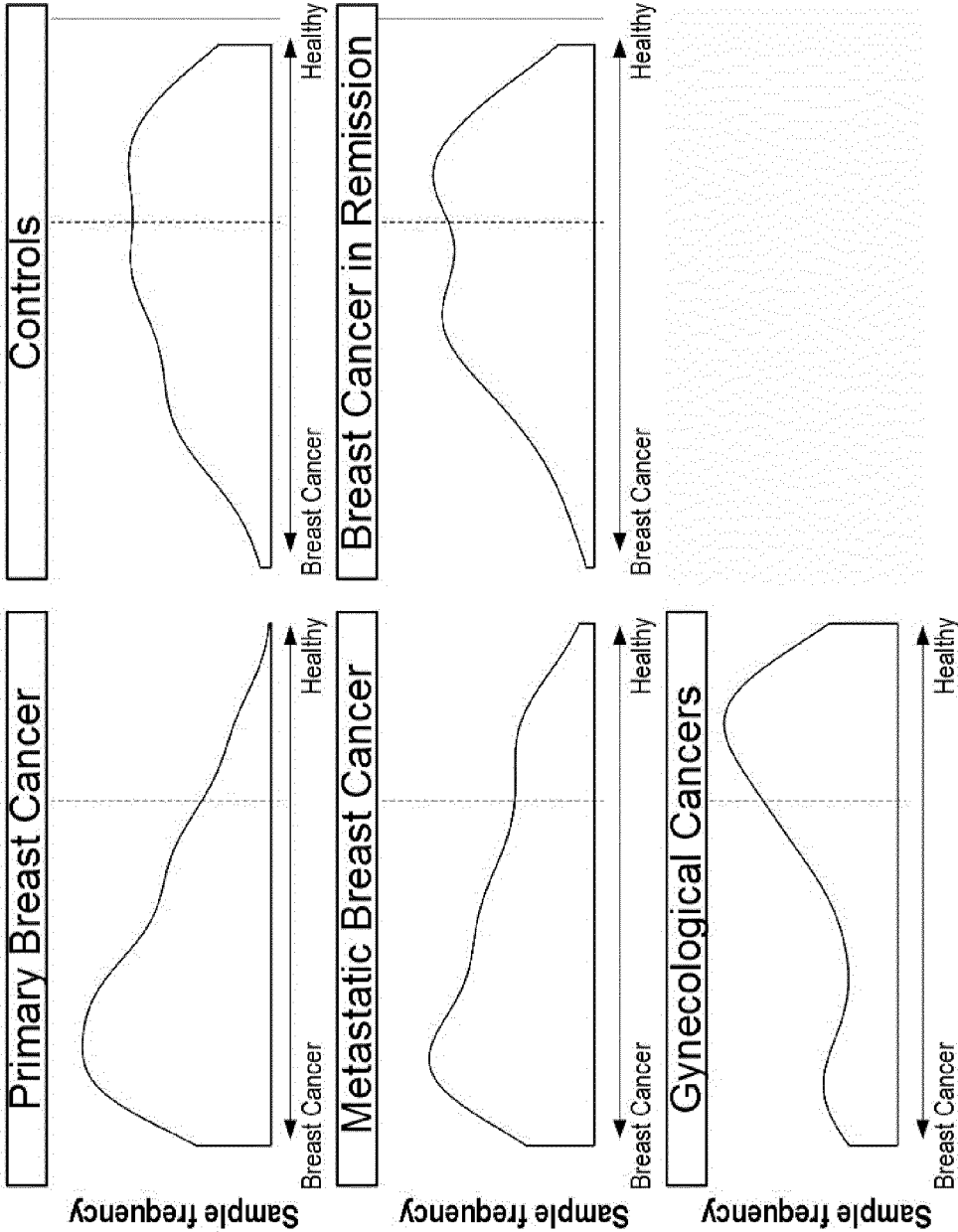


Fig. 4A

