



(51) International Patent Classification:

C12Q 1/68 (2018.01) A61P 35/00 (2006.01)

(21) International Application Number:

PCT/US2018/012982

(22) International Filing Date:

09 January 2018 (09.01.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/444,085 09 January 2017 (09.01.2017) US

(71) Applicants: **DANA-FARBER CANCER INSTITUTE, INC.** [US/US]; 450 Brookline Avenue, Boston, MA 02215 (US). **THE BRIGHAM AND WOMEN'S HOSPITAL, INC.** [US/US]; 75 Francis Street, Boston, MA 02115 (US).

(72) Inventors: **CHOWDHURY, Dipanjan**; 151 Gardner Road, Brookline, MA 02445 (US). **ELIAS, Kevin M.**; 55 Harrison Street, #2, Brookline, MA 02446 (US). **FENDLER, Wojciech**; c/o Department of Biostatistics and Translational Medicine, Medical University of Lodz, Poland, 15 Mazowiecka St., 92-215 Lodz (PL). **STAWISKI, Konrad**; C/o Department of Biostatistics and Translational Medicine, Medical University of Lodz, Poland, 15 Mazowiecka St., 92-215 Lodz (PL).

(74) Agent: **CUSTER, Andrew W.** et al.; Lathrop Gage LLP, 28 State Street, Boston, MA 02109 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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).

(54) Title: CIRCULATING MICRORNA SIGNATURES FOR OVARIAN CANCER

(57) Abstract: Provided herein are methods and kits for determining the presence or absence of certain microRNA biomarkers in a blood sample of a female patient. The microRNA biomarkers are associated with ovarian cancer. Also provided are methods for screening a female subject for the presence or absence of certain microRNA biomarkers, as well as methods for treating a female subject having an ovarian cancer.



CIRCULATING MICRORNA SIGNATURES FOR OVARIAN CANCER**STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR
DEVELOPMENT**

This invention was made with Government support under grant number
5 K12HD13015 awarded by the National Institutes of Health. The Government has
certain rights in the invention.

BACKGROUND

Ovarian cancer kills more American women annually than any other
gynecologic malignancy. The five-year relative survival rate is about 80-92% when
10 diagnosed and treated in an earlier stage (stage I or II), and is about 17-40% when
diagnosed and treated in a later stage (stage III or IV). Unfortunately, most ovarian
cancers are diagnosed at a later stage due to the lack of an effective screening tool.

The difficulty in developing an effective screening tool is attributed to the low
prevalence of ovarian cancer in the general population, which limits the positive
15 predictive value of screening tests. Conventional screening algorithms based on
transvaginal ultrasounds tend to diagnose slow-growing benign masses such as
endometriomas or non-invasive tumors of low malignant potential (so-called
“borderline tumors”), rather than the rapidly proliferating high grade neoplasms
which cause nearly all ovarian cancer deaths. CA-125, a conventional biomarker of
20 ovarian cancer, is not accurate enough for ovarian cancer screening in all women
because it is affected by many noncancerous conditions. Moreover, epithelial ovarian
cancer is not a single biologic entity, but instead describes a heterogeneous set of
malignancies that manifest as ovarian masses. This heterogeneity further decreases
the sensitivity and specificity in diagnosing ovarian cancer using a single marker.

25 MicroRNAs (miRNAs) are small regulatory RNA molecules that control gene
expression by RNA silencing and post-transcriptional regulation. They are often
tissue-specific and are dysregulated in many cancers. MicroRNAs have double-
stranded hairpin structures and are more stable than messenger RNAs. Some
miRNAs can be detected in the blood and the amounts remain stable in blood samples
30 for years or even decades, providing a practical possibility for using them as

biomarkers for noninvasive cancer diagnosis. However, most studies focus on miRNAs aberrantly expressed in tumor samples rather than blood samples. There is a need in the art to identify circulating miRNAs that can amount to accurate and robust diagnosis of early stage ovarian cancer.

5

SUMMARY

The instant disclosure provides methods for determining the presence or absence and/or the amount of microRNAs in a sample (e.g., blood sample) from a female subject (e.g., human subject), as well as kits comprising probes to micro RNAs. The instant disclosure also describes methods for treating a female subject, as
10 well as methods for screening blood samples of female subjects for the presence or absence of certain microRNAs.

In one aspect, the instant disclosure provides a method for determining the presence or absence and/or amounts of microRNAs in a sample from a female subject (e.g., human subject). In some embodiments, the method comprises: (a) obtaining a
15 sample collected from the subject; and (b) determining the presence or absence and/or amounts of one or more (e.g., two or more, three or more, four or more, five or more, or six or more) test microRNAs selected from the group consisting of: hsa-miR-29a-3p (SEQ ID NO: 39); hsa-miR-92a-3p (SEQ ID NO: 47); hsa-miR-200c-3p (SEQ ID NO: 105); hsa-miR-320c (SEQ ID NO: 115); hsa-miR-335-5p (SEQ ID NO: 122);
20 hsa-miR-450b-5p (SEQ ID NO: 149); and hsa-miR-1307-5p (SEQ ID NO: 182) in the sample, thereby determining the presence or absence and/or amounts of the microRNAs in the sample. In some embodiments, step (b) comprises determining the presence or absence and/or amounts of seven test microRNAs of: hsa-miR-29a-3p (SEQ ID NO: 39); hsa-miR-92a-3p (SEQ ID NO: 47); hsa-miR-200c-3p (SEQ ID NO: 105); hsa-miR-320c (SEQ ID NO: 115); hsa-miR-335-5p (SEQ ID NO: 122);
25 hsa-miR-450b-5p (SEQ ID NO: 149); and hsa-miR-1307-5p (SEQ ID NO: 182) in the sample. In some embodiments, the presence or absence and/or amounts of no more than 2,500 (e.g., no more than 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,500, or
30 2,000) test microRNAs are determined. In some embodiments, step (b) consists of determining the presence or absence and/or amounts of seven test microRNAs of: hsa-miR-29a-3p (SEQ ID NO: 39); hsa-miR-92a-3p (SEQ ID NO: 47); hsa-miR-200c-3p

(SEQ ID NO: 105); hsa-miR-320c (SEQ ID NO: 115); hsa-miR-335-5p (SEQ ID NO: 122); hsa-miR-450b-5p (SEQ ID NO: 149); and hsa-miR-1307-5p (SEQ ID NO: 182) in the sample.

In another aspect, the instant disclosure provides a method for diagnosing ovarian cancer in a female subject (e.g., human subject), the method comprising: (a) obtaining a sample collected from the subject; (b) determining the amounts of one or more (e.g., two or more, three or more, four or more, five or more, or six or more) test microRNAs selected from the group consisting of hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-450b-5p (SEQ ID NO: 149), and hsa-miR-1307-5p (SEQ ID NO: 182) in the sample; and (c) comparing the amounts of the test microRNAs determined in step (b) to a statistical model, thereby diagnosing ovarian cancer in the subject. In some embodiments, step (b) comprises determining the amounts of seven test microRNAs of: hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-450b-5p (SEQ ID NO: 149), and hsa-miR-1307-5p (SEQ ID NO: 182) in the sample. In some embodiments, the presence or absence and/or amounts of no more than 2,500 (e.g., no more than 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,500, or 2,000) test microRNAs are determined. In some embodiments, step (b) consists of determining the presence or absence and/or amounts of seven test microRNAs of: hsa-miR-29a-3p (SEQ ID NO: 39); hsa-miR-92a-3p (SEQ ID NO: 47); hsa-miR-200c-3p (SEQ ID NO: 105); hsa-miR-320c (SEQ ID NO: 115); hsa-miR-335-5p (SEQ ID NO: 122); hsa-miR-450b-5p (SEQ ID NO: 149); and hsa-miR-1307-5p (SEQ ID NO: 182) in the sample.

In some embodiments, step (b) of a method disclosed herein further comprises determining the presence or absence and/or amounts of one or more (e.g., two or more, three or more, four or more, five or more, or six or more) additional test microRNAs selected from the group consisting of: hsa-miR-23b-3p (SEQ ID NO: 29); hsa-miR-32-5p (SEQ ID NO: 46); hsa-miR-150-5p (SEQ ID NO: 83); hsa-miR-200a-3p (SEQ ID NO: 104); hsa-miR-203a (SEQ ID NO: 106); hsa-miR-320d (SEQ ID NO: 116); and hsa-miR-1246 (SEQ ID NO: 178) in the sample. In some

embodiments, step (b) of a method disclosed herein further comprises determining the presence or absence and/or amounts of seven additional test microRNAs of: hsa-miR-23b-3p (SEQ ID NO: 29); hsa-miR-32-5p (SEQ ID NO: 46); hsa-miR-150-5p (SEQ ID NO: 83); hsa-miR-200a-3p (SEQ ID NO: 104); hsa-miR-203a (SEQ ID NO: 106);
 5 hsa-miR-320d (SEQ ID NO: 116); and hsa-miR-1246 (SEQ ID NO: 178) in the sample. In some embodiments, the presence or absence and/or amounts of no more than 2,500 (e.g., no more than 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,500, or 2,000) test microRNAs are determined.

10 In some embodiments, step (b) comprises determining the presence or absence and/or amounts of fourteen test microRNAs of hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-23b-3p (SEQ ID
 15 NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR-203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR-1246 (SEQ ID NO: 178) in the sample. In some embodiments, the presence or absence and/or amounts of no more than 2,500 (e.g., no more than 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190,
 20 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,500, or 2,000) test microRNAs are determined. In some embodiments, step (b) consists of determining the presence or absence and/or amounts of fourteen test microRNAs of hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-
 25 450b-5p (SEQ ID NO: 149), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR-203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR-1246 (SEQ ID NO: 178) in the sample.

In some embodiments, step (b) of a method disclosed herein comprises
 30 determining the presence or absence and/or amounts of the fourteen test microRNAs of: hsa-miR-29a-3p (SEQ ID NO: 39); hsa-miR-92a-3p (SEQ ID NO: 47); hsa-miR-200c-3p (SEQ ID NO: 105); hsa-miR-320c (SEQ ID NO: 115); hsa-miR-335-5p (SEQ ID NO: 122); hsa-miR-450b-5p (SEQ ID NO: 149); hsa-miR-1307-5p (SEQ ID

NO: 182); hsa-miR-23b-3p (SEQ ID NO: 29); hsa-miR-32-5p (SEQ ID NO: 46); hsa-miR-150-5p (SEQ ID NO: 83); hsa-miR-200a-3p (SEQ ID NO: 104); hsa-miR-203a (SEQ ID NO: 106); hsa-miR-320d (SEQ ID NO: 116); and hsa-miR-1246 (SEQ ID NO: 178) in the sample. In some embodiments, the presence or absence and/or
 5 amounts of only these fourteen test microRNAs are determined.

In some embodiments, the method disclosed herein further comprises the step of: (d) determining the presence or absence and/or amounts of one or more (e.g., two, three, four, two or more, three or more, four or more) normalizing microRNAs in the sample. In some embodiments, the method disclosed herein further comprises the
 10 step of (e) normalizing the determined presence or absence and/or amounts of the test microRNAs using the presence or absence and/or amounts of the normalizing microRNAs determined in step (d). In some embodiments, the normalizing microRNAs are selected from the group consisting of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and
 15 hsa-miR-191-5p (SEQ ID NO: 196). In some embodiments, step (d) comprises determining the presence or absence and/or amounts of four normalizing microRNAs of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196) in the sample. In some embodiments, step (d) consists of determining the presence or absence and/or
 20 amounts of four normalizing microRNAs of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196) in the sample. In some embodiments, step (d) comprises determining the presence or absence and/or amounts of two normalizing microRNAs of hsa-miR-103a-3p (SEQ ID NO: 55) and hsa-miR-423-3p (SEQ ID NO: 142) in the
 25 sample. In some embodiments, step (d) consists of determining the presence or absence and/or amounts of two normalizing microRNAs of hsa-miR-103a-3p (SEQ ID NO: 55) and hsa-miR-423-3p (SEQ ID NO: 142) in the sample. In some embodiments, the presence or absence and/or amounts of no more than 2,500 (e.g., no more than 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190,
 30 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,500, or 2,000) normalizing microRNAs are determined. In some embodiments, the presence or absence and/or amounts of no normalizing microRNAs are determined in step (d).

The determined presence or absence and/or amounts of the normalizing microRNAs can be used to normalize the determined presence or absence and/or amounts of any test microRNAs. In some embodiments, the determined presence or absence and/or amounts of the normalizing microRNAs are used to normalize the

5 determined presence or absence and/or amounts of the seven microRNAs of hsa-miR-29a-3p (SEQ ID NO: 39); hsa-miR-92a-3p (SEQ ID NO: 47); hsa-miR-200c-3p (SEQ ID NO: 105); hsa-miR-320c (SEQ ID NO: 115); hsa-miR-335-5p (SEQ ID NO: 122); hsa-miR-450b-5p (SEQ ID NO: 149); and hsa-miR-1307-5p (SEQ ID NO: 182). In some embodiments, steps (b) and (d) in a method disclosed herein comprises

10 determining the presence or absence and/or amounts of these seven test microRNAs and the four normalizing microRNAs of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196). In some embodiments, steps (b) and (d) in a method disclosed herein consists of determining the presence or absence and/or amounts of

15 these seven test microRNAs and the four normalizing microRNAs of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196). In some embodiments, steps (b) and (d) in a method disclosed herein comprises determining the presence or absence and/or amounts of these seven test microRNAs and the two normalizing microRNAs

20 of hsa-miR-423-3p (SEQ ID NO: 142) and hsa-miR-103a-3p (SEQ ID NO: 55). In some embodiments, steps (b) and (d) in a method disclosed herein consists of determining the presence or absence and/or amounts of these seven test microRNAs and the two normalizing microRNAs of hsa-miR-423-3p (SEQ ID NO: 142) and hsa-miR-103a-3p (SEQ ID NO: 55).

25 In some embodiments, the determined presence or absence and/or amounts of the normalizing microRNAs are used to normalize the determined presence or absence and/or amounts of the fourteen test microRNAs of: hsa-miR-29a-3p (SEQ ID NO: 39); hsa-miR-92a-3p (SEQ ID NO: 47); hsa-miR-200c-3p (SEQ ID NO: 105); hsa-miR-320c (SEQ ID NO: 115); hsa-miR-335-5p (SEQ ID NO: 122); hsa-miR-450b-5p (SEQ ID NO: 149); hsa-miR-1307-5p (SEQ ID NO: 182); hsa-miR-23b-3p (SEQ ID NO: 29); hsa-miR-32-5p (SEQ ID NO: 46); hsa-miR-150-5p (SEQ ID NO: 83); hsa-miR-200a-3p (SEQ ID NO: 104); hsa-miR-203a (SEQ ID NO: 106); hsa-miR-320d (SEQ ID NO: 116); and hsa-miR-1246 (SEQ ID NO: 178). In some

30

embodiments, steps (b) and (d) in a method disclosed herein comprises determining the presence or absence and/or amounts of these fourteen test microRNAs and the two normalizing microRNAs of hsa-miR-423-3p (SEQ ID NO: 142) and hsa-miR-103a-3p (SEQ ID NO: 55). In some embodiments, steps (b) and (d) in a method disclosed

5 herein consists of determining the presence or absence and/or amounts of these fourteen test microRNAs and the two normalizing microRNAs of hsa-miR-423-3p (SEQ ID NO: 142) and hsa-miR-103a-3p (SEQ ID NO: 55). In some embodiments, steps (b) and (d) in a method disclosed herein comprises determining the presence or absence and/or amounts of these fourteen test microRNAs and the four normalizing

10 microRNAs of: hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196). In some embodiments, steps (b) and (d) in a method disclosed herein consists of determining the presence or absence and/or amounts of these fourteen test microRNAs and the four normalizing microRNAs of: hsa-miR-103a-3p (SEQ ID NO:

15 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196). With the amounts of these fourteen microRNAs, the amounts of normalizing microRNAs are dispensable. Accordingly, in some embodiments, steps (b) and (d) in a method disclosed herein consists of determining the presence or absence and/or amounts of these fourteen test microRNAs and no

20 normalizing microRNAs.

In another aspect, the instant disclosure provides a method for determining the presence or absence and/or amounts of microRNAs in a blood sample from a subject (e.g., human subject), the method comprising: (a) obtaining a blood sample collected from the subject; and (b) determining the presence or absence and/or amounts of one

25 or more of the microRNAs listed in Table 2 in the sample, thereby determining the presence or absence and/or amounts of the microRNAs in the sample. In some embodiments, step (b) comprises determining the presence or absence and/or amounts of at least 1 (e.g., at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150,

30 160, 170, 180, or 190) of the microRNAs listed in Table 2 in the sample. In some embodiments, step (b) comprises determining the presence or absence and/or amounts of no more than 190 (e.g., no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110,

120, 130, 140, 150, 160, 170, or 180) of the microRNAs listed in Table 2 in the sample. In some embodiments, step (b) consists of determining the presence or absence and/or amounts of the 192 microRNAs listed in Table 2 in the sample.

Any one of the foregoing methods (e.g., methods for diagnosing ovarian cancer, or methods for determining the presence or absence and/or amounts of microRNAs) can be used to diagnose ovarian cancer at any stage, and is particularly advantageous in the ability to diagnose early-stage ovarian cancer. Accordingly, in some embodiments, the ovarian cancer diagnosed is a stage I (e.g., stage IA, stage IB, stage IC, stage IC1, stage IC2, or stage IC3) ovarian cancer. In some embodiments, the ovarian cancer diagnosed is a stage II (e.g., stage IIA or stage IIB) ovarian cancer. In some embodiments, the ovarian cancer diagnosed is a stage III (e.g., stage IIIA, IIIA1(i), IIIA1(ii), IIIA2, IIIB, or IIIC) ovarian cancer. In some embodiments, the ovarian cancer diagnosed is a stage IV (e.g., stage IVA or stage IVB) ovarian cancer. In some embodiments, the ovarian cancer diagnosed of the female subject is an invasive serous adenocarcinoma, an invasive clear cell adenocarcinoma, or an invasive endometrioid adenocarcinoma.

In another aspect, the instant disclosure provides a method for treating a female subject suspected of having ovarian cancer, the method comprising the steps of any one of the foregoing methods (e.g., methods for diagnosing ovarian cancer, or methods for determining the presence or absence and/or amounts of microRNAs); and treating the subject for ovarian cancer or monitoring the subject for ovarian cancer, based on the presence or absence and/or amounts of microRNAs determined in steps (b) and/or (d), and/or the comparison of step (c). In some embodiments, treating the subject is performed using one or more treatments selected from the group consisting of: surgery, chemotherapy, targeted antibody therapy, radiation therapy, and hormone therapy, and stem cell transplant.

In yet another aspect, the disclosure provides methods for screening a female subject for the presence or absence and/or amounts of microRNAs, the method comprising the steps of any one of the foregoing methods (e.g., methods for diagnosing ovarian cancer, or methods for determining the presence or absence and/or amounts of microRNAs). In some embodiments, the screening method further comprises the step of determining the presence or absence and/or amounts of the same microRNAs in a second sample. In some embodiments, the second sample is a

second sample from the same subject, a sample of the same tissue or organ from a different subject known to have ovarian cancer, or a sample of the same tissue or organ from a different subject known not to have ovarian cancer.

In some embodiments of any one of the preceding aspects, step (b) and/or step
5 (d) are performed by detecting binding of the sample to at least one probe capable of specifically hybridizing to each of the microRNAs or a complementary DNA (cDNA) thereof. In some embodiments, step (b) and/or step (d) are performed by detecting binding of the sample to at least two probes capable of specifically hybridizing to each of the microRNAs or a cDNA thereof. In some embodiments, step (b) and/or
10 step (d) are performed using a nucleic acid detection assay, e.g., microarray, reverse-transcription (RT)-PCR, and reverse-transcription quantitative PCR (RT-qPCR). In some embodiments, at least one of the probes comprises a detectable label. The detectable label can be a non-natural label that permits detection of a microRNA in a sample, such as fluorescent label, a fluorescence donor, a fluorescence quencher, or
15 any other non-natural molecule that may be used to detect the nucleic acid polymer or detect the hybridization of the nucleic acid polymer with the microRNA. In some embodiments, each one of the probes comprises a detectable label. In some embodiments, step (b) and/or step (d) are performed by reverse-transcribing the microRNA molecules in the sample, thereby obtain a cDNA sample; and sequencing
20 the cDNA sample. In some embodiments, step (b) and/or step (d) further comprises amplifying the DNA molecules in the cDNA sample before sequencing the cDNA sample. In some embodiments, the sequencing is performed by next-generation sequencing.

In some embodiments of any one of the preceding aspects, the sample is
25 selected from a sample of blood, lymph node, or ovary. In some embodiments, the sample is a blood sample, e.g., plasma, serum, or whole blood. In some embodiments, the sample is a sample of lymph node, e.g., a draining lymph node for an ovary. In some embodiments, the sample is an ovary sample, e.g., an ovary suspected to contain a tumor.

30 In some embodiments of any one of the preceding aspects, the female subject is at the risk of developing ovarian cancer. Certain factors may increase a subject's risk of ovarian cancer. For example, ovarian cancer rates are highest in women aged 55-64 years. Half of all ovarian cancers are found in women 63 years of age or older

(the median age of diagnosis). Family history of ovarian cancer, colorectal cancer (such as mutations in the gene MUTYH), and/or breast cancer also increases the risk of ovarian cancer in a subject. Mutations in the genes BRCA1 and BRCA2 are associated with increased risk of ovarian cancer. Increased ovarian cancer risk is also
5 seen in Lynch syndrome, characterized by germline mutations in DNA mismatch repair (MMR) genes, including mutL homolog 1 (MLH1), mutS homolog 2 (MSH2), MSH3, MSH6, postmeiotic segregation increased 1 (PMS1), and PMS2 (see Nakamura et al., Mol Clin Oncol. 2014 Nov; 2(6): 909–916). Accumulated mutations in PTEN also increase the risk of ovarian cancer (see Merritt and Cramer, Cancer
10 Biomark. 2010;9(1-6):287-305.). Female subjects with Peutz-Jeghers syndrome, an inherited genetic variant of the gene STK11, also have an increases risk of ovarian cancer. Accordingly, in some embodiments, the subject has any one of the foregoing risk factors for ovarian cancer. In some embodiments, the subject has one or more ovaries. In some embodiments, the subject has at least one mutation in the BRCA1 or
15 BRCA2 gene. In some embodiments, the subject has at least one mutation in the MMR genes. In some embodiments, the subject is at least 55 years (e.g., at least 56, 57, 58, 59, 60, 61, 62, 63, 64, or 65 years) of age. In some embodiments, the subject is no more than 65 years (e.g., no more than 55, 56, 57, 58, 59, 60, 61, 62, 63, or 64 years) of age.

20 In some embodiments, the subject has been suspected of having ovarian cancer. In some embodiments, the subject has been diagnosed of having ovarian cancer, and the method disclosed herein is used to at least confirm the diagnosis. Methods of diagnosing ovarian cancer include without limitation physical examination, ultrasound imaging, computed tomography scan, magnetic resonance
25 imaging (MRI) scan, positron emission tomography (PET) scan, and blood tests. In some embodiments, the blood test is determines the level of CA-125, wherein the subject has a higher level of CA-125 antigen in the blood than a diagnostic threshold level (e.g., 35 U/mL).

The statistical model used in the methods disclosed herein can be any
30 statistical models known in the art. In some embodiments, the statistical model comprises one or more models selected from the group consisting of linear discriminant analysis, logistic regression, multivariate adaptive regression splines, naive Bayes, neural network, support vector machine, functional tree, LAD tree,

Bayesian network, elastic net regression, and random forest. In some embodiments, the statistical model comprises a neural network, optionally further comprising one or more other statistical models. In some embodiments, the statistical model consists of a neural network.

5 In another aspect, the disclosure provides a kit comprising at least one test probe capable of specifically hybridizing to a microRNAs selected from the group consisting of: hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-320c (SEQ ID NO: 115), and hsa-miR-200c-3p (SEQ ID NO: 105), or a cDNA thereof. In some embodiments, the kit
10 comprises at least one test probe (e.g., at least two test probes) capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-320c (SEQ ID NO: 115), and hsa-miR-200c-3p (SEQ ID NO: 105), or a cDNA thereof. In some embodiments, the test probes consist of seven
15 probes to test microRNAs with one test probe capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-320c (SEQ ID NO: 115), and hsa-miR-200c-3p (SEQ ID NO: 105), or a cDNA thereof. In some embodiments, the test probes consist of fourteen probes to test
20 microRNAs with two test probes capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-320c (SEQ ID NO: 115), and hsa-miR-200c-3p (SEQ ID NO: 105), or a cDNA thereof.

In some embodiments of the kits disclosed herein, the kit further comprises at least one additional test probe capable of hybridizing to a microRNA selected from
30 the group consisting of: hsa-miR-23b-3p (SEQ ID NO: 29); hsa-miR-32-5p (SEQ ID NO: 46); hsa-miR-150-5p (SEQ ID NO: 83); hsa-miR-200a-3p (SEQ ID NO: 104); hsa-miR-203a (SEQ ID NO: 106); hsa-miR-320d (SEQ ID NO: 116); and hsa-miR-1246 (SEQ ID NO: 178), or a cDNA thereof. In some embodiments, the kit further

comprises at least one test probe (e.g., at least two test probes) capable of hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR1246 (SEQ ID NO: 178), or a cDNA thereof.

In some embodiments, the kit comprises at least one test probe (e.g., at least two test probes) capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR1246 (SEQ ID NO: 178), or a cDNA thereof. In some embodiments, the test probes consist of fourteen probes to test microRNAs with one test probe capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR1246 (SEQ ID NO: 178), or a cDNA thereof. In some embodiments, the test probes consist of 28 probes to test microRNAs with two test probes capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR1246 (SEQ ID NO: 178), or a cDNA thereof. In some embodiments, the kit comprises no normalizing probes.

In some embodiments, the kit further comprises at least one normalizing probe capable of hybridizing to a microRNA selected from the group consisting of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196), or a cDNA thereof. In some embodiments, the kit further comprises at least one normalizing probe (e.g., at least two normalizing probes) capable of hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196), or a cDNA thereof. In some embodiments, the kit further comprises at least one normalizing probe (e.g., at least two normalizing probes) capable of hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-423-3p (SEQ ID NO: 142) and hsa-miR-103a-3p (SEQ ID NO: 55). In some embodiments, the normalizing probes consist of four probes to normalizing microRNAs with one normalizing probe capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196), or a cDNA thereof. In some embodiments, the normalizing probes consist of eight probes to normalizing microRNAs with two test probes capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196), or a cDNA thereof. In some embodiments, the kit comprises no normalizing probes.

In some embodiments, the kit comprises at least one test probe (e.g., at least two test probes) capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-320c (SEQ ID NO: 115), and hsa-miR-200c-3p (SEQ ID NO: 105), or a cDNA thereof, and at least one normalizing probe (e.g., at least two normalizing probes) capable of hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196), or a cDNA thereof. In some

embodiments, the probes consist of eleven probes to microRNAs with one probe capable of specifically hybridizing to each of the microRNAs or a cDNA thereof. In some embodiments, the test probes consist of 22 probes to microRNAs with two probes capable of specifically hybridizing to each of the microRNAs or a cDNA thereof.

In some embodiments of any of the kits disclosed herein, at least one of the probes comprises a detectable label. The detectable label can be a non-natural label that permits detection of a microRNA in a sample, such as fluorescent label, a fluorescence donor, a fluorescence quencher, or any other non-natural molecule that may be used to detect the nucleic acid polymer or detect the hybridization of the nucleic acid polymer with the microRNA. In some embodiments, each one of the probes comprises a detectable label.

In certain embodiment, the kit further comprises one or more reagents suitable for the determination of the presence or absence and/or amounts of microRNAs (e.g., buffers) and instructions for use. The kit may also include reagents for amplification of a nucleic acid sequence, such as a DNA sequence or an RNA sequence. In some embodiments, such reagents may include: a heat-stable DNA polymerase; deoxynucleotide triphosphates (dNTPs); a buffer solution to provide a suitable chemical environment for optimum activity, binding kinetics, and stability of the DNA polymerase; bivalent cations such as magnesium or manganese ions; and/or monovalent cations, such as potassium ions. In some embodiments, such reagents may include: a reverse transcriptase; an RNase inhibitor; a primer that hybridizes to a nucleic acid sequence (such as RNA or DNA); a primer that hybridizes to an adenosine oligonucleotide; and a buffer solution that provides a suitable chemical environment for optimum activity, binding kinetics, and stability of the reverse transcriptase. In some embodiments, the kit further comprises standard samples of one or more of the microRNAs capable of hybridizing to the probes. The reagents for amplification of the nucleic acid sequence may be provided in the form of a solution, a concentrated solution, or a powder. In some embodiments, the kit further comprises one or more reagents suitable of extracting and/or purifying microRNA molecules from a samples (e.g., blood sample).

In certain embodiments, kits of the instant disclosure may also include instructions describing the use of the kit for determining the presence or absence

and/or amounts of the test microRNAs in a sample (e.g., blood sample). In some embodiments, the kit further comprises access to a computer program to make diagnosis based on the presence or absence and/or amounts of the test microRNAs and/or normalizing microRNAs in a sample. In certain embodiments, the instructions
5 are provided as an insert sheet. In certain embodiments, the instructions are provided as a computer-readable form carried on a device or transmitted or obtainable from a location on the Internet.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A shows a flow chart of the protocol for miRNA sequencing, filtering,
10 batch adjustment and separation into the training and testing sets.

FIG. 1B shows a flow chart of the model development and testing.

FIG. 2A is a graph depicting the performance of various disease classification models created on the subset of miRNA biomarkers selected using a significance-based filter. The solid dark gray bars denote the sensitivity of the models, the dotted
15 light gray bars denote the specificity of the classification models; the error bars denote 95% confidence intervals.

FIG. 2B is a graph depicting the performance of various machine learning models created on the subset of miRNA biomarkers selected using a correlation based feature selection (CFS). The solid dark gray bars denote the sensitivity of the models,
20 the dotted light gray bars denote the specificity of the classification models; the error bars denote 95% confidence intervals.

FIG. 2C shows a graph depicting the performance of various machine learning models created on a subset of miRNA biomarkers by a fold change-based filter. The solid dark gray bars denote the sensitivity of the models, the dotted light
25 gray bars denote the specificity of the classification models; the error bars denote 95% confidence intervals. The arrow indicates the neural network analysis using fold change-based filter variables (fourteen miRNA biomarkers).

FIG. 3A is a graph of receiver operating characteristic (ROC) curves for a neural network analysis using miRNA biomarkers selected using a fold change-based
30 filter. The ROC curves of FIG. 3A compares the performance of the neural network using fourteen miRNA biomarkers on a training set of raw, non-batch-adjusted data

(dotted line) with the performance of the neural network using fourteen miRNA biomarkers on a training set of batch-adjusted data (solid line). The diagonal dashed line represents a classifier without diagnostic utility.

FIG. 3B is a graph of ROC curves for a neural network analysis using fourteen miRNA biomarkers. The ROC curves of FIG. 3B compare the performance of the neural network and fourteen miRNA biomarkers on a testing set of raw non-batch adjusted data (dotted line) with the performance of the neural network and fourteen miRNA biomarkers on a testing set of batch-adjusted data (solid line). The diagonal dashed line represents a classifier without diagnostic utility.

FIG. 4A is a graph of ROC curves comparing the performance of the neural network using a set of fourteen biomarkers (solid line) with the performance of the CA-125 biomarker (dotted line), based on 120 patients. The diagonal line represents a classifier without diagnostic utility.

FIG. 4B is a graph of ROC curves comparing the performance of the neural network in classifying ovarian cancer (solid line) with the performance of the CA-125 biomarker (dotted line). The neural network used a set of fourteen biomarkers and a training set (N=94). The diagonal line represents a classifier without diagnostic utility. The neural network using the fourteen miRNA biomarkers has an area under the curve (AUC) of 0.92 (95%CI; 0.87-0.98), while the CA-125 biomarker has an AUC of 0.78 (95%CI; 0.62-0.84).

FIG. 4C shows a graph of ROC curves comparing the performance of the neural network in classifying ovarian cancer using a set of fourteen biomarkers (solid line) with the performance of the CA-125 biomarker (dotted line). The neural network used a testing set (N=26). The diagonal line represents a classifier without diagnostic utility. The neural network using the fourteen miRNA biomarkers has an AUC of 0.94 (95%CI; 0.89-0.99), while the CA-125 biomarker has an AUC of 0.78 (95%CI; 0.62-0.84), with $p=0.12$.

FIG. 5A is a graph of ROC curves for neural network analysis of qPCR data using a fourteen miRNA signature with minimal values imputed for undetectable miRNAs.

FIG. 5B is a graph of ROC curves for neural network analysis of qPCR data built on a seven miRNA signature after a global sensitivity analysis.

FIG. 6 is a ROC curve for neural network analysis using qPCR inputs from an different clinical test set gathered in Poland.

FIG. 7A is a line graph of change in mir-200c-3p expression in blood samples from patients both before and after surgical cytoreduction.

5 **FIG. 7B** is a line graph of change in mir-200a-3p expression in blood samples from patients both before and after surgical cytoreduction.

FIG. 7C is a line graph of change in mir-92a-3p expression in blood samples from patients both before and after surgical cytoreduction.

10 **FIG. 7D** is a line graph of change in mir-29a-3p expression in blood samples from patients both before and after surgical cytoreduction.

FIG. 8 is a series of photomicrographs showing *in situ* expression of selected miRNAs from a serum signature in sections of fallopian tubes showing serous tubal intraepithelial carcinoma (STIC) lesions and Stage I high grade serous ovarian cancer (HGSOC). Lesional cells are indicated by TP53 and Ki-67 staining.

15 **FIG. 9** is a graph of an ROC curve showing that in a cohort of 454 patient samples with a wide array of medical diagnoses, including 70 healthy controls, a fourteen miRNA signature accurately identifies ovarian cancer cases against all other diagnoses (i.e. healthy controls or other cancers). The diagonal line represents a classifier without diagnostic utility.

20 **FIG. 10A** is a graph of a ROC curve of a fourteen miRNA signature for ovarian cancer compared with pancreatic ductal cancer in a certain dataset. The diagonal line represents a classifier without diagnostic utility.

25 **FIG. 10B** is a graph of a ROC curve of a fourteen miRNA signature for ovarian cancer compared with other prostate cancer in a certain dataset. The diagonal line represents a classifier without diagnostic utility.

FIG. 10C is a graph of a ROC curve of a fourteen miRNA signature for ovarian cancer compared with stomach cancer in a certain dataset. The diagonal line represents a classifier without diagnostic utility.

30 **FIG. 10D** is a graph of a ROC curve of a fourteen miRNA signature for ovarian cancer compared with pancreatic cancer in a certain dataset. The diagonal line represents a classifier without diagnostic utility.

FIG. 10E is a graph of a ROC curve of a fourteen miRNA signature for ovarian cancer compared with melanoma in a certain dataset. The diagonal line represents a classifier without diagnostic utility.

5 **FIG. 10F** is a graph of a ROC curve of a fourteen miRNA signature for ovarian cancer compared with lung cancer in a certain dataset. The diagonal line represents a classifier without diagnostic utility.

FIG. 10G is a graph of a ROC curve of a fourteen miRNA signature for ovarian cancer compared with periodontitis in a certain dataset. The diagonal line represents a classifier without diagnostic utility.

10 **FIG. 10H** is a graph of a ROC curve of a fourteen miRNA signature for ovarian cancer compared with pancreatitis in a certain dataset. The diagonal line represents a classifier without diagnostic utility.

FIG. 10I is a graph of a ROC curve of a fourteen miRNA signature for ovarian cancer compared with multiple sclerosis in a certain dataset. The diagonal
15 line represents a classifier without diagnostic utility.

FIG. 10J is a graph of a ROC curve of a fourteen miRNA signature for ovarian cancer compared with acute myocardial infarction (MI) in a certain dataset. The diagonal line represents a classifier without diagnostic utility.

FIG. 10K is a graph of a ROC curve of a fourteen miRNA signature for
20 ovarian cancer compared with chronic obstructive pulmonary disease (COPD) in a certain dataset. The diagonal line represents a classifier without diagnostic utility.

FIG. 10L is a graph of a ROC curve of a fourteen miRNA signature for ovarian cancer compared with sarcoidosis in a certain dataset. The diagonal represents a classifier without diagnostic utility.

25 **DETAILED DESCRIPTION**

It is to be understood that the methods described in this disclosure are not limited to particular methods and experimental conditions disclosed herein; as such methods and conditions may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not
30 intended to be limiting.

Furthermore, the experiments described herein, unless otherwise indicated, use conventional molecular and cellular biological and immunological techniques within the skill of the art. Such techniques are well known to the skilled worker, and are explained fully in the literature. See, e.g., Ausubel, *et al.*, ed., Current Protocols in Molecular Biology, John Wiley & Sons, Inc., NY, N.Y. (1987-2008), including all supplements, Molecular Cloning: A Laboratory Manual (Fourth Edition) by MR Green and J. Sambrook and Harlow *et al.*, Antibodies: A Laboratory Manual, Chapter 14, Cold Spring Harbor Laboratory, Cold Spring Harbor (2013, 2nd edition).

Unless otherwise defined herein, scientific and technical terms used herein have the meanings that are commonly understood by those of ordinary skill in the art. In the event of any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic definition. Unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. The use of “or” means “and/or” unless stated otherwise. The use of the term “including”, as well as other forms, such as “includes” and “included”, is not limiting.

Generally, nomenclatures used in connection with cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques provided herein are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. Enzymatic reactions and purification techniques are performed according to manufacturer’s specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

That the disclosure may be more readily understood, select terms are defined below.

As used herein, the term “nucleic acid” refers to a polymer of two or more nucleotides or nucleotide analogues (such as ribonucleic acid having methylene bridge between the 2'-O and 4'-C atoms of the ribose ring) capable of hybridizing to a complementary nucleic acid. As used herein, this term includes, without limitation,
5 DNA, RNA, LNA, and PNA.