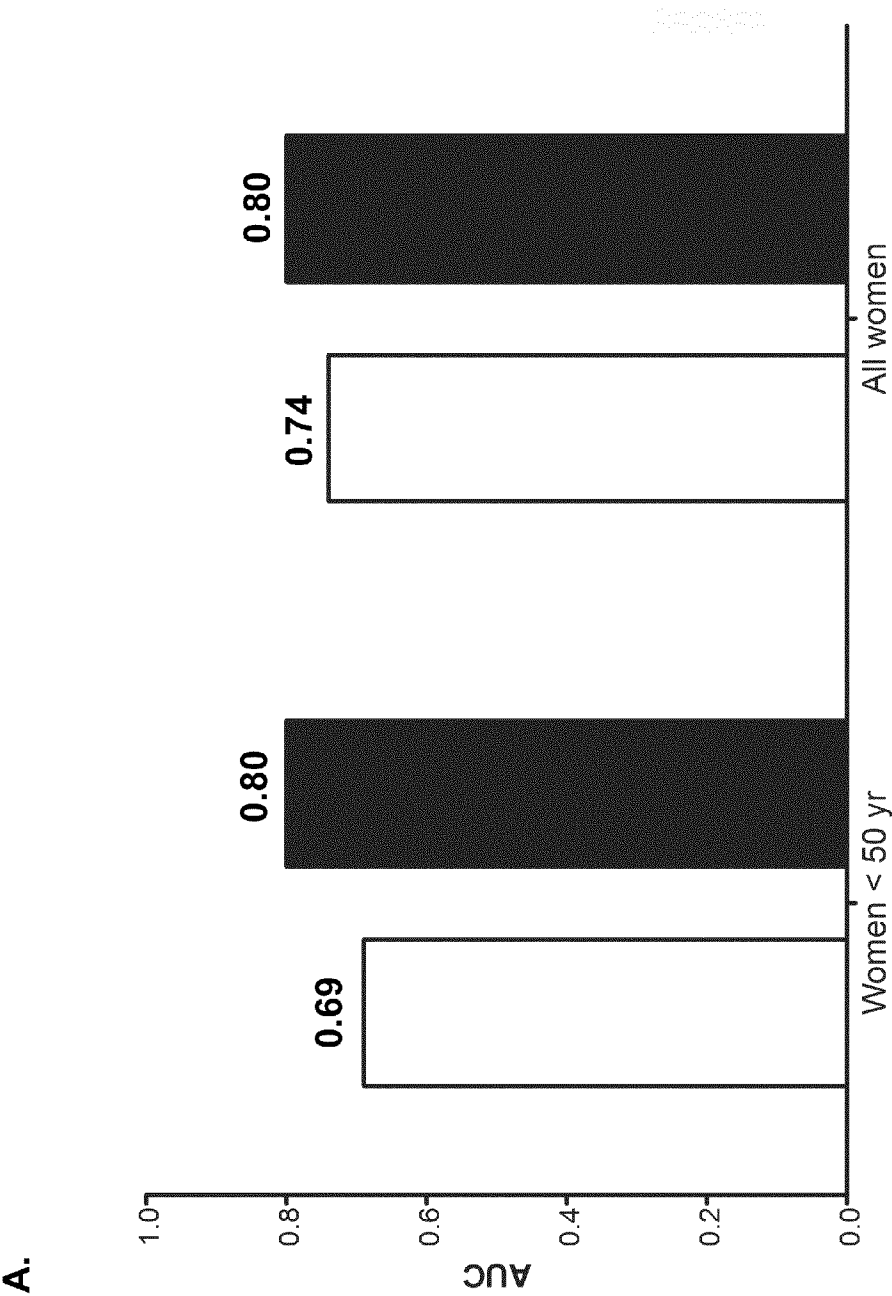
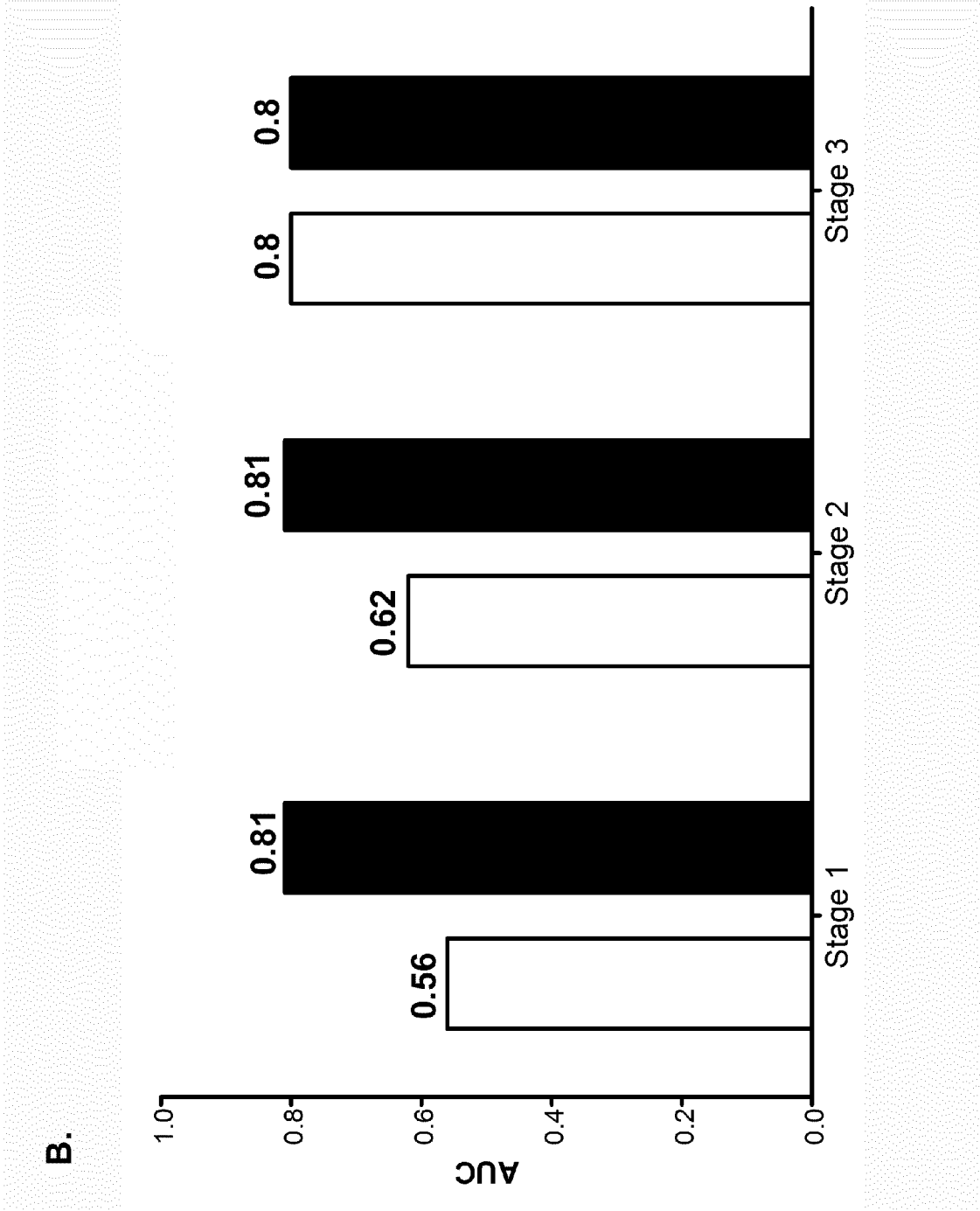