The term “microRNAs” or “miRNAs” as used herein, refers to small noncoding ribonucleic acid (RNA) gene products between 19 and 26 nucleotides long that form a hairpin secondary structure. MicroRNAs described herein are named using the nomenclature set forth in Ambros et al., RNA. 2003 Mar;9(3):277-9,
10 incorporated herein by reference, and sequences may be found at mirbase.org.

The term “subject”, as used herein, refers to a mammal, e.g. a human, a domestic animal or a livestock including a cat, a dog, a cattle and a horse.

As used herein, the phrase “determining the presence or absence” refers to assessing whether an analyte, such as a microRNA is detectable or undetectable in a
15 biological sample (e.g. a blood sample) using one or more detection techniques for detecting the analyte (such as qPCR, microarray detection, etc.). An analyte that is detected in a biological sample using a detection technique is considered “present”. An analyte that is not detected in a biological sample using a detection technique is considered “absent”.

20 As used herein, the term “bind” or “binding” refers to non-covalent or covalent interaction between two molecules, such as between two complementary nucleic acids.

As used herein, the term “specifically hybridizing” refers to non-covalent interaction between a first nucleic acid molecule (e.g. a nucleic acid probe having a
25 certain nucleotide sequence) and a second nucleic acid molecule (e.g. a microRNA having a nucleotide sequence complementary to that of the nucleic acid probe). Hybridization conditions have been described in the art and are known to one of skill in the art. In some embodiments, the condition for detecting the hybridization is a suitable condition of a nucleic acid detection assay (e.g., microarray, RT-PCR, or RT-
30 qPCR). The likelihood of hybridization between two nucleic acids correlates with the nucleotide sequence complementary between the two nucleic acids.

The term “hybridize” as used herein, refers to annealing of a first single-stranded nucleic acid to a second complementary single-stranded nucleic, in which complementary nucleotides of the first and second nucleic acids pair by hydrogen bonding.

5 The phrase “detecting binding of a probe”, as used herein, refers to use of a detection method allowing determination that a probe (e.g. a nucleic acid probe) has non-covalently or covalently interacted with a target molecule (e.g. a target nucleic acid in a sample). For example, detecting binding of probe in qPCR may include optical detection of fluorescence of a self-quenching probe following binding to the
10 complementary sequence of a target nucleic acid in the sample. In some embodiments, detecting binding of a probe may include detection of a nucleic acid intercalating agent to detect amplified double-stranded nucleic acid, such as a fluorescent intercalating agent used in qPCR.

 As used herein, the term “probe” refers to a molecule or complex that is used
15 to determine the presence or absence and/or amount of a microRNA in a sample (e.g. a blood sample). In certain embodiments, the probe comprises a nucleic acid moiety (e.g., DNA, modified DNA, or modified RNA) that is capable of specifically hybridizing to the microRNA or a complementary DNA (cDNA) thereof. In certain
20 embodiments, the probe comprises a sequence of at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 contiguous nucleotides identical or complementary to the microRNA. In certain embodiments, a probe further comprises a detectable label that is conjugated, covalently or non-covalently, to the nucleic acid moiety. Exemplary
25 detectable labels include without limitation a fluorophore, a small molecule (e.g., a small molecule of the avidin family), an enzyme, an antibody or antibody fragment, or a nucleic acid sequence not present in the subject in a form that is linked to the microRNA (e.g., a barcode sequence). Accordingly, the probe may be a fluorophore-labeled nucleic acid having a nucleotide sequence that is complementary to a nucleotide sequence of a microRNA.

 The term “PCR”, as used herein, refers to polymerase chain reaction for
30 amplifying an amount of target DNA. PCR relies on thermal cycling, which consists of cycles of repeated heating and cooling of a reaction for DNA denaturation, annealing and enzymatic elongation of the amplified DNA. First, the strands of the DNA are separated at a high temperature in a process called DNA melting or

denaturing. Next, the temperature is lowered, allowing the primers and the strands of target DNA to selectively bind or anneal, creating templates for DNA polymerase to amplify the target DNA. Next, at a working temperature of the DNA polymerase, template-dependent DNA synthesis occurs. These steps are repeated to create many
5 copies of the target DNA.

A “primer”, as used herein, refers to a short, single-stranded DNA sequence that selectively binds to a target DNA sequence and enables addition of new deoxyribonucleotides by DNA polymerase at the 3' end. According to certain embodiments, the forward primer is 18-35, 19-32 or 21-31 nucleotides in length. The
10 nucleotide sequence of the forward primer is not limited, so long as it specifically hybridizes with part of or an entire target site, and its T_m value may be within a range of 50 °C to 72 °C, in particular may be within a range of 58 °C to 61 °C, and may be within a range of 59 °C to 60 °C. The nucleotide sequence of the primer may be manually designed to confirm the T_m value using a primer T_m prediction tool.
15 Primer nucleotides may include nucleotide analogues and/or modified nucleotides, such as LNA or PNA.

As used herein, the term “RT-PCR” refers to reverse transcription polymerase chain reaction, a process for amplifying RNA. RNA molecules are reverse transcribed to complementary DNA (cDNA) using reverse transcriptase and then
20 using PCR to amplify the resulting cDNA.

As used herein, the term “RT-qPCR” refers to reverse transcription quantitative polymerase chain reaction, a variant of RT-PCR in which amplification of cDNA during the RT-PCR process is quantitatively detected in real time using a probe that detects amplified target DNA. For example, in some embodiments, self-
25 quenching nucleic acid probes are added to the reaction mixture. The self-quenching nucleic acid probes only fluoresce when they bind a target sequence. As each cycle of PCR is completed, the self-quenching probes bind to the amplified DNA, unquench and fluoresce with exposure to a light excitation source. As DNA is amplified, increased probe and target binding results in increased fluorescence of the self-
30 quenching nucleic acid probe. Detection of the fluorescing probes after each amplification cycle allows real-time measurement of the amplification process, as increasing amounts of the nucleic acid probe bind with amplified target DNA and fluoresce. In some embodiments, an intercalating dye probe is added to the reaction

mixture that fluoresces upon interaction with double-stranded nucleic acids. The increase in dye fluorescence during the amplification process allows the measurement of DNA amplification in real-time, as increasing amounts of the dye probe intercalate with the increasing amounts of target DNA being amplified.

5 As used herein, the term “normalize” or “normalizing” refers to adjusting a first measured value (e.g., level of a gene of interest) relative to a second measured value (e.g., level of a housekeeping gene), wherein the first and second measured values are measured from the same sample (e.g., different portions of the same homogenous sample), and wherein the second measured value is correlated to the
10 quantity and/or quality of the sample. Normalization allows obtaining a relative amount of the first value that is not affected by the quantity and/or quality of the sample that may vary from individual sample preparation.

 As used herein, the term “normalizing microRNA” refers to a microRNA that is known to have a stable amount in a sample (e.g. a blood sample) and is used to
15 normalize the measured value of a test microRNA in the sample. A single normalizing microRNA may be used to normalize the measured amount of a target microRNA in a sample, or an averaged value of multiple microRNAs may be used for normalization. In certain embodiments, normalization may be calculated by: Number of amplification cycles (average of the normalizer microRNA) – number of
20 amplification cycles (miR of interest).

 As used herein, the term “test microRNA” refers to a microRNA the presence or absence and/or amount of which is determined, for example, for diagnosis purpose (e.g., using an algorithm). In some embodiments, the presence or absence and/or amount of one or more test microRNAs can be used additionally for normalization
25 purpose.

 As used herein, the term “normalizing probe” refers to a probe that is used to determine the presence or absence and/or amount of a normalizing microRNA in a sample (e.g. a blood sample). In certain embodiments, the normalizing probe comprises a nucleic acid moiety (e.g., DNA, modified DNA, or modified RNA) that is
30 capable of specifically hybridizing to a normalizing microRNA or a complementary DNA (cDNA) thereof.

As used herein, the term “test probe” refers to a probe that is used to determine the presence or absence and/or amount of a test microRNA in a sample (e.g. a blood sample). In certain embodiments, the test probe comprises a nucleic acid moiety (e.g., DNA, modified DNA, or modified RNA) that is capable of specifically hybridizing to a test microRNA or a complementary DNA (cDNA) thereof.

The phrase “a reagent for amplification of a DNA sequence” includes, but is not limited to: (1) a heat-stable DNA polymerase; (2) deoxynucleotide triphosphates (dNTPs); (3) a buffer solution, providing a suitable chemical environment for optimum activity, binding kinetics, and stability of the DNA polymerase; (4) bivalent cations such as magnesium or manganese ions; and (5) monovalent cations, such as potassium ions. The reagents may be provided in the form of a solution, a concentrated solution, or powder.

The phrase “a reagent for reverse transcription of an RNA molecule” encompasses, but is not limited to: a reverse transcriptase; an RNase inhibitor; a primer that hybridizes to a nucleic acid sequence (such as RNA or DNA); a primer that hybridizes to an adenosine oligonucleotide; and a buffer solution that provides a suitable chemical environment for optimum activity, binding kinetics, and stability of the reverse transcriptase. The reagents may be provided in the form of a solution, a concentrated solution, or powder.

As used herein, the term “blood sample” refers to an amount of blood taken from a subject, such as whole blood, or a component portion of blood taken from a subject, such as plasma, which lacks cells normally contained in whole blood (e.g. erythrocytes, leukocytes, and platelets), or serum which is plasma that lacks fibrinogen and some clotting factors.

As used herein, the term “nucleic acid detection method” encompasses any method that may be used to detect the presence of a nucleic acid, including methods of sequencing (e.g. Gilbert sequencing, Sanger sequencing, SMRT sequencing or next-generation sequencing), microarray detection, PCR, RT-PCR, real-time qPCR, real-time RT-qPCR.

As used herein, the term “next-generation sequencing” refers to high-throughput parallel sequencing of short fragments of single-stranded nucleic acids attached to slides or beads, such as techniques by ILLUMINA, ROCHE (454

sequencing), or ION TORRENT, THERMOFISHER. The incorporation of individual nucleotides onto single-stranded nucleic acids may be detected optically (via fluorescence of incorporated nucleotides) or by detection of hydrogen ions released during nucleotide incorporation (e.g. ion semiconductor sequencing).

5 As used herein, the term “microarray detection” refers to methods of detecting target nucleic acids using single-stranded nucleic acid probes attached to discrete areas of a solid surface (e.g. spots on a slide or beads in microwells). Hybridization of the probes to specific nucleic acids may be detected by a variety of methods, such as using optical detection (e.g. fluorophores, chemiluminescent molecules) or
10 radiographic detection.

 As used herein, the term “non-natural label” encompasses, without limitation, one or more labeling molecules that may be bound, attached to, or associated with a biological molecule (such as a nucleic acid, nucleotide, protein, peptide, amino acid, carbohydrate, lipid, primary/secondary metabolites, or chemical product produced by
15 a living organism) to allow detection of the molecule when associated with the biological molecule; non-natural labels are not normally associated with the biological molecule. Exemplary non-natural labels include, without limitation: antigenic tags (e.g. digoxigenin); radioisotopes (e.g. ³²P); enzymes catalyzing chemiluminescent or colorimetric chemical reactions (e.g. horseradish peroxidase or alkaline phosphatase);
20 nucleic acid dyes (e.g. Hoechst 33342, DAPI, ethidium bromide); organic fluorophores (e.g. 6-carboxyfluorescein, tetrachlorofluorescein, fluorescein, rhodamine, or cyanine); fluorophore quenchers (e.g. tetramethylrhodamine, dimethylaminoazobenzenesulfonic acid, BLACK HOLE QUENCHERS, or IOWA BLACK dyes); protein fluorophores (e.g. green fluorescent protein); donor and
25 acceptor fluorophores for fluorescence resonance energy transfer (e.g. fluorescein and tetramethylrhodamine, or NowGFP and mOrange); quantum dot fluorophores (e.g. metal chalcogenides, core shell semiconducting nanocrystals, or alloyed semiconductor quantum dots); and immune system-based molecules bound, attached to, or associated with non-natural labels described herein (e.g. antibodies or antibody
30 fragments labeled with a fluorophore or catalytic enzyme).

 As used herein, the term “artificial neural network” refers to a forecasting model based on a linked collection of neural units *in silico* that loosely model a simple mathematical model of the brain. Artificial neural networks allow

identification of complex nonlinear relationships between its response variable and its predictor variables. An artificial neural network may have one or more hidden layers that each include one or more neurons that interact to produce a prediction given two or more variables.

5 The term “ovarian cancer”, as used herein, refers to a group of malignancies affecting the ovary, that have developed from epithelial cells, sex cord-stromal cells (e.g. granulosa, theca, and hilus cells), or germ cells (e.g. oocytes). About 60% of ovarian tumors are of epithelial origin and account for 90% of ovarian cancers (see Karst et al., “Ovarian Cancer Pathogenesis: A Model in Evolution,” Journal of
10 Oncology, vol. 2010, Article ID 932371, 13 pages, 2010. doi:10.1155/2010/932371, incorporated herein by reference). Such epithelium-derived ovarian carcinomas are heterogeneous in character, with differences in tumor morphology, clinical symptoms, and genetic alterations. The World Health Organization (WHO) lists eight different tumor histologies, including serous, endometrioid, mucinous, clear cell, transitional
15 cell, squamous cell, mixed epithelial, and undifferentiated. Tumors of each of these subtypes may be classified as benign (having low malignant potential and/or indolence), malignant, or borderline, as well as low-grade (Type I) or high-grade (Type II).

 Type I cancers (such as endometrioid and mucinous carcinomas) are of low
20 histological grade, and exhibit low-grade nuclei with infrequent mitotic figures (see Vang et al., Adv Anat Pathol. 2009 Sep; 16(5): 267–282, incorporated herein by reference). Type I ovarian serous cancers often have mutations of the KRAS, BRAF or ERBB2 genes, and lack TP53 mutations. Type I tumors are usually indolent and have better patient outcomes, although they can develop into Type II malignancies.
25 In contrast, Type II ovarian serous cancers (such as serous carcinoma) are considered a higher histological grade, having high-grade nuclei and numerous mitotic figures, and are characterized by rapid tumor development. Genetically, Type II ovarian serous tumors usually include TP53 mutation and lack mutations of KRAS, BRAF, or ERBB2.

30 Ovarian tumors may also be surgically staged to indicate malignancy. Stage I indicates confinement to the ovary. Stage IA indicates that the cancer is completely inside one ovary, and stage IB indicates that the cancer is completely inside both ovaries. Stage IC indicates that the cancer is in both ovaries and one of the following:

some cancer is on the surface of an ovary; cancer cells are present in fluid inside the abdomen (fluid may be taken during surgery); or that an ovary ruptures before or during surgery. Stage IC1 indicates that the tumor capsule broke during surgery to cause a surgical spill of cancer cells into the abdomen and/or pelvis. Stage IC2
5 indicates that at least one ovary or fallopian tube, or the tumor capsule ruptured before surgery. Stage IC3 indicates detection of cancer cells in fluid or washings from the abdomen.

Stage II indicates tumor extension beyond the ovary to other pelvic structures (e.g. fallopian tube or uterus). Stage IIA indicates that the cancer has grown into the
10 fallopian tubes or into the uterus. Stage IIB indicates that the cancer has grown into other pelvic tissues, such as the bladder or rectum. Stage IIC indicates that the cancer has grown into pelvic tissues and that cancer cells are present in fluid taken from inside the abdomen.

Stage III indicates metastasis to the peritoneum and/or regional lymph nodes.
15 Stage IIIA indicates that cancer growths are visible microscopically in tissue from the abdominal lining. Stage IIIA1 indicates that the cancer is in one or both ovaries or fallopian tubes, and may have spread or grown into the pelvis or nearby organs. Stage IIIA1(i) indicates that the areas of cancer spread in the lymph nodes are 10mm across or smaller. Stage IIIA1(ii) indicates that the areas of cancer spread in the lymph
20 nodes is great than 10mm across. Stage IIIA2 indicates that the cancer is in one or both ovaries or fallopian tubes, and may have spread or grown into the pelvis or nearby organs, and that microscopic examination of biopsies from the upper lining of the abdomen show small cancer deposits. Stage IIIB indicates that tumor growths 2cm or smaller are visible on the abdominal lining or the capsule of the liver or
25 spleen. Stage IIC indicates that tumor growths 2cm or larger are visible on the abdominal lining, and/or that cancer is detected in the lymph nodes of the upper abdomen, groin, or behind the uterus.

Stage IV indicates metastasis of the ovarian cancer past the peritoneum. Stage IVA indicates that the cancer has caused pleural effusion (build up fluid in the lining
30 of the lungs). Stage IVB indicates that the cancer has spread inside the liver or spleen, to lymph nodes in the groin, or outside the abdomen and/or to other organs (e.g. the lungs).

As used herein, the term “treating” or “treatment” refers to relieving, reducing, or alleviating at least one symptom in a subject or effecting a delay of progression of a disease. For example, treatment can be the diminishment of one or several symptoms of a disorder or complete eradication of a disorder, such as cancer. Within the
5 meaning of the present disclosure, the term “treat” also denotes to arrest and/or reduce the risk of worsening a disease, or prevention of at least one symptom associated with or caused by the state, disease or disorder being prevented. For example, treatments may relieve, reduce or alleviate at least one symptom of ovarian cancer.

The course of treatment of ovarian cancer is often similar for many patients in
10 order to account for the heterogeneity of response by the different cancer types. Such treatments include cytoreductive surgery (e.g. debulking to remove the tumor, salpingo-oophorectomy, hysterectomy, lymphadenectomy, omentectomy), followed by platinum-based chemotherapy (e.g. cisplatin or carboplatin and a taxane). In some instances, other chemotherapeutic agents may be used if the ovarian cancer is resistant
15 to platinum-based drugs either alone or in combination, such as liposomal doxorubicin, paclitaxel, docetaxel, nab-paclitaxel, gemcitabine, etoposide, pemetrexed, cyclophosphamide, topotecan, vinorelbine, or irinotecan.

Other treatments of ovarian cancer used in place of or in addition to surgery and chemotherapy may include administration of a targeted antibody, such as
20 bevacizumab or olaparib. In some instances, radiation therapy may be used to treat ovarian cancer, such as external beam radiation therapy or brachytherapy (implantation of a radioactive material near the affected) in order to target the affected area. In some instances, hormone therapy may be used to treat ovarian cancer, such as administration of luteinizing-hormone-releasing hormone (LHRH) agonists (e.g.
25 goserelin and/or leuprolide), tamoxifen, or aromatase inhibitors (e.g. letrozole, anastrozole, and exemestane).

It will be readily apparent to those skilled in the art that other suitable modifications and adaptations of the methods described herein may be made using suitable equivalents without departing from the scope of the embodiments disclosed
30 herein. Having now described certain embodiments in detail, the same will be more clearly understood by reference to the following examples, which are included for purposes of illustration only and are not intended to be limiting.

The following examples which should not be construed as further limiting.

Example 1: Neural networks for identifying ovarian cancer patients using miRNA sequencing data.

The goal of this study was to produce a circulating miRNA signature from human sera for the diagnosis of ovarian cancer. FIG. 1A shows a flow chart of the protocol for miRNA sequencing, filtering, batch adjustment and separation into the training and testing sets. FIG. 1B shows a flow chart of the model development and testing. A single heterogeneous patient cohort of pre-treatment (prior to either surgery or chemotherapy) blood samples was constructed from 180 women enrolled in two independent prospective cohort studies, the Pelvic Mass Protocol (PMP; see Materials and methods) consisting of 120 patients and the ERASMOS study (see Materials and Methods) comprising 60 patients. Both of the cohorts consisted of patients presenting for surgical management of a pelvic mass. The smaller cohort consisted of sequential cases, reflecting the natural frequency of different ovarian tumor subtypes in the clinical population, including the fact that most women with invasive ovarian cancer presented with advanced stage diseases. The larger cohort was derived from a case-control study and allowed enrichment of the study population for less common clinical cases that would be expected to confound a conventional screening algorithm, including benign complex ovarian masses, borderline tumors, early stage cancers, and non-serous histologic subtypes. To this latter cohort, healthy race and age-matched controls were added from the NECC study (see Materials and Methods). Table 1 below lists the demographics of the patients in the study populations.

Table 1: Demographics of patients in model study populations			
	ERASMOS (n=60)	PMP/NECC (n=119*)	p-value
Age, years, median (SD) †	57 (9.8)	56 (7.1)	0.44
CA-125, units/ml, median (SD) †	155 (689.8)	88.1 (1335.5)	0.72
Histology, n (%)‡			
Control	0 (0)	15 (12.6)	<0.0001
Serous cystadenoma / cystadenofibroma	7 (11.7)	14 (11.8)	
Endometrioma	0 (0)	15 (12.6)	
Other benign lesion	9 (15.0)	0 (0)	
Borderline mucinous tumor	2 (3.3)	0 (0)	
Borderline serous tumor	5 (8.3)	15 (12.6)	

Table 1: Demographics of patients in model study populations				
		ERASMOS (n=60)	PMP/NECC (n=119*)	p-value
	Stage I/II serous adenocarcinoma	5 (8.3)	20 (16.8)	
	Stage III/IV serous adenocarcinoma	19 (31.2)	10 (8.4)	
	Stage I/II clear cell / endometrioid adenocarcinoma	6 (10.0)	20 (16.8)	
	Stage III/IV clear cell/endometrioid adenocarcinoma	0 (0)	10 (8.4)	
	Mucinous adenocarcinoma	1 (1.7)	0 (0)	
	Other ovarian cancer	10 (10.0)	0 (0)	
Stage, n (%)‡				
	Not applicable	16 (26.7)	59 (49.6)	<0.0001
	I	9 (15.0)	22 (18.5)	
	II	8 (13.3)	18 (15.1)	
	III	19 (31.2)	18 (15.1)	
	IV	8 (13.3)	2 (1.7)	
Grade, n (%)‡				
	Not applicable	16 (26.7)	44 (37.0)	0.07
	Borderline	7 (11.7)	15 (12.6)	
	1 (well-differentiated)	6 (10.0)	12 (10.1)	
	2 (moderately differentiated)	3 (5.0)	12 (10.1)	
	3 (poorly differentiated)	28 (46.7)	36 (30.3)	
ERASMOS – Effects of Regional Analgesia on Serum miRNA after Oncology Surgery Study				
PMP – Pelvic Mass Protocol				
NECC – New England Case Control study				
* 15 samples from NECC, 114 samples from PMP				
† Student's t-test				
‡ chi-square test				

Total RNAs from the sera were extracted, converted into small RNA cDNA libraries, and sequenced, resulting in 2578 known or putative miRNA sequences. Filtering for miRNAs present in at least 50% of both cohorts at a detection threshold of 10 transcripts per million reads (tpm) left 192 miRNAs (see Table 2 below) to test in the models. One patient was excluded due to an unusual miRNA profile that was connected to recent cardiac ischemia.

Table 2: 192 miRNAs that can be reproducibly detected in human serum by miRNA sequencing		
<u>SEQ ID NO.</u>	<u>MicroRNA miRBase ID</u>	<u>Sequence</u>
1	hsa-let-7a-3p	CUAUACAAUCUACUGUCUUUC
2	hsa-let-7a-5p	UGAGGUAGUAGGUUGUAUAGUU

Table 2: 192 miRNAs that can be reproducibly detected in human serum by miRNA sequencing		
<u>SEQ ID NO.</u>	<u>MicroRNA miRBase ID</u>	<u>Sequence</u>
3	hsa-let-7b-3p	CUAUACAACCUACUGCCUUC
4	hsa-let-7b-5p	UGAGGUAGUAGGUUGUGUGGUU
5	hsa-let-7c-5p	UGAGGUAGUAGGUUGUAUGGUU
6	hsa-let-7d-3p	CUAUACGACCUGCUGCCUUUCU
7	hsa-let-7d-5p	AGAGGUAGUAGGUUGCAUAGUU
8	hsa-let-7e-5p	UGAGGUAGGAGGUUGUAUAGUU
9	hsa-let-7f-5p	UGAGGUAGUAGAUUGUAUAGUU
10	hsa-let-7g-5p	UGAGGUAGUAGUUUGUACAGUU
11	hsa-let-7i-5p	UGAGGUAGUAGUUUGUGCUGUU
12	hsa-miR-1-5p	UGGAUGUAAAGAAGUAUGUAU
13	hsa-miR-7-5p	UGGAAGACUAGUGAUUUUGUUGU
14	hsa-miR-10a-5p	UACCCUGUAGAUCCGAAUUUGUG
15	hsa-miR-10b-5p	UACCCUGUAGAACCGAAUUUGUG
16	hsa-miR-15b-3p	CGAAUCAUUAUUUGCUGCUCUA
17	hsa-miR-15b-5p	UAGCAGCACAUCAUGGUUUACA
18	hsa-miR-16-2-3p	CCAAUAUUACUGUGCUGCUUUA
19	hsa-miR-16-5p	UAGCAGCACGUAAAUAUUGGCG
20	hsa-miR-17-5p	CAAAGUGCUUACAGUGCAGGUAG
21	hsa-miR-19a-3p	UGUGCAAUAUCUAUGCAAACUGA
22	hsa-miR-19b-3p	UGUGCAAUCCAUGCAAACUGA
23	hsa-miR-20a-5p	UAAAGUGCUUAUAGUGCAGGUAG
24	hsa-miR-20b-5p	CAAAGUGCUCAUAGUGCAGGUAG
25	hsa-miR-21-5p	UAGCUUAUCAGACUGAUGUUGA
26	hsa-miR-22-3p	AAGCUGCCAGUUGAAGAACUGU
27	hsa-miR-22-5p	AGUUCUUCAGUGGCAAGCUUUA
28	hsa-miR-23a-3p	AUCACAUUGCCAGGGAUUUCC
29	hsa-miR-23b-3p	AUCACAUUGCCAGGGAUUACC
30	hsa-miR-24-2-5p	UGCCUACUGAGCUGAAACACAG
31	hsa-miR-24-3p	UGGCUCAGUUCAGCAGGAACAG
32	hsa-miR-25-3p	CAUUGCACUUGUCUCGGUCUGA
33	hsa-miR-26a-5p	UUCAAGUAAUCCAGGAUAGGCU
34	hsa-miR-26b-5p	UUCAAGUAAUUCAGGAUAGGU
35	hsa-miR-27a-3p	UUCACAGUGGCUAAGUUCGCG
36	hsa-miR-27b-3p	UUCACAGUGGCUAAGUUCUGC
37	hsa-miR-28-3p	CACUAGAUUGUGAGCUCCUGGA
38	hsa-miR-28-5p	AAGGAGCUCACAGUCUAUUGAG

Table 2: 192 miRNAs that can be reproducibly detected in human serum by miRNA sequencing		
<u>SEQ ID NO.</u>	<u>MicroRNA miRBase ID</u>	<u>Sequence</u>
39	hsa-miR-29a-3p	UAGCACCAUCUGAAAUCGGUUA
40	hsa-miR-30a-3p	CUUUCAGUCGGAUGUUUGCAGC
41	hsa-miR-30a-5p	UGUAAACAUCCUCGACUGGAAG
42	hsa-miR-30c-5p	UGUAAACAUCCUACACUCUCAGC
43	hsa-miR-30d-5p	UGUAAACAUCCCCGACUGGAAG
44	hsa-miR-30e-3p	CUUUCAGUCGGAUGUUUACAGC
45	hsa-miR-30e-5p	UGUAAACAUCCUUGACUGGAAG
46	hsa-miR-32-5p	UAUUGCACAUUACUAAGUUGCA
47	hsa-miR-92a-3p	UAUUGCACUUGUCCCCGGCCUGU
48	hsa-miR-92b-3p	UAUUGCACUCGUCCCCGGCCUCC
49	hsa-miR-93-5p	CAAAGUGCUGUUCGUGCAGGUAG
50	hsa-miR-98-5p	UGAGGUAGUAAGUUGUAUUGUU
51	hsa-miR-99a-5p	AACCCGUAGAUCCGAUCUUGUG
52	hsa-miR-99b-5p	CACCCGUAGAACCGACCUUGCG
53	hsa-miR-100-5p	AACCCGUAGAUCCGAACUUGUG
54	hsa-miR-101-3p	UACAGUACUGUGAUAAACUGAA
55	hsa-miR-103a-3p	AGCAGCAUUGUACAGGGCUAUGA
56	hsa-miR-106b-3p	CCGCACUGUGGGUACUUGCUCG
57	hsa-miR-106b-5p	UAAAGUGCUGACAGUGCAGAU
58	hsa-miR-107	AGCAGCAUUGUACAGGGCUAUCA
59	hsa-miR-122-5p	UGGAGUGUGACAAUGGUGUUUG
60	hsa-miR-125a-5p	UCCCUGAGACCCUUUAACCUUGUGA
61	hsa-miR-125b-5p	UCCCUGAGACCCUAACUUGUGA
62	hsa-miR-126-3p	UCGUACCGUGAGUAAUAAUGCG
63	hsa-miR-126-5p	CAUUAAUACUUUUGGUACGCG
64	hsa-miR-128-3p	UCACAGUGAACCGGUCUCUUU
65	hsa-miR-130a-3p	CAGUGCAAUGUUAAAAGGGCAU
66	hsa-miR-130b-5p	ACUCUUUCCCUGUUGCACUAC
67	hsa-miR-134-5p	UGUGACUGGUUGACCAGAGGGG
68	hsa-miR-139-3p	UGGAGACGCGGCCCGUUGGAGU
69	hsa-miR-139-5p	UCUACAGUGCACGUGUCUCCAGU
70	hsa-miR-140-3p	UACCACAGGGUAGAACCACGG
71	hsa-miR-140-5p	CAGUGGUUUUACCCUAUGGUAG
72	hsa-miR-142-3p	UGUAGUGUUUCCUACUUUAUGGA
73	hsa-miR-142-5p	CAUAAAGUAGAAAGCACUACU
74	hsa-miR-143-3p	UGAGAUGAAGCACUGUAGCUC

Table 2: 192 miRNAs that can be reproducibly detected in human serum by miRNA sequencing		
<u>SEQ ID NO.</u>	<u>MicroRNA miRBase ID</u>	<u>Sequence</u>
75	hsa-miR-144-3p	UACAGUAUAGAUGAUGUACU
76	hsa-miR-144-5p	GGAUAUCAUCAUUAACUGUAAG
77	hsa-miR-145-3p	GGAUUCCUGGAAAUACUGUUCU
78	hsa-miR-146a-5p	UGAGAACUGAAUUCCAUGGGUU
79	hsa-miR-146b-5p	UGAGAACUGAAUUCCAUGGCU
80	hsa-miR-148a-3p	UCAGUGCACUACAGAACUUUGU
81	hsa-miR-148a-5p	AAAGUUCUGAGACACUCCGACU
82	hsa-miR-148b-3p	UCAGUGCAUCACAGAACUUUGU
83	hsa-miR-150-5p	UCUCCCAACCCUUGUACCAGUG
84	hsa-miR-151a-3p	CUAGACUGAAGCUCCUUGAGG
85	hsa-miR-152-3p	UCAGUGCAUGACAGAACUUGG
86	hsa-miR-155-5p	UUA AUGCUAAUCGUGAUAGGGGU
87	hsa-miR-181a-2-3p	ACCACUGACCGUUGACUGUACC
88	hsa-miR-181a-5p	AACAUUCAACGCUGUCGGUGAGU
89	hsa-miR-181b-5p	AACAUUCAUUGCUGUCGGUGGGU
90	hsa-miR-181d-5p	AACAUUCAUUGUUGUCGGUGGGU
91	hsa-miR-182-5p	UUUGGCAAUGGUAGAACUCACACU
92	hsa-miR-183-5p	UAUGGCACUGGUAGAAUUCACU
93	hsa-miR-185-3p	AGGGGCUGGCUUCCUCUGGUC
94	hsa-miR-185-5p	UGGAGAGAAAGGCAGUUCCUGA
95	hsa-miR-186-5p	CAAAGAAUUCUCCUUUUGGGCU
96	hsa-miR-191-5p	CAACGGAAUCCCAAAGCAGCUG
97	hsa-miR-192-5p	CUGACCUAUGAAUUGACAGCC
98	hsa-miR-193a-5p	UGGGUCUUUGCGGGCGAGAUGA
99	hsa-miR-194-5p	UGUAAACAGCAACUCCAUGUGGA
100	hsa-miR-197-3p	UUCACCACCUUCUCCACCCAGC
101	hsa-miR-199a-3p	ACAGUAGUCUGCACAUUGGUUA
102	hsa-miR-199a-5p	CCCAGUGUUCAGACUACCUGUUC
103	hsa-miR-199b-3p	ACAGUAGUCUGCACAUUGGUUA
104	hsa-miR-200a-3p	UAACACUGUCUGGUAACGAUGU
105	hsa-miR-200c-3p	UAAUACUGCCGGGUAUGAUGGA
106	hsa-miR-203a-5p	GUGAAAUGUUUAGGACCACUAG
107	hsa-miR-215-5p	AUGACCUAUGAAUUGACAGAC
108	hsa-miR-221-3p	AGCUACAUUGUCUGCUGGGUUUC
109	hsa-miR-222-3p	AGCUACAUCUGGCUACUGGGU
110	hsa-miR-223-3p	UGUCAGUUUGUCAAUACCCCA

Table 2: 192 miRNAs that can be reproducibly detected in human serum by miRNA sequencing

<u>SEQ ID NO.</u>	<u>MicroRNA miRBase ID</u>	<u>Sequence</u>
111	hsa-miR-223-5p	CGUGUAUUUGACAAGCUGAGUU
112	hsa-miR-224-5p	CAAGUCACUAGUGGUUCCGUU
113	hsa-miR-320a	AAAAGCUGGGUUGAGAGGGCGA
114	hsa-miR-320b	AAAAGCUGGGUUGAGAGGGCAA
115	hsa-miR-320c	AAAAGCUGGGUUGAGAGGGU
116	hsa-miR-320d	AAAAGCUGGGUUGAGAGGA
117	hsa-miR-323a-3p	CACAUUACACGGUCGACCUCU
118	hsa-miR-323b-3p	CCCAAUACACGGUCGACCUCUU
119	hsa-miR-328-3p	CUGGCCCUCUCUGCCCUUCCGU
120	hsa-miR-330-3p	GCAAAGCACACGGCCUGCAGAGA
121	hsa-miR-335-3p	UUUUUCAUUAUUGCUCUGACC
122	hsa-miR-335-5p	UCAAGAGCAAU AACGAAAAAUGU
123	hsa-miR-338-5p	AACAAUAUCCUGGUGCUGAGUG
124	hsa-miR-339-3p	UGAGCGCCUCGACGACAGAGCCG
125	hsa-miR-340-3p	UCCGUCUCAGUUACUUUAUAGC
126	hsa-miR-340-5p	UUAUAAAGCAAUGAGACUGAUU
127	hsa-miR-342-5p	AGGGGUGCUAUCUGUGAUUGA
128	hsa-miR-345-5p	GCUGACUCCUAGUCCAGGGCUC
129	hsa-miR-361-3p	UCCCCAGGUGUGAUUCUGAUUU
130	hsa-miR-361-5p	UUAUCAGAAUCUCCAGGGGUAC
131	hsa-miR-363-3p	AAUUGCACGGUAUCCAUCUGUA
132	hsa-miR-370-3p	GCCUGCUGGGGUGGAACCUGGU
133	hsa-miR-374a-5p	UUAUAAUACAACCUGAUAAAGUG
134	hsa-miR-375	UUUGUUCGUUCGGCUCGCGUGA
135	hsa-miR-378a-3p	ACUGGACUUGGAGUCAGAAGGC
136	hsa-miR-378c	ACUGGACUUGGAGUCAGAAGAGUGG
137	hsa-miR-379-5p	UGGUAGACUAUGGAACGUAGG
138	hsa-miR-381-3p	UAUACAAGGGCAAGCUCUCUGU
139	hsa-miR-382-5p	GAAGUUGUUCGUGGUGGAUUCG
140	hsa-miR-409-3p	GAAUGUUGCUCGGUGAACCCCU
141	hsa-miR-421	AUCAACAGACAUAUAAUUGGGCGC
142	hsa-miR-423-3p	AGCUCGGUCUGAGGCCCCUCAGU
143	hsa-miR-423-5p	UGAGGGGCAGAGAGCGAGACUUU
144	hsa-miR-424-3p	CAAAACGUGAGGCGCUGCUAU
145	hsa-miR-425-3p	AUCGGGAAUGUCGUGUCCGCC
146	hsa-miR-425-5p	AAUGACACGAUCACUCCCGUUGA

Table 2: 192 miRNAs that can be reproducibly detected in human serum by miRNA sequencing

<u>SEQ ID NO.</u>	<u>MicroRNA miRBase ID</u>	<u>Sequence</u>
147	hsa-miR-432-5p	UCUUGGAGUAGGUCAUUGGGUGG
148	hsa-miR-450a-5p	UUUUGCGAUGUGUCCUAAUUAU
149	hsa-miR-450b-5p	UUUUGCAAUAUGUCCUGAAUA
150	hsa-miR-451a	AAACCGUUACCAUACUGAGUU
151	hsa-miR-483-5p	AAGACGGGAGGAAAGAAGGGAG
152	hsa-miR-484	UCAGGCUCAGUCCCCUCCCGAU
153	hsa-miR-486-3p	CGGGGCAGCUCAGUACAGGAU
154	hsa-miR-486-5p	UCCUGUACUGAGCUGCCCCGAG
155	hsa-miR-493-5p	UUGUACAUGGUAGGCUUUCAU
156	hsa-miR-500a-3p	AUGCACCUGGGCAAGGAUUCUG
157	hsa-miR-501-3p	AAUGCACCCGGGCAAGGAUUCU
158	hsa-miR-502-3p	AAUGCACCUGGGCAAGGAUUCA
159	hsa-miR-503-5p	UAGCAGCGGGAACAGUUCUGCAG
160	hsa-miR-532-5p	CAUGCCUUGAGUGUAGGACCGU
161	hsa-miR-543	AAACAUUCGCGGUGCACUUCUU
162	hsa-miR-550a-5p	AGUGCCUGAGGGAGUAAGAGCCC
163	hsa-miR-574-5p	UGAGUGUGUGUGUGUGAGUGUGU
164	hsa-miR-576-3p	AAGAUGUGGAAAAAUUGGAAUC
165	hsa-miR-584-5p	UUAUGGUUUGCCUGGGACUGAG
166	hsa-miR-589-5p	UGAGAACCACGUCUGCUCUGAG
167	hsa-miR-625-3p	GACUAUAGAACUUUCCCCCUCA
168	hsa-miR-629-5p	UGGGUUUACGUUGGGAGAACU
169	hsa-miR-652-3p	AAUGGCGCCACUAGGGUUGUG
170	hsa-miR-654-3p	UAUGUCUGCUGACCAUCACCUU
171	hsa-miR-660-5p	UACCCAUUGCAUAUCGGAGUUG
172	hsa-miR-664a-5p	ACUGGCUAGGGAAAAUGAUUGGAU
173	hsa-miR-744-5p	UGCGGGGCUAGGGCUAACAGCA
174	hsa-miR-769-5p	UGAGACCUCUGGGUUCUGAGCU
175	hsa-miR-941	CACCCGGCUGUGUGCACAUUGUC
176	hsa-miR-942-5p	UCUUCUCUGUUUUGGCCAUGUG
177	hsa-miR-1180-3p	UUUCCGGCUCGCGUGGGUGUGU
178	hsa-miR-1246	AAUGGAUUUUUGGAGCAGG
179	hsa-miR-1285-3p	UCUGGGCAACAAAGUGAGACCU
180	hsa-miR-1301-3p	UUGCAGCUGCCUGGGAGUGACUUC
181	hsa-miR-1307-3p	ACUCGGCGUGGCGUCGGUCGUG
182	hsa-miR-1307-5p	UCGACCGGACCUCGACCGGCU

Table 2: 192 miRNAs that can be reproducibly detected in human serum by miRNA sequencing		
<u>SEQ ID NO.</u>	<u>MicroRNA miRBase ID</u>	<u>Sequence</u>
183	hsa-miR-1908-5p	CGGCGGGGACGGCGAUUGGUC
184	hsa-miR-2110	UUGGGGAAACGGCCGCUGAGUG
185	hsa-miR-3158-3p	AAGGGCUUCCUCUCUGCAGGAC
186	hsa-miR-3613-5p	UGUUGUACUUUUUUUUUUGUUC
187	hsa-miR-3615	UCUCUCGGCUCUCGCGGCUC
188	hsa-miR-4433b-3p	CAGGAGUGGGGGUGGGACGU
189	hsa-miR-4443	UUGGAGGCGUGGGUUUU
190	hsa-miR-4732-5p	UGUAGAGCAGGGAGCAGGAAGCU
191	hsa-miR-6842-3p	UUGGCUGGUCUCUGCUCGCCAG
192	hsa-miR-7706	UGAAGCGCCUGUGCUCUGCCGAGA

The data for the remaining 179 patients were then batch-adjusted using ComBat to account for the two different study cohorts. Subject samples were randomized into “training” and “testing” sets in an approximate 3:1 ratio (see Table 3 below). There were no differences in the distribution of histopathological diagnoses between the training and testing sets ($p=1.0$).