Fig. 4B



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/056028

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, WPI Data, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2014/152622 A1 (UNIV TEXAS [US]; BETH ISRAEL HOSPITAL [US]) 25 September 2014 (2014-09-25)	20-24
Y	pages 62-63; claim 6; table 5 paragraphs [0004] - [0005], [0044], [0086], [0136] paragraph [0105] -----	1-7, 11-19,25
X	WO 2014/129975 A1 (SINGAPORE HEALTH SERV PTE LTD [SG]) 28 August 2014 (2014-08-28)	20-24
A	abstract; claims 1-2; tables 2-3 page 27 -----	1-19
Y	US 2013/190386 A1 (CROCE CARLO M [US] ET AL) 25 July 2013 (2013-07-25) paragraph [0013]; claims 1-2 -----	1-7, 11-19,25
	-/--	



Further documents are listed in the continuation of Box C.



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

24 November 2015

Date of mailing of the international search report

07/12/2015

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Aguilera, Miguel

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2015/056028

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WANG XIAOPAI ET AL: "Serum miR-103 as a potential diagnostic biomarker for breast cancer", J SOUTH MED UNIV, vol. 32, 1 January 2012 (2012-01-01), pages 631-638, XP055230993, DOI: 10.3969/j.issn.1673-4254.2012.05.009 abstract -----	1-7, 11-19,25

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2015/056028

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2014152622 A1	25-09-2014	AU 2014239309 A1	08-10-2015
		CA 2905081 A1	25-09-2014
		WO 2014152622 A1	25-09-2014

WO 2014129975 A1	28-08-2014	NONE	

US 2013190386 A1	25-07-2013	AU 2013209477 A1	11-09-2014
		CA 2866052 A1	25-07-2013
		CN 104685065 A	03-06-2015
		EP 2804960 A1	26-11-2014
		JP 2015511121 A	16-04-2015
		US 2013190386 A1	25-07-2013
		US 2014378535 A1	25-12-2014
		WO 2013110053 A1	25-07-2013