Table 3: Demographics of patients after stratified random sampling into training and testing sets.			
	Training (n=135)	Testing (n=44)	p-value
Age, years, median (SD) †	56 (8.1)	56 (8.3)	1.0
CA-125, units/ml, median (SD) †	126.5 (1193.5)	105.6 (577.8)	0.91
Pathology, n (%)‡			1.0
Control	11 (8.1)	4 (9.1)	
Benign lesions	34 (25.2)	11 (25.0)	
Borderline tumors	16 (11.9)	5 (11.4)	
Stage I/II invasive cancers	41 (30.4)	12 (27.3)	
Stage III/IV invasive cancers	33 (24.4)	12 (27.3)	
† student's t-test			
‡ chi-square test			

As the dataset included more variables than cases, direct model development on the full dataset would have resulted in overfitted results. Therefore the variables for disease classification model development were preselected using three different methods: a significance filter, a group-stratified fold change filter, and a correlation-based feature selection (CFS) (see results in Table 4 below). Eleven different types of

- machine learning algorithms were then deployed (linear discriminant analysis, logistic regression, multivariate adaptive regression splines, naive Bayes, neural network, support vector machine, functional tree, Least Absolute Deviation regression trees (LAD tree), Bayesian network, elastic net regression, and random forest) on the three
- 5 sets of variables to separate the cases of invasive cancer from the healthy controls or benign/borderline masses. Although borderline tumors are not strictly benign, they are clinically indolent and seldom fatal, and were therefore grouped with benign lesions, as the goal of the analysis was to diagnose the tumors most contributing to mortality.

Table 4: Variables (biomarkers) selected using a significance filter, a correlation-based feature selection (CFS), and a group-stratified fold change filter		
<u>Significance-based selection</u>	<u>CFS variable selection</u>	<u>Expression fold selection</u>
hsa-miR-29a-3p (SEQ ID NO: 39)	hsa-miR-16-2-3p (SEQ ID NO: 18)	hsa-miR-23b-3p (SEQ ID NO: 29)
hsa-miR-30d-5p (SEQ ID NO: 43)	hsa-miR-200a-3p (SEQ ID NO: 104)	hsa-miR-29a-3p (SEQ ID NO: 39)
hsa-miR-200a-3p (SEQ ID NO: 104)	hsa-miR-200c-3p (SEQ ID NO: 105)	hsa-miR-32-5p (SEQ ID NO: 46)
hsa-miR-200c-3p (SEQ ID NO: 105)	hsa-miR-320b (SEQ ID NO: 114)	hsa-miR-92a-3p (SEQ ID NO: 47)
hsa-miR-320d (SEQ ID NO: 116)	hsa-miR-320d (SEQ ID NO: 116)	hsa-miR-150-5p (SEQ ID NO: 83)
hsa-miR-320c (SEQ ID NO: 115)		hsa-miR-200a-3p (SEQ ID NO: 104)
hsa-miR-450b-5p (SEQ ID NO: 149)		hsa-miR-200c-3p (SEQ ID NO: 105)
hsa-miR-203a-5p (SEQ ID NO: 106)		hsa-miR-203a-5p (SEQ ID NO: 106)
hsa-miR-486-3p (SEQ ID NO: 153)		hsa-miR-320c (SEQ ID NO: 115)
hsa-miR-1246 (SEQ ID NO: 178)		hsa-miR-320d (SEQ ID NO: 116)
hsa-miR-1307-5p (SEQ ID NO: 182)		hsa-miR-335-5p (SEQ ID NO: 122)
		hsa-miR-450b-5p (SEQ ID NO: 149)

Table 4: Variables (biomarkers) selected using a significance filter, a correlation-based feature selection (CFS), and a group-stratified fold change filter		
<u>Significance-based selection</u>	<u>CFS variable selection</u>	<u>Expression fold selection</u>
		hsa-miR-1246 (SEQ ID NO: 178)
		hsa-miR-1307-5p (SEQ ID NO: 182)

The machine learning algorithms were graded in terms of receiver operating characteristic area under the curve (ROC AUC) (see Table 5 below). Without using a true population with a defined prevalence of ovarian cancer, it was impossible to

5 assign a positive or negative predictive value for each test set (e.g. linear discriminant analysis using the biomarkers selected using the significance filter). However, it was possible to estimate the clinical utility of the algorithms by defining the sensitivity and specificity for each selected test set (see FIGS. 2A-2C). Descriptions of the variable selection methods and algorithms appear below in the Materials and Methods

10 section.

Table 5: Performance of the eleven classifier models on the testing set			
<u>Model/Method</u>	<u>Significance-based</u>	<u>Correlation-based feature selection subset</u>	<u>Fold change-based</u>
Linear discriminant analysis	0.80 (0.66-0.93)	0.76 (0.62-0.90)	0.78 (0.64-0.92)
Logistic regression	0.81 (0.68-0.94)	0.75 (0.61-0.90)	0.82 (0.70-0.94)
Neural network	0.84 (0.72-0.96)	0.75 (0.60-0.89)	0.90 (0.81-0.99)
Support vector machine	0.77 (0.63-0.91)	0.73 (0.58-0.87)	0.77 (0.63-0.91)
Multivariate adaptive regression splines	0.57 (0.40-0.74)	0.66 (0.49-0.82)	0.73 (0.58-0.88)
Naive Bayes classifier	0.75 (0.60-0.89)	0.68 (0.52-0.84)	0.75 (0.60-0.89)
Least Absolute Deviation regression tree	0.77 (0.63-0.91)	0.61 (0.44-0.78)	0.69 (0.53-0.84)
Functional tree	0.78 (0.64-0.91)	0.77 (0.63-0.91)	0.68 (0.52-0.84)
Bayesian network	0.72 (0.56-0.87)	0.67 (0.52-0.83)	0.72 (0.56-0.87)
Random forest	0.78 (0.64-0.91)	0.71 (0.56-0.86)	0.76 (0.62-0.90)
Elastic net	0.80 (0.67-0.93)	0.76 (0.62-0.90)	0.79 (0.66-0.92)
Values represent receiver operating characteristic (ROC) areas under curves (AUC) 95% confidence intervals shown in parentheses			

FIGS. 2A-2C each show a graph depicting the performance of the eleven disease classification models; the left-hand bars denote the sensitivity of the classification models, the right-hand bars denote the specificity of the classification models. The whiskers on the bars denote 95% confidence intervals. The graph of FIG. 2A shows the performance of the eleven disease classification models using the miRNA biomarkers selected using the significance-based filter. The graph of FIG. 2B shows the performance of the eleven disease classification models using a CFS subset algorithm. The graph of FIG. 2C shows the performance of the eleven disease classification models using the fold change-based filter. The arrow in FIG. 2C indicates the model with the largest sensitivity and specificity performance characteristics, which is the neural network analysis using the fourteen miRNA biomarkers (see biomarker listing in Table 4 above and Table 6 below).

The optimal machine learning algorithm proved to be a neural network analysis employing miRNA fold changes with a ROC AUC of 0.90 (95% CI: 0.81-0.99) (see arrow in FIG. 2C). This network was selected as the best among 5000 different neural networks constructed for each variable selection method (for a total of 15,000 different neural networks) through empirical optimization of the number of hidden layer neurons used and their activation functions. The network consisted of the fourteen individual miRNAs listed in Table 6 below.

Table 6: Fourteen miRNA biomarkers using fold changes in neural network algorithm		
<u>SEQ ID NO.</u>	<u>MicroRNA miRBase ID</u>	<u>Sequence</u>
29	hsa-miR-23b-3p	AUCACAUUGCCAGGGAUUACC
39	hsa-miR-29a-3p	UAGCACCAUCUGAAAUCGGUUA
46	hsa-miR-32-5p	UAUUGCACAUUACUAAGUUGCA
47	hsa-miR-92a-3p	UAUUGCACUUGUCCCGGCCUGU
83	hsa-miR-150-5p	UCUCCCAACCCUUGUACCAGUG
104	hsa-miR-200a-3p	UAACACUGUCUGGUAACGAUGU
105	hsa-miR-200c-3p	UAAUACUGCCGGGUAUGAUGGA
106	hsa-miR-203a-5p	GUGAAAUGUUUAGGACCACUAG
115	hsa-miR-320c	AAAAGCUGGGUUGAGAGGGU
116	hsa-miR-320d	AAAAGCUGGGUUGAGAGGA
122	hsa-miR-335-5p	UCAAGAGCAAUAACGAAAAAUGU
149	hsa-miR-450b-5p	UUUUGCAAUAUGUUCUGAAUA
178	hsa-miR-1246	AAUGGAUUUUUGGAGCAGG

Table 6: Fourteen miRNA biomarkers using fold changes in neural network algorithm

<u>SEQ ID NO.</u>	<u>MicroRNA miRBase ID</u>	<u>Sequence</u>
182	hsa-miR-1307-5p	UCGACCGGACCUCGACCGGCU

Seven neurons were used in the hidden layer of the neural network to calculate the prediction. The relationships between individual miRNA species were non-linear, so these relationships would likely have been obscured if a simple hierarchical clustering of the statistically significant miRNAs had been employed from the univariate analysis classifier model.

To examine whether the neural network suffered from overfitting or if its performance was due to the batch effect adjustment, the original raw sequencing data was analyzed using the neural network. Based on the fourteen selected miRNAs, the neural network worked equally well on the unadjusted raw datasets, with an AUC of 0.93 (95%CI: 0.89-0.98) in the training set and 0.90 (95%CI 0.80-0.99) in the testing set. FIG. 3A shows a graph of ROC curves, comparing the performance of the neural network on a training set of raw, non-batch-adjusted data (dotted line) with the performance of the neural network on a training set of batch-adjusted data (solid line). FIG. 3B shows a graph of ROC curves comparing the performance of the neural network on a testing set of raw non-batch adjusted data (dotted line) with the performance of the neural network on a testing set of batch-adjusted data (solid line). Again, this approach outperformed all the other classification methods examined.

Example 2: Comparison between neural networks with the CA-125 biomarker.

Next, the fourteen miRNA biomarkers were compared to the classification performance of CA-125. CA-125 is the most common circulating biomarker currently used in the diagnosis of ovarian cancer. While an elevated CA-125 level is a sensitive serum biomarker for advanced serous invasive epithelial ovarian cancer, it misses about 50% of cases of early stage ovarian cancer and non-serous histological type of ovarian cancer. Moreover, CA-125 has limited specificity, particularly for premenopausal women, where it can be elevated in a range of benign conditions, such as pregnancy, endometriosis, and menses. Among the 179 study subjects, preoperative serum CA-125 data was available for 120 patients. CA-125 levels were similarly distributed in the two patient cohorts. The neural network using the selected

fourteen miRNA biomarkers (AUC 0.93; 95% CI; 0.88-0.97) outperformed using the single biomarker CA-125 (AUC 0.74; 95% CI 0.65-0.83) in overall operating characteristics for the whole group ($p=0.001$). FIG. 4A is a graph of ROC curves comparing the performance of the neural network using a set of the selected fourteen biomarkers (solid line) with the performance of the CA-125 biomarker (dotted line), based on 120 patients.

Using the neural network to diagnose patients, the fourteen miRNA biomarkers also outperformed the CA-125 biomarker when the patients were separately analyzed as training and test sets, although dividing the patients into smaller sets reduced the power of the analysis (see FIGS. 4B-4C). FIG. 4B is a graph of ROC curves for the training data set, comparing the performance of the neural network classifying ovarian cancer in patients using the selected fourteen biomarkers (solid line) with the performance of the CA-125 biomarker (dotted line) ($N=94$). FIG. 4C shows a graph of ROC curves for the testing data set, comparing the performance of the neural network in classifying ovarian cancer using the selected fourteen biomarkers (solid line) with the performance of the CA-125 biomarker (dotted line) ($N=26$). In the training set ($N=94$), the fourteen biomarkers had an AUC of 0.92 (95% CI; 0.87-0.98) versus an AUC of 0.73 (95% CI; 0.59-0.96) for CA-125 alone ($p=0.0018$). In the test set ($N=26$), the fourteen miRNA biomarkers had an AUC of 0.94 (95% CI 0.89-0.99) versus an AUC of 0.78 (95% CI 0.62-0.84) for CA-125 alone ($p=0.12$).

The primary advantage of the neural network reliant on the fourteen miRNA biomarkers, over the single biomarker CA-125 is the improved ability to avoid false positives (65% fewer; $p=0.002$) and to identify more non-serous ovarian cancers (22% more; $p=0.002$). Table 7 below shows misclassification data for CA-125 and for the neural network prediction using the selected fourteen biomarkers. Notably, the neural network miRNA biomarkers and CA125 levels were independent of one another, as illustrated by the general lack of linear correlations between expression of the fourteen miRNAs used in the network and CA125 levels in cancer and non-cancer cases.

Table 7: Misclassification matrices for the neural network and CA-125 predictions with detailed histopathological data							
	CA-125 Prediction			Neural Network Prediction			p-value
Diagnosis	Correct	False Positive	False Negative	Correct	False Positive	False Negative	
Other benign mass	8	7	--	11	4	--	
Endometrioma	4	4	--	8	0	--	
Cystadenoma	2	1	--	2	1	--	
Borderline tumors	6	11	--	14	3	--	
Total false positive rate	23/43 (53%)			8/43 (19%)			0.002
Stage I/II Invasive Serous Adenocarcinoma	18	--	4	19	--	3	
Stage III/IV Invasive Serous Adenocarcinoma	21	--	1	20	--	2	
Total false negative rate	5/44 (11%)			5/44 (11%)			1.0
Stage I/II Invasive Clear Cell/Endometrioid Adenocarcinoma	11	--	7	14	--	4	
Stage III/IV Invasive Clear Cell/Endometrioid Adenocarcinoma	7	--	1	8	--	0	
Total false negative rate	8/26 (31%)			4/26 (15%)			0.002
Invasive cancer (other histologies)	4	--	3	5	--	2	
Total false negative rate	3/7 (43%)			2/7 (29%)			1.0
Correct – accurate classification as benign, borderline or control versus invasive cancer False positive – incorrectly assigned benign, borderline or control to invasive cancer False negative – incorrectly assigned invasive cancer to benign, borderline or control							

Example 3: Neural networks for identifying ovarian cancer patients using miRNA qPCR data.

It was important to know whether the miRNAs selected would only identify invasive ovarian cancer cases using sequencing data, or if the miRNA biomarker signature would be as robust if a different type of input for miRNA quantification was used. To validate the algorithm on a second platform, the expression levels of the fourteen miRNAs were measured (as well as an additional nine reference miRNAs derived from the sequencing data) by quantitative PCR (qPCR), and then repeated using the neural network construction. This produced a ROC curve (see FIG. 5A) with an AUC of 1.00 (95% CI; 1.00-1.00) on the training set and an AUC of 0.85 (95%CI; 0.71-0.99) on the testing set, respectively. FIG. 5A depicts a graph of ROC curves for the neural network analysis of qPCR data using the fourteen miRNA signature with minimal values imputed for undetectable miRNAs. As qPCR has a lower sensitivity than sequencing, miRNA values were undetectable for some miRNAs in some samples using this technique. This might account for a lower AUC for the testing set using qPCR than had been observed when using sequencing data.

To minimize the impact of missing data points on classifier performance, a global sensitivity analysis was performed on the best neural network for qPCR data, and the variables which contributed the least to the classifier's performance were iteratively removed. Following this analysis, the number of miRNAs used by the neural network was reduced to seven, listed below in Table 8.

Table 8: Seven miRNA biomarkers used in the final neural network model reliant on qPCR data		
<u>SEQ ID NO.</u>	<u>MicroRNA miRBase ID</u>	<u>Sequence</u>
39	hsa-miR-29a-3p	UAGCACCAUCUGAAAUCGGUUA
47	hsa-miR-92a-3p	UAUUGCACUUGUCCCGGCCUGU
105	hsa-miR-200c-3p	UAAUACUGCCGGGUAAUGAUGGA
115	hsa-miR-320c	AAAAGCUGGGUUGAGAGGGU
122	hsa-miR-335-5p	UCAAGAGCAAUAACGAAAAAUGU
149	hsa-miR-450b-5p	UUUUGCAAUAUGUUCUGAAUA
182	hsa-miR-1307-5p	UCGACCGGACCUCGACCGGCU

The number of normalizers used with these seven miRNAs was reduced to two: hsa-miR-423-3p (SEQ ID NO: 142) and hsa-miR-103a-3p (SEQ ID NO: 55), selected using the NormFinder method (see Materials and Methods below). The resulting network was composed of four neurons in the hidden layer maintained the performance of the miRNA sequence and qPCR-based networks, with an ROC AUC of 0.97; (95%CI; 0.91-1.00) on the training set, and a ROC AUC of 0.91 (95%CI; 0.85-1.00) on the testing set, while being a more robust tool with simpler input requirements. FIG. 5B is a graph of ROC curves for neural network analysis of qPCR data built on a seven miRNA signature after a global sensitivity analysis.

10 Example 4: External Validation of qPCR

The clinical performance of the final, locked-down diagnostic test was then tested on a completely independent external sample set collected from 51 preoperative patients treated in Lodz, Poland (see Table 9 below for clinical characteristics of the external validation set). In this population, the neural network had a positive predictive value of 91.3% (95% CI: 73.3–97.6%) and a negative predictive value of 78.6% (95% CI: 64.2–88.2%) with an AUC of 0.85 (see FIG. 6 for ROC AUC curve).

Table 9: Clinical Characteristics of the external validation set		
Characteristic		Polish external validation set (n=51)
Age, years, median (SD) †		55.5 (16.1)
Grade, n (%)‡		
	Borderline	4 (7.8)
	1	2 (3.9)
	2	7 (13.7)
	3	13 (25.5)
	Unspecified	3 (5.9)
	Benign	22 (43.1)
FIGO Stage, n (%)‡		
	I	7 (13.7)
	II	3 (5.9)
	III	18 (35.3)
	IV	1 (2.0)
	Benign	22 (43.1)
Histology, n (%)		
	Serous cystadenoma / cystadenofibroma	6 (11.8)
	Endometrioma/endometriosis	10 (19.6)
	Mature teratoma	6 (11.8)
	Borderline serous tumor	2 (3.9)
	Borderline seromucinous tumor	2 (3.9)

Table 9: Clinical Characteristics of the external validation set		
Characteristic		Polish external validation set (n=51)
	Serous adenocarcinoma	4 (7.8)
	Mucinous adenocarcinoma	1 (2.0)
	Endometrioid adenocarcinoma	1 (2.0)
	Clear Cell Adenocarcinoma	9 (17.6)
	Mixed adenocarcinoma	3 (5.9)
	Adenocarcinoma unspecified	7 (13.7)

Ideally, a serum biomarker should have biologic relevance to the clinical disease. To assess biologic relevance, the expression levels of the miRNAs from the ERASMOS patient set were examined to determine any changes in the cancer patients after surgical cytoreduction. Among the patients with ovarian cancer in the study, 27 had both preoperative and postoperative serum miRNAs profiled. These included 4/7 target miRNAs in the qPCR neural network model. Circulating levels of all three miRNAs decreased within 72 hr of tumor removal, with significant changes for miR-200a-3p and miR-200c-3p (see FIGS. 7A-7D).

To test if the miRNAs were in fact coming from the earliest lesions of this disease, paraffin-embedded tissue sections were assembled from independent sets of 15 cases of serous tubal intraepithelial carcinomas and 15 Stage I high grade (serous or Grade three endometrioid) epithelial ovarian cancers. Immunohistochemistry was performed on sequential sections for TP53 and Ki67 to highlight the lesions. In situ hybridization was then performed for three of the miRNAs in the neural network; mir-200c-3p, mir-335-5 p, and mir-92a-3p (see FIG. 8). FIG. 8 shows a series of photomicrographs taken at 10x, showing in situ expression of selected miRNAs from a serum signature, in sections of fallopian tubes showing serous tubal intraepithelial carcinoma (STIC) lesions and Stage I high grade serous ovarian cancer (HGSOC). Lesional cells are indicated by TP53 and Ki-67 staining. The top row of photomicrographs shows STIC lesion in continuity with normal fallopian tube at 20x. The middle row of photomicrographs shows STIC lesions in continuity with normal fallopian tube and invasive cancer with p53-null lesion at 10x. The bottom row of photomicrographs shows HGSOC intraluminal to the fallopian tube at 10x. The first column of photomicrographs (far left side) shows sections stained with hematoxylin and eosin (H+E); the second column shows staining for TP53; the third column shows

staining for Ki-67. The fourth, fifth and sixth columns show in situ hybridization for three different markers (miR-92a-3p, miR-200c-3-, and miR-335-5p, respectively).

In 100% of the samples, there was complete overlap between lesional cells and the miRNAs crucial for neural network performance, indicating that the miRNAs detected in the serum are present even in early lesions in the fallopian tube epithelium, and that pre-metastatic disease can be detected.

Example 5: Neural networks for identifying ovarian cancer patients using miRNA microarray data.

For external validation of the fourteen miRNA signature, both the miRNA signature and the neural network algorithm were tested on an independent, publicly available dataset previously published by Keller, *et al.* (Keller, A. *et al.* (2011) Toward the blood-borne miRNome of human diseases. *Nat Methods* **8**, 841-843). In that study, the authors collected blood samples from 454 individuals, among whom 15 had ovarian cancer and 70 were healthy controls. The data were generated using a probe-based microarray, which contained all fourteen miRNAs from the signature described above, allowing for one-to-one mapping of the fourteen microRNAs without any exclusion. Data were normalized as in the original manuscript by Keller, *et al.* The dataset was divided randomly into a training set and a testing set, using a 70:30 ratio. A neural network was developed using the fourteen miRNA signature to differentiate patients with cancer from controls. Using the original miRNA-sequence data set from Keller *et al.* (GSE31568 accession to the Gene Expression Omnibus database), 5000 networks were generated using the same protocol described herein, with the five best networks manually assessed for their complexity and performance.

The best neural network in terms of highest performance and lowest complexity had four neurons in the hidden layer. This neural network perfectly classified patients in the training set (AUC 1.00, 95% CI; 1.00-1.00) and provided good discriminatory power on the testing set (AUC 0.93, 95% CI; 0.81-1.00), with an overall sensitivity of 75% and specificity of 100%. Since the Keller dataset included a variety of other non-ovarian cancer diagnoses, the algorithm using the fourteen biomarker signature was analyzed to determine whether the algorithm was specific to ovarian cancer or would identify individuals with six other cancer types or six benign diagnoses.

FIG. 9 depicts a ROC curve showing that in the Keller et al. cohort of 454 patient samples with a wide array of medical diagnoses, including 70 healthy controls, a fourteen miRNA signature accurately identifies ovarian cancer cases against all other diagnoses (i.e. healthy controls or other cancers). This dataset is deposited in the Gene Expression Omnibus database (ncbi.nlm.nih.gov/geo/) under entry number GSE31568.

The fourteen miRNA signature was found to be unique to ovarian cancer over a variety of conditions, including pancreatic ductal cancer, other pancreatic cancer, prostate cancer, stomach cancer, melanoma, lung cancer, periodontitis, pancreatitis, multiple sclerosis, acute myocardial infarction, chronic obstructive pulmonary disease, and sarcoidosis. The neural network 14 miRNA signature did not separate any other diagnoses from the control group in the published dataset by Keller. The number of subjects (n) denotes the number of cases of the given diagnosis in the Keller dataset as follows: pancreatic ductal cancer (n = 45; FIG. 10A); prostate cancer (n = 23; FIG. 10B); stomach cancer (n = 13; FIG. 10C); other pancreatic cancers (n = 48; FIG. 10D); melanoma (n = 35; FIG. 10E); lung cancer (n = 32; FIG. 10F); periodontitis (n = 18; FIG. 10G); pancreatitis (n = 38; FIG. 10H); multiple sclerosis (n = 23; FIG. 10I); acute myocardial infarction (MI) (n = 20; FIG. 10J); chronic obstructive pulmonary disease (COPD) (n = 24; FIG. 10K); and sarcoidosis (n = 45; FIG. 10L). FIGS. 10A-10L show ROC curves illustrating the sensitivity and specificity for the each of the diseases listed above.

Materials and methods

Study Subjects

The model was developed from two ‘phase II’ specimen sets (i.e. samples collected from women prior to surgery or chemotherapy): one set was from Effects of Regional Analgesia on Serum microRNAs after Oncology Surgery (ERASMOS), and the other set from the Pelvic Mass Protocol (PMP) (Cramer et al., 2010, Correlates of the preoperative level of CA125 at presentation of ovarian cancer. *Gynecologic Oncology* 119:462–468; Elias et al., 2015, A prospective phase 0 study on the effects of anesthetic selection on serum miRNA profiles during primary cytoreductive surgery for suspected ovarian cancer. *Gynecologic Oncology* 137:1). Healthy subjects from the New England Case-Control (NECC) study were selected as controls. The NECC study was a large epidemiologic study matching cases of

ovarian cancer to geographically situated controls (Rice et al., 2013, Tubal ligation, hysterectomy and epithelial ovarian cancer in the New England Case-Control Study. International Journal of Cancer 133:2415–2421). These studies were approved by the Dana-Farber Cancer Institute Institutional Review Board Protocol 05–060 (NECC study), Brigham and Women’s Hospital Institutional Review Board Protocol 2000-P-001678 (Pelvic Mass Protocol), and Dana-Farber/Harvard Cancer Center Institutional Review Board Protocol 12–532 (ERASMOS). All subjects were enrolled after signing informed consent, and samples were collected fresh in 13 x 75 mm BD Vacutainer Plus Plastic Serum tubes (BD Life Sciences, Franklin Lakes, NJ) with spray-coated silica. Samples were allowed to clot 1 hr at room temperature before processing, then spun down by centrifugation at 1300 x g x 10 min, aliquoted into 1.5 ml vials and stored at – 80 C. Samples from the other studies were thawed and aliquoted for the current study and then refrozen.

ERASMOS Study

The Effects of Regional Analgesia on Serum microRNAs after Oncology Surgery (ERASMOS) study was an observational Phase 0 study that enrolled 60 patients from 03/2013 – 05/2015, conducted as a follow-up study to a prior publication associating regional analgesia exposure with improved ovarian cancer outcomes. Patients were approached consecutively for enrollment. Eligible patients were scheduled to undergo exploratory laparotomy for a pelvic mass suspicious for invasive epithelial ovarian cancer. All patients were enrolled after signing informed consent. The final patient cohort consisted of 60 patients with 3 patients excluded for inadequate sample collection. As the diagnosis for patients was not known *a priori*, a few cases of invasive ovarian cancer of other histologies (for example granulosa cell tumor and leiomyosarcoma) were identified but retained in the cohort analysis as classifying these as “benign” or excluding them would reduce the clinical utility of the testing strategy. The serum blood samples used in the present study were collected immediately preoperatively for each patient.

Pelvic Mass Protocol (PMP)

The Pelvic Mass Protocol (PMP) enrolled women referred to the DFCI/BWH Gynecologic Oncology service over the period 1992 to 2013 (Williams et al., 2014, Prognostic significance and predictors of the neutrophil-to-lymphocyte ratio in ovarian cancer. Gynecologic Oncology 132:542–550.). The PMP was a prospective

study performed at BWH collecting serum from women scheduled to undergo surgery for a known adnexal mass. Of some 455 women with a pelvic mass enrolled, a total of 120 samples were selected from the following categories: serous cystadenoma (Samuel and Carter, 2016, The diagnostic and prognostic potential of micrnas in epithelial ovarian carcinoma. *Molecular Diagnosis & Therapy* 21:59–73), serous borderline tumor (Samuel and Carter, 2016, The diagnostic and prognostic potential of micrnas in epithelial ovarian carcinoma. *Molecular Diagnosis & Therapy* 21:59–73), Stage I/II invasive serous adenocarcinoma (Häusler et al., 2010, Whole blood-derived miRNA profiles as potential new tools for ovarian cancer screening. *British Journal of Cancer* 103:693–700), and Stage III/IV invasive serous adenocarcinoma (Wang et al., 2016, Diagnostic potential of tumor DNA from ovarian cyst fluid. *eLife* 5:e15175), endometrioma (Samuel and Carter, 2016, The diagnostic and prognostic potential of micrnas in epithelial ovarian carcinoma. *Molecular Diagnosis & Therapy* 21:59–73), Stage I/II invasive clear cell or endometrioid adenocarcinoma (Häusler et al., 2010, Whole blood-derived miRNA profiles as potential new tools for ovarian cancer screening. *British Journal of Cancer* 103:693–700), or Stage III/IV invasive clear cell or endometrioid adenocarcinoma (Wang et al., 2016, Diagnostic potential of tumor DNA from ovarian cyst fluid. *eLife* 5:e15175). Overall, 37% of the subjects had benign disease, 12.6% had borderline tumors, 10.1% had low grade carcinomas, and 40.4% had high grade carcinomas. One sample of serous cystadenoma was excluded as an outlier due to a recent cardiovascular event as evidenced by extreme elevation of myocardial ischemia-associated miRNAs. From the most recent phase (2004–2008) of the NECC study, fifteen age and race matched healthy controls were selected; the controls were matched to the demographics of the EOC cases and benign disease controls from the PMP study. There was no overlap of subjects between the two studies. The samples sizes were based on a plan for a 2:1 ratio of early stage (Stage I/II) cancer cases to advanced stage (Stage III/IV) cases, a 1:1 ratio of invasive cancer cases: benign/borderline/control subjects, and for balanced numbers of healthy control: benign serous: benign endometrioid: borderline serous subjects. Borderline endometrioid or clear cell tumors were exceedingly rare and thus not included. For the qPCR model, 113 epithelial ovarian cancer cases and 113 healthy controls were added, matched for age and collection year. Twenty samples failed quality control, leaving 206 additional samples to add to the 119

samples originally profiled from PMP and creating a 325 sample set for qPCR-based model building and cut-off calibration.

Study Subjects for External Validation

Serum samples were collected from consecutive women undergoing surgical evaluation at the Medical University of Lodz, Poland, for a pelvic mass in association with an IRB-approved tumor collection protocol. All subjects were enrolled after signing informed consent, and samples were collected fresh in 13 x 75 mm BD Vacutainer Plus Plastic Serum tubes (BD Life Sciences, Franklin Lakes, NJ) with spray-coated silica. Samples were allowed to clot 1 hr at room temperature before processing, then spun down by centrifugation at 1300 x g for 10 min, aliquoted into 1.5 ml vials and stored at – 80°C. Samples were thawed only for the present study.

Outcome

Samples were classified as either invasive cancer or benign/borderline/controls. Although borderline tumors are not strictly benign, they are clinically indolent and seldom fatal; thus borderline samples were grouped with benign lesions, as the goal was to diagnose the tumors most contributing to mortality. For each patient, an estimated probability of >0.5 was classified as predicting invasive ovarian cancer.

Next Generation Sequencing (NGS)

Sample preparation, library construction, and miRNA sequencing were performed by Exiqon, Inc. (Vedbæk, Denmark). 500 µl of human serum from each sample were analyzed in duplicate. RNA from each sample was isolated using the miRCURY RNA isolation kit (Exiqon, Vedbæk, Denmark) according to the manufacturer's protocol optimized for serum. The quality of the isolated RNA was checked using qPCR. Total RNA was converted into microRNA NGS libraries using the NEBNext library generation kit (New England Biolabs Inc., Ipswich, MA) according to the manufacturer's instructions. Each individual RNA sample had adaptors ligated to its 3' and 5' ends and converted into cDNA. Then the cDNA was pre-amplified with specific primers containing sample specific indices. After 18 cycles of pre-PCR the libraries were purified on QiaQuick columns and the insert efficiency evaluated by a Bioanalyzer 2100 instrument on a high sensitivity DNA chip (Agilent Inc., Lexington, MA). The microRNA cDNA libraries were size-fractionated

on a LabChip XT (PerkinElmer, Waltham, MA) and a band representing adaptors and 15-40 bp insert excised using the manufacturer's instructions. Samples were then quantified using qPCR and concentration standards.

Based on quality of the inserts and the concentration measurements, the
5 libraries were pooled in equimolar concentrations (all concentrations of libraries to be pooled were of the same concentration). The library pools were finally quantified again with qPCR and the optimal concentration of the library pool used to generate the clusters on the surface of a flowcell before sequencing using v3 sequencing methodology according to the manufacturer instructions (Illumina Inc., Dedham,
10 MA). Samples were sequenced on the Illumina NextSeq 500 system (Illumina Inc., Dedham, MA) using a single-end read length of 50 nucleotides at an average of 10 million reads per sample. Sequence tags were mapped to miRbase 20 (mirbase.org/). After sequencing adapters were trimmed off as part of the base calling, trimming of adapters from the dataset revealed distinct peaks representing microRNA (~18-22nt).

15 Novel microRNAs not in standard miRBase or Rfam classification were identified based on the prediction algorithm miRPara. Expression levels were quantified in tags per million (TPM). TPM is a unit used to measure expression in NGS experiments. The number of reads for a particular microRNA is divided by the total number of mapped reads and multiplied by 1 million. Primary sequencing data
20 have been deposited in the Gene Expression Omnibus (GEO) database (ncbi.nlm.nih.gov/geo) in Accession No. GSE94533. The most stable miRNAs from the sequencing data were selected as normalizers using the NormFinder algorithm (Andersen et al., 2004).

qPCR Validation

25 Nine potential reference miRNAs were selected from the miRNA sequencing data using the NormFinder algorithm. Both the fourteen miRNAs from the diagnostic set and 9 potential reference miRNAs were quantified by qPCR using Exiqon (Vedbæk, Denmark) LNA-containing miRNA-specific probes. Briefly, miRNAs were polyadenylated and reverse transcribed into cDNA in a single reaction step, then
30 transferred to pre-loaded plates of primers using a pipetting robot. cDNA and ExiLent SYBR Green master mix were transferred to qPCR panels pre-loaded with primers using a pipetting robot. Amplification was performed on a Roche Lightcycler 480 (Roche, Basel, Switzerland). Amplification quality was determined by

generating melting curves; reactions with low efficiency or multiple peaks on the melting curve were discarded. Raw Cq values and melting points, detected by the Lightcycler software, were exported. Assays with several melting points or with melting points deviating from assay specifications were flagged and removed from the dataset. Reactions with amplification efficiency below 1.6 were also removed. Assays giving Cq values within 5 Cq values of the negative control sample were also removed from the dataset. Spike-in positive controls and no template negative controls were included. Minimum detection values for qPCR were established at 42 cycles; miRNAs with no amplification before that number of qPCR cycles were assumed to have their expression undetectable, and a quantification cycle (Cq) value of 42 was imputed as a substitute value. Data were normalized to the average of the assays detected in all samples (n = 120 samples). The nine selected reference miRNAs were re-evaluated after profiling for their stability across the arrays and the average of the four most stable probes (hsa-miR-423-3p, SEQ ID NO: 142; hsa-miR-191-5p, SEQ ID NO: 196; hsa-miR-221-3p, SEQ ID NO: 108; and hsa-miR-103a-3p, SEQ ID NO: 55) was used as the reference for the change in Cq (dCq) calculations.

Comparison of Preoperative and Postoperative samples

Individual miRNAs measurements from preoperative and postoperative serum samples from the ERASMOS study had been measured previously using multiplexed miRNA hydrogel probes (Fire-Plex, Abcam, Cambridge, MA) on a flow cytometer. Samples were profiled in duplicate, then replicates were merged. Fluorescence intensity values across all samples were normalized with Firefly Analysis Workbench (Abcam, Cambridge, MA) using the geNorm algorithm to identify appropriate normalizers (Vandesompele et al., 2002).

Pathology Samples

Paraffin blocks were selected from the surgical pathology files of the Brigham and Women's Hospital per BWH IRB Protocol #2016P002742. Hematoxylin and eosin sections of the cases were reviewed by a gynecologic pathologist (CC). The tissues had been routinely fixed in 10% neutral formalin and embedded in paraffin. Immunohistochemistry for TP53 and Ki-67 were performed using commercially available antibodies as previously described (Perets et al., 2013). Appropriate positive and negative (without primary antibodies) controls were used simultaneously for each antibody. In situ hybridization was performed using commercially available RNA

probes from Exiqon (Vedbæk, Denmark) according to the manufacturer's instructions. All probe concentrations were 1 nM. A probe for the small nuclear RNA U6 served as a positive control while a non-targeting scramble RNA probe served as negative control.

5 *Sample Size Estimation*

A testing set was sought showing a superiority of 0.1 in the area under the receiver operating characteristic curve (AUC) against a value of 0.75 (assumed as a null hypothesis for a clinically useful biomarker) with a statistical power of 80% and a type 1 error probability <0.05 (Hanley and McNeil, 1982). For statistical power
 10 estimation purposes, it was assumed that the model predictions would be moderately correlated with CA-125 levels ($r > 0.3$). The calculation yielded a required testing set of 44 patients (22 with invasive cancer and 22 without invasive cancer). To train the classifiers, it was assumed that the training set would require 3-fold more patients ($N = 132$) bringing the total number of required patient samples to 176 samples. The
 15 sample size was increase to 180 to account for potential clinical or technical outliers.

Statistical analysis

Basic statistics

Differences in the distribution of histopathologic diagnoses, grade, and stage between the cohorts and datasets were calculated using chi-square tests. Differences
 20 in false-positive and false-negative assignment were compared using Fisher's exact test. Differences in age and CA-125 levels between the cohorts and datasets were calculated using a Mann-Whitney U test. For all tests, a two-tailed p-value <0.05 was considered significant. For the ROC curves, cut-off values for prediction with the best diagnostic performance were established using the Youden index $\{ \text{sensitivity}_c +$
 25 $\text{specificity}_c - 1 \}$. Tests were performed in R.

Variable selection methods

For significance-based criterion, a student's t test was used, assuming miRNAs with $p < 0.05$ and false discovery rate < 0.05 for cancer versus benign/borderline/controls as significant.

30 For the fold change filter, miRNAs were selected that showed fold changes < 0.8 or > 1.2 for cancer versus benign/borderline/control comparisons in both the NECC and ERASMOS cohorts.

Correlation-based Feature Subset Selection (CFS) is a wrapper feature selection method that evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that were highly correlated with the class while
5 having low intercorrelation were preferred in the process. Search of the space of attribute subsets was performed by greedy hillclimbing augmented with a backtracking facility. This method of searching, called “Best First”, started with the empty set of attributes and searched the set forward.

All three sets of variables were analyzed using 11 different models. Six
10 models (linear discriminant analysis, logistic regression, multivariate adaptive regression splines, naive Bayes, neural network, and support vector machine) were developed using STATISTICA Data Miner 12.5 (StatSoft, Tulsa, OK, USA). The remaining five models (functional tree, LAD tree, Bayesian network, elastic net regression, and random forest) were created using Weka 3.9.0 (University of Waikato,
15 New Zealand). Descriptions of the classification models are presented below.

Linear discriminant analysis (LDA)

The linear discriminant analysis method creates a new set of spatial coordinates that allow for linear separation of the groups. The most discriminative features were extracted on the basis of their correlations and the model used a
20 backward stepwise variable selection algorithm only retaining in the model variables that showed final F values >5. This two-step filtering (variable selection after one of the three initial variable filtering algorithms) of the variables used in sample classification was aimed at the reduction of the number of miRNAs required for the model to work.

25 Depending on the number of variables selected by the filters, the discriminatory function of the LDA was based on a reduced set of miRNAs that passed the F value threshold and were retained in the model. For the subset of miRNAs filtered by statistical significance, the model used three miRNAs: hsa-miR-30d-5p (SEQ ID NO: 43), hsa-miR-200c-3p (SEQ ID NO: 105) and hsa-miR-320d
30 (SEQ ID NO: 116).

For CFS variable selection the model used three miRNAs: hsa-miR-320d (SEQ ID NO: 116), hsa-miR-200a-3p (SEQ ID NO: 104), and hsa-miR-16-2-3p (SEQ

ID NO: 18). The variable selection method based on stratified fold change used a yet another different set of miRNAs: hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR-150-5p (SEQ ID NO: 83).

Logistic regression

5 As above, the logistic regression model was built using a backward stepwise variable selection procedure, with variables showing $p < 0.15$ being retained in the final model. The procedure allowed for second order interactions between the variables to detect potential subgroup-specific effects. A standard quasi-Newton estimation procedure was performed in model development.

10 After exclusion of variables with p values > 0.15 in the multivariate model, the miRNAs remaining in the classifier were hsa-miR-30d-5p (SEQ ID NO: 43), hsa-miR-320d (SEQ ID NO: 116), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-1246 (SEQ ID NO: 178), and an interaction of hsa-miR-200c-3p* hsa-miR-1246 (SEQ ID NO: 178). A logistic regression model based on miRNAs selected by the CFS variable
15 algorithm required only two miRNAs to work: hsa-miR-200c-3p (SEQ ID NO: 105) and hsa-miR-320d (SEQ ID NO: 116). A logistic regression classifier built on the fold change filter-selected miRNAs used three miRNAs: hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-320d (SEQ ID NO: 116), hsa-miR-1246 (SEQ ID NO: 178), and an interaction between hsa-miR-200c-3p (SEQ ID NO: 105)* hsa-miR-1246 (SEQ ID
20 NO: 178). Results of all three models were convergent and the crucial role of hsa-miR-200c-3p (SEQ ID NO: 105)/ hsa-miR-320d (SEQ ID NO: 116) was confirmed by all models. The logistic regression model achieved a similar performance to that of the neural net in the CFS-selected variable subset. This was a logical consequence of a strong variable filtering leaving too few input variables for the network to
25 identify subtle patterns.

Multivariate Adaptive Regression Splines (MARS)

An alternative approach to modeling of the classification function was the MARS model – a modification of a multivariate joint-point regression which estimates a number of basal functions most appropriate for data from specific
30 fragments of the multidimensional dataset. The method is used in complex function modeling of non-monotonous or non-linear associations. Within our analysis we used a MARS model that allowed for up to third degree interactions between the variables, allowing for up to $1.5 \times (n \text{ variables})$ basal function in each model and penalizing the

introduction of additional basal functions by a factor of 2. Interactions between variables were tested for improvement of model performance up to the degree of three. During the model building procedure we iteratively removed variables absent in any of the basal functions until only miRNAs used in at least one basal function
 5 remained in the MARS model.

Using 11 miRNAs filtered on the basis of significance, a MARS model was created, which included 14 basal functions. All functions were transformation of five, single miRNAs: hsa-miR-30d-5p (SEQ ID NO: 43), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-200a-3p (SEQ ID NO: 104), and
 10 hsa-miR-1307-3p (SEQ ID NO: 181). The MARS model built on CFS-filtered variables consisted of seven basal functions based on four miRNAs: hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320d (SEQ ID NO: 116), hsa-miR-16-2-3p (SEQ ID NO: 18), and hsa-miR-320b (SEQ ID NO: 114). The final MARS model built on 14 miRNAs filtered by the stratified fold change threshold was optimized at 10 basal
 15 functions based on five miRNAs: hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR-92-3p (SEQ ID NO: 47), hsa-miR-203a (SEQ ID NO: 106), and hsa-miR-320c (SEQ ID NO: 115). All MARS models showed relatively poor performance hinting at issues with model overfitting and low specificity (for example, the ROC AUC for the significance-based
 20 and CFS variable selection inputs did not meet statistical significance).

Elastic Net regression

An elastic-net regularized generalized linear model is a linear regression using coordinate descent. In order to train this model we have used Java implementation of a component of the R package 'glmnet' in Weka software. As a regression method for
 25 classification was desired, class was binarized and one regression model was built for each class value (i.e. meta-scheme classification via regression). The alpha elastic-net mixing parameter was chosen to be 0.001 while the epsilon value for generating the lambda sequence was set to 10^{-4} . Additionally, a covariance update method was used. This resulted in the following formula:

30 *weka.classifiers.meta.ClassificationViaRegression -W*
weka.classifiers.functions.ElasticNet -- -m2 y -alpha 0.001 -lambda_seq -thr 1.0E-7 -
mxit 10000000 -numModels 100 -infolds 10 -eps 1.0E-4 -sparse n -stderr_rule n -

addStats n. Reproduction of model induction may require installing additional packages from WEKA package manager.

Neural Network

Five thousand (5000) neural networks were built for each variable selection
5 method (15,000 networks in total), and the best neural network (in terms of
performance in properly assigning classes to the test set) was retained. The networks
were built in a semi-automated way. Their structure was of a multilayer perceptron
with a number of neurons in the hidden layer iteratively optimized from $(n$
variables)/3 to $(n$ variables)*1.5 to avoid overfitting. Admissible linking functions
10 between the layers were linear, logistic, hyperbolic tangential, and exponential.
Neuron weights were calculated using the BFGS (Broyden-Fletcher-Goldfarb-
Shanno) algorithm and the network was trained in each epoch using an error back-
propagation algorithm to optimize weights in each pass.

Support Vector Machine (SVM)

15 This classifier was built with a set of different entry parameters: kernel
function types, function parameters, and hinge loss function. Admissible kernel
functions were linear, polynomial (2nd and 3rd order) and radial basis function (gamma
from 0.1 to 1 tested in 0.1 increments). The models performed worse than simpler
classification tools (logistic regression/linear discriminant analysis), possibly due to a
20 small number of cases available for testing.

Naïve Bayes classifier

A priori class probabilities were estimated empirically on the basis of class
frequencies in the dataset, normal distribution was assumed for all log-10 transformed
miRNA expression values quantified as transcripts per million. The exact probability
25 estimator of the naïve Bayes classifier showed similar performance on all three
variable subsets, achieving accuracy comparable to that of the SVM model

LAD tree

LADTree is completely deterministic tree that allows decision making by
counting respective probabilities on the pathway through the tree. Multi-class
30 alternating decision tree using the LogitBoost strategy (LAD Tree; ECML '02
Proceedings of the 13th European Conference on Machine Learning, pp. 161-172,

Aug.19-23, 2002). The number of boosting iterations to use, which determined the size of the trees, was set to be 10.

Formula: *weka.classifiers.trees.LADTree -B 10*. Reproduction of model induction may require installing additional packages from WEKA package manager.

5 *Functional Tree*

Functional trees are logistic classification decision trees that have logistic regression functions at the inner nodes or leaves. Training of models was performed by WEKA software. As in default settings, minimum number of instances at which a node is considered for splitting was 15, number of iterations for LogitBoost was also 10 15 and no weight trimming was applied.

Formula: *weka.classifiers.trees.FT -I 15 -F 0 -M 15 -W 0.0*. Please note that reproduction of model induction may require installing additional packages from WEKA package manager.

All functional trees were models with one node. In order to infer how this 15 model works, evaluation of values for linear combination function at each node for every class has to be done. For example, for cancer in the CFS-processed dataset the formula is:

$$F1 = -1.75 + [\text{hsa-miR-16-2-3p}] * -0.29 + [\text{hsa-miR-200a-3p}] * 0.08 + [\text{hsa-miR-200c-3p}] * 1.07 + [\text{hsa-miR-320b}] * -0.21 + [\text{hsa-miR-320d}] * 1.29$$

20 As the classifiers are binary, the result for the second class (F2) should be an opposite number (F1 = -F2). In the next step the value of the following formula should be calculated and compared to threshold of the node:

$$\frac{e^{F1}}{e^{F1} + e^{F2}}$$

Bayesian Network

A Bayes Network was trained using a K2 search algorithm, which is a hill 25 climbing algorithm restricted by an order on the variables. The initial network used for structure learning was a Naive Bayes Network and there could be only one parent a node. Conditional probability tables of a Bayes network were driven directly from data once the structure has been learned (with alpha value equal to 0.5). Formula: *weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -* 30 *P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5*.

Reproduction of model induction may require installing additional packages from WEKA package manager.

Random Forest

- Random forest is a technique of random decision forests that considers K
 5 randomly chosen attributes at each node. K was calculated as integer of 1 plus binary
 logarithm of number of predictors. Minimum proportion of the variance needed at a
 node in order for splitting to be performed was set to 0.001. No backfitting was
 performed.

- Formula: *weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0*
 10 *-M 1.0 -V 0.001 -S 1*. Reproduction of model induction may require installing
 additional packages from WEKA package manager.

CLAIMS

What is claimed is:

1. A method for diagnosing ovarian cancer in a female subject, the method comprising:
 - (a) obtaining a sample collected from the subject;
 - (b) determining the amounts of one or more test microRNAs selected from the group consisting of hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-450b-5p (SEQ ID NO: 149), and hsa-miR-1307-5p (SEQ ID NO: 182) in the sample; and
 - (c) comparing the amounts of the test microRNAs determined in step (b) to a statistical model,thereby diagnosing ovarian cancer in the subject.
2. The method of claim 1, wherein step (b) comprises determining the amounts of seven test microRNAs of hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-450b-5p (SEQ ID NO: 149), and hsa-miR-1307-5p (SEQ ID NO: 182) in the sample.
3. The method of claim 2, wherein step (b) consists of determining the amounts of seven test microRNAs of hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-450b-5p (SEQ ID NO: 149), and hsa-miR-1307-5p (SEQ ID NO: 182) in the sample.
4. The method of claim 1 or 2, wherein step (b) further comprises determining the amounts of one or more additional test microRNAs selected from the group consisting of hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46),

hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR-203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR-1246 (SEQ ID NO: 178) in the sample.

5. The method of claim 1 or 2, wherein step (b) further comprises determining the amounts of seven additional test microRNAs of hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR-203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR-1246 (SEQ ID NO: 178) in the sample.

6. The method of claim 5, wherein step (b) consists of determining the amounts of fourteen test microRNAs of hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR-203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR-1246 (SEQ ID NO: 178) in the sample.

7. The method of claim 1, 2, 4, or 5, wherein the amounts of no more than 20 test microRNAs are determined.

8. The method of any one of claims 1-7, further comprising:

(d) determining the amounts of one or more normalizing microRNA selected from the group consisting of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196) in the sample; and

(e) normalizing the determined amounts of the test microRNAs using the amounts of the normalizing microRNAs determined in step (d).

9. The method of claim 8, where step (d) comprises determining the amounts of two normalizing microRNAs of hsa-miR-103a-3p (SEQ ID NO: 55) and hsa-miR-423-3p (SEQ ID NO: 142) in the sample.
10. The method of claim 8 or 9, wherein step (d) comprises determining the amounts of four normalizing microRNAs of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196) in the sample.
11. The method of claim 10, wherein step (d) consists of determining the amounts of four normalizing microRNAs of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196) in the sample.
12. The method of any one of claims 1-7, wherein the amounts of the test microRNAs are not normalized against the amount of any additional microRNA.
13. The method of any one of claims 1-12, wherein step (b) and/or step (d) are performed by detecting binding of the sample to at least one probe capable of specifically hybridizing to each of the microRNAs or a cDNA thereof.
14. The method of claim 13, wherein step (b) and/or step (d) are performed using a nucleic acid detection assay, optionally wherein the assay is selected from the group consisting of microarray, RT-PCR, and RT-qPCR.
15. The method of claim 14, wherein step (b) and/or step (d) are performed using RT-qPCR.

16. The method of any one of claims 13-15, wherein at least one of the probes comprises a detectable label.
17. The method of any one of claims 13-16, wherein each one of the probes comprises a detectable label.
18. The method of any one of claims 1-12, wherein step (b) and/or step (d) are performed by:
- reverse-transcribing the microRNA molecules in the sample, thereby obtain a cDNA sample; and
- sequencing the cDNA sample.
19. The method of claim 18, further comprising amplifying the DNA molecules in the cDNA sample before sequencing the cDNA sample.
20. The method of any one of the preceding claims, wherein the sample is selected from a sample of blood, lymph node, or ovary.
21. The method of claim 20, wherein the sample is a blood sample, optionally wherein the blood sample is selected from the group consisting of plasma, serum, and whole blood.
22. The method of claim 20, wherein the sample is a sample of lymph node, optionally wherein the lymph node is a draining lymph node for an ovary.
23. The method of claim 20, wherein the sample is an ovary sample, optionally wherein the ovary is suspected to contain a tumor.

24. The method of any one of the preceding claims, wherein the subject is at the risk of developing ovarian cancer.
25. The method of claim 24, wherein the subject has one or more ovaries.
26. The method of claim 24 or 25, wherein the subject has a higher level of CA-125 antigen in the blood than a diagnostic threshold level, optionally wherein the diagnostic threshold level is 35 U/mL.
27. The method of any one of claims 24-26, wherein the subject has at least one mutation in the BRCA1 or BRCA2 gene.
28. The method of any one of claims 24-27, wherein the subject is at least 55 years of age.
29. The method of any one of claims 1-28, wherein the subject is a human subject.
30. A kit comprising at least one test probe capable of specifically hybridizing to a microRNA selected from the group consisting of hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-320c (SEQ ID NO: 115), and hsa-miR-200c-3p (SEQ ID NO: 105), or a cDNA thereof.
31. The kit of claim 30, comprising at least one test probe capable of specifically hybridizing to each of the microRNAs of hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-320c (SEQ ID NO: 115), and hsa-miR-200c-3p (SEQ ID NO: 105), or a cDNA thereof.

32. The kit of claim 31, wherein the test probes consist of seven probes to microRNAs with one test probe capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-320c (SEQ ID NO: 115), and hsa-miR-200c-3p (SEQ ID NO: 105), or a cDNA thereof.

33. The kit of claim 30 or 31, further comprising at least one additional test probe capable of specifically hybridizing to a microRNA selected from the group consisting of hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR1246 (SEQ ID NO: 178), or a cDNA thereof.

34. The kit of claim 30 or 31, further comprising at least one additional test probe capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR1246 (SEQ ID NO: 178), or a cDNA thereof.

35. The kit of claim 34, wherein the test probes consist of fourteen probes to microRNAs with one test probe capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR1246 (SEQ ID NO: 178), or a cDNA thereof..

36. The kit of any one of claims 30-35, further comprising at least one normalizing probe capable of specifically hybridizing to a microRNA selected from the group consisting of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196), or a cDNA thereof.

37. The kit of any one of claims 30-35, further comprising at least one normalizing probe capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-103a-3p (SEQ ID NO: 55) and hsa-miR-423-3p (SEQ ID NO: 142), or a cDNA thereof.

38. The kit of any one of claims 30-35, further comprising at least one normalizing probe capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196), or a cDNA thereof.

39. The kit of claim 38, wherein the normalizing probes consist of four probes to microRNAs with one normalizing probe capable of specifically hybridizing to each of the microRNAs selected from the group consisting of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196), or a cDNA thereof.

40. The kit of any one of claims 30-35, wherein the kit comprises no normalizing probes.

41. The kit of any one of claims 30-40, wherein at least one of the probes comprises a detectable label.

42. The kit of any one of claims 30-41, wherein each one of the probes comprises a detectable label.

43. The kit of any one of claims claim 30-42, further comprising a reagent for reverse transcription of a microRNA molecule.

44. A method for treating a female subject suspected of having an ovarian cancer, the method comprising:

(a) obtaining a sample collected from the subject;

(b) determining the amounts of one or more test microRNAs selected from the group consisting of hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-450b-5p (SEQ ID NO: 149), and hsa-miR-1307-5p (SEQ ID NO: 182) in the sample;

(c) comparing the amounts of the test microRNAs determined in step (b) to a statistical model; and

(f) treating the subject for ovarian cancer, or monitor the subject for ovarian cancer, based on the comparison of step (c).

45. The method of claim 44, wherein step (b) comprises determining the amounts of seven test microRNAs of hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-450b-5p (SEQ ID NO: 149), and hsa-miR-1307-5p (SEQ ID NO: 182) in the sample.

46. The method of claim 45, wherein step (b) consists of determining the amounts of seven test microRNAs of hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-450b-5p (SEQ ID NO: 149), and hsa-miR-1307-5p (SEQ ID NO: 182) in the sample.

47. The method of claim 44 or 45, wherein step (b) further comprises determining the amounts of one or more additional test microRNAs selected from the group consisting of hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR-203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR-1246 (SEQ ID NO: 178) in the sample.

48. The method of claim 44 or 45, wherein step (b) further comprises determining the amounts of seven additional test microRNAs of hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR-203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR-1246 (SEQ ID NO: 178) in the sample.

49. The method of claim 48, wherein step (b) consists of determining the amounts of fourteen test microRNAs of hsa-miR-29a-3p (SEQ ID NO: 39), hsa-miR-92a-3p (SEQ ID NO: 47), hsa-miR-200c-3p (SEQ ID NO: 105), hsa-miR-320c (SEQ ID NO: 115), hsa-miR-335-5p (SEQ ID NO: 122), hsa-miR-450b-5p (SEQ ID NO: 149), hsa-miR-1307-5p (SEQ ID NO: 182), hsa-miR-23b-3p (SEQ ID NO: 29), hsa-miR-32-5p (SEQ ID NO: 46), hsa-miR-150-5p (SEQ ID NO: 83), hsa-miR-200a-3p (SEQ ID NO: 104), hsa-miR-203a (SEQ ID NO: 106), hsa-miR-320d (SEQ ID NO: 116), and hsa-miR-1246 (SEQ ID NO: 178) in the sample.

50. The method of claim 44, 45, 47, or 48, wherein the amounts of no more than 20 test microRNAs are determined.

51. The method of any one of claims 44-50, further comprising:

(d) determining the amounts of one or more normalizing microRNA selected from the group consisting of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196) in the sample; and

(e) normalizing the determined amounts of the test microRNAs using the amounts of the normalizing microRNAs determined in step (d).

52. The method of claim 51, where step (d) comprises determining the amounts of two normalizing microRNAs of hsa-miR-103a-3p (SEQ ID NO: 55) and hsa-miR-423-3p (SEQ ID NO: 142) in the sample.

53. The method of claim 51 or 52, wherein step (d) comprises determining the amounts of four normalizing microRNAs of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196) in the sample.

54. The method of claim 53, wherein step (d) consists of determining the amounts of four normalizing microRNAs of hsa-miR-103a-3p (SEQ ID NO: 55), hsa-miR-221-3p (SEQ ID NO: 108), hsa-miR-423-3p (SEQ ID NO: 142), and hsa-miR-191-5p (SEQ ID NO: 196) in the sample.

55. The method of any one of claims 44-50, wherein the amounts of the test microRNAs are not normalized against the amount of any additional microRNA.

56. The method of any one of claims 44-55, wherein step (b) and/or step (d) are performed by detecting binding of the sample to at least one probe capable of specifically hybridizing to each of the microRNAs or a cDNA thereof.

57. The method of claim 56, wherein step (b) and/or step (d) are performed using a nucleic acid detection assay, optionally wherein the assay is selected from the group consisting of microarray, RT-PCR, and RT-qPCR.

58. The method of claim 57, wherein step (b) and/or step (d) are performed using RT-qPCR.

59. The method of any one of claims 56-58, wherein at least one of the probes comprises a detectable label.

60. The method of any one of claims 56-59, wherein each one of the probes comprises a detectable label.

61. The method of any one of claims 44-55, wherein step (b) and/or step (d) are performed by:

reverse-transcribing the microRNA molecules in the sample, thereby obtain a cDNA sample; and

sequencing the cDNA sample.

62. The method of claim 61, wherein step (b) and/or step (d) further comprises amplifying the DNA molecules in the cDNA sample before sequencing the cDNA sample.

63. The method of any one of claims 44-62, wherein the sample is selected from a sample of blood, lymph node, or ovary.

64. The method of claim 63, wherein the sample is a blood sample, optionally wherein the blood sample is selected from the group consisting of plasma, serum, and whole blood.

65. The method of claim 63, wherein the sample is a sample of lymph node, optionally wherein the lymph node is a draining lymph node for an ovary.

66. The method of claim 63, wherein the sample is an ovary sample, optionally wherein the ovary is suspected to contain a tumor.
67. The method of any one of claims 44-66, wherein the subject is at the risk of developing ovarian cancer.
68. The method of claim 67, wherein the subject has one or more ovaries.
69. The method of claim 67 or 68, wherein the subject has a higher level of CA-125 antigen in the blood than a diagnostic threshold level, optionally wherein the diagnostic threshold level is 35 U/mL.
70. The method of any one of claims 67-69, wherein the subject has at least one mutation in the BRCA1 or BRCA2 gene.
71. The method of any one of claims 67-70, wherein the subject is at least 55 years of age.
72. The method of any one of claims 44-71, wherein the subject is a human subject.
73. The method of any one of claims 44-72, wherein treating the subject is performed using one or more treatments selected from the group consisting of: surgery, chemotherapy, targeted antibody therapy, radiation therapy, and hormone therapy.
74. The method of any one of claims 1-29 and 44-73, wherein the statistical model comprises one or more models selected from the group consisting of linear discriminant analysis, logistic regression, multivariate adaptive regression splines,

naive Bayes, neural network, support vector machine, functional tree, LAD tree, Bayesian network, elastic net regression, and random forest.

75. The method of claim 74, wherein the statistical model comprises a neural network.

76. The method of claim 75, wherein the statistical model consists of a neural network.

1/20

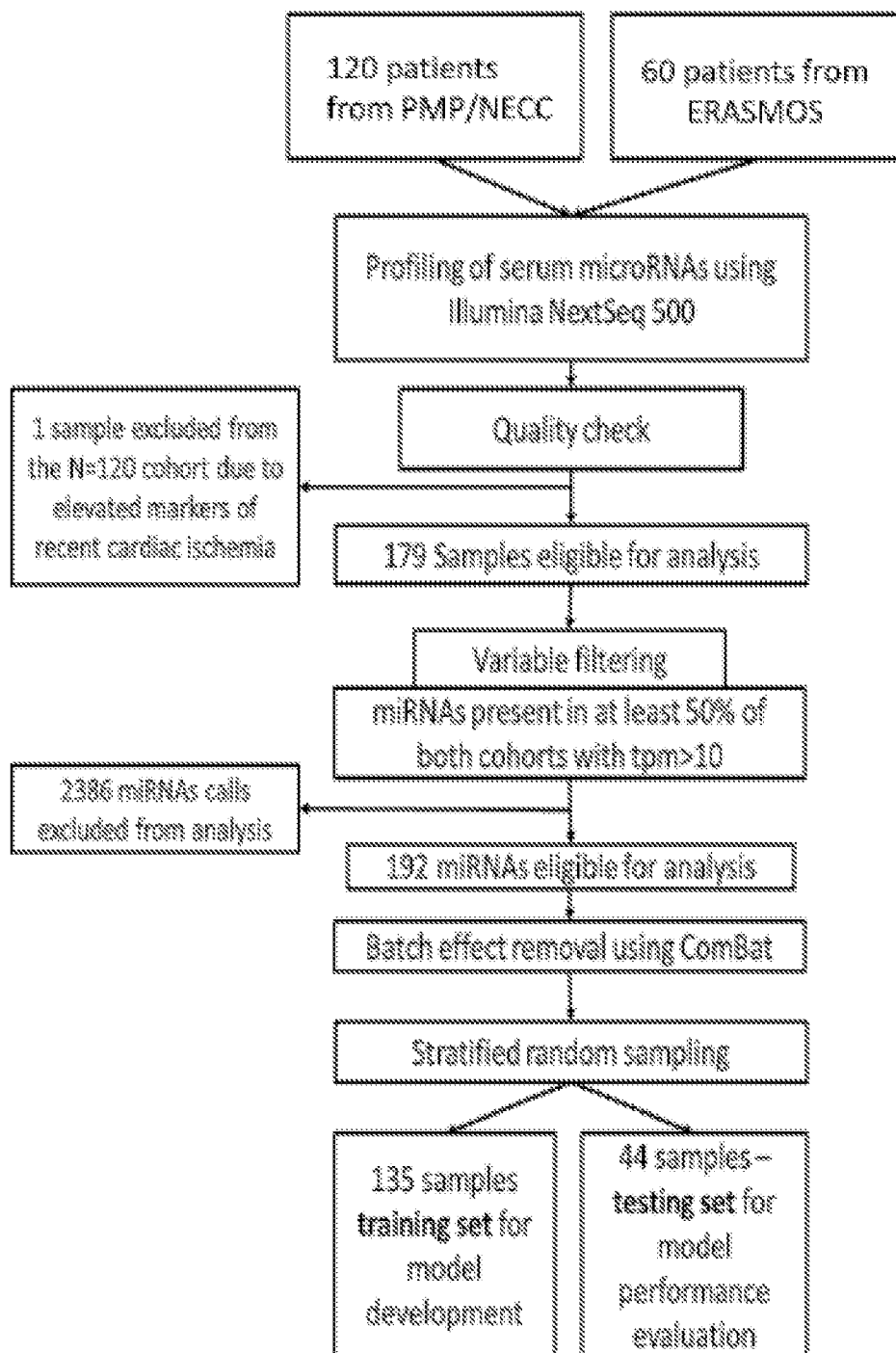


FIG. 1A

2/20

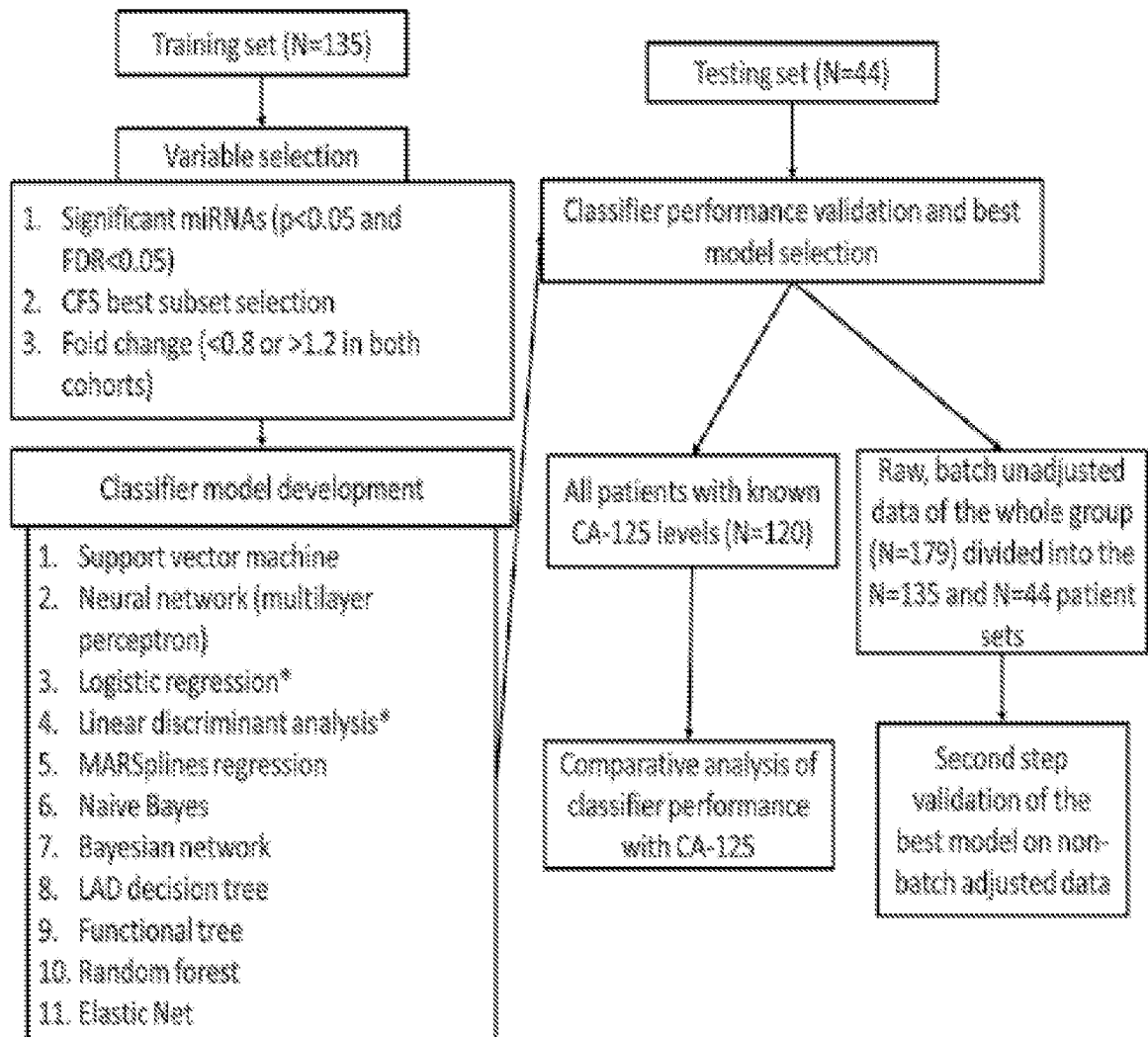


FIG. 1B

3/20

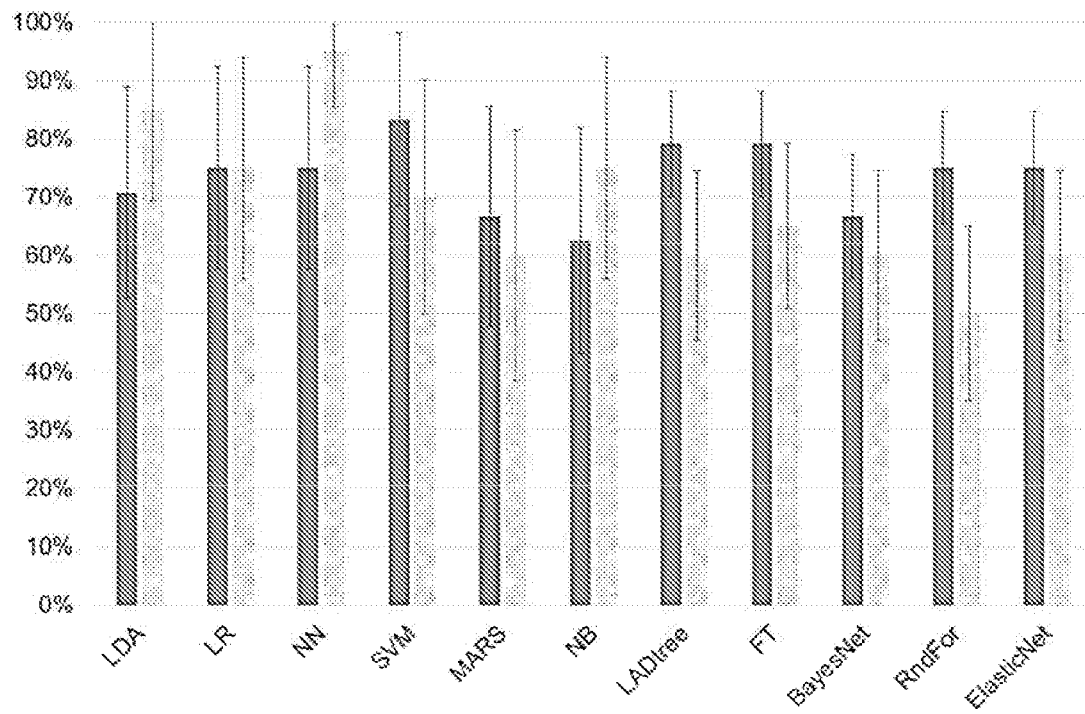


FIG. 2A

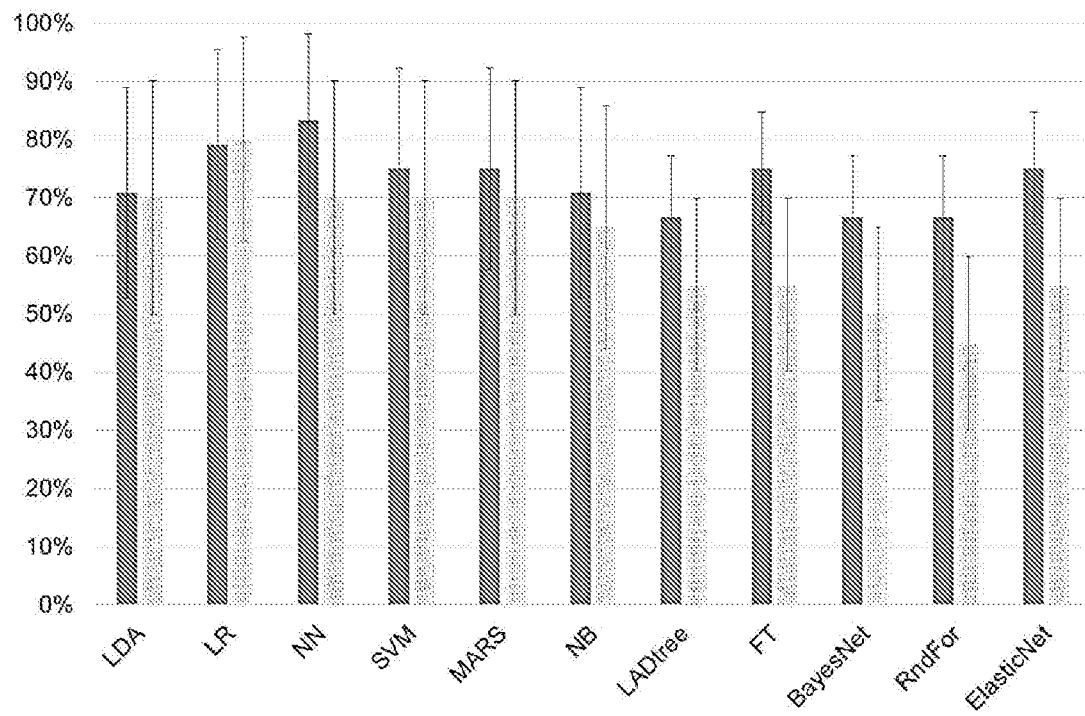


FIG. 2B

4/20

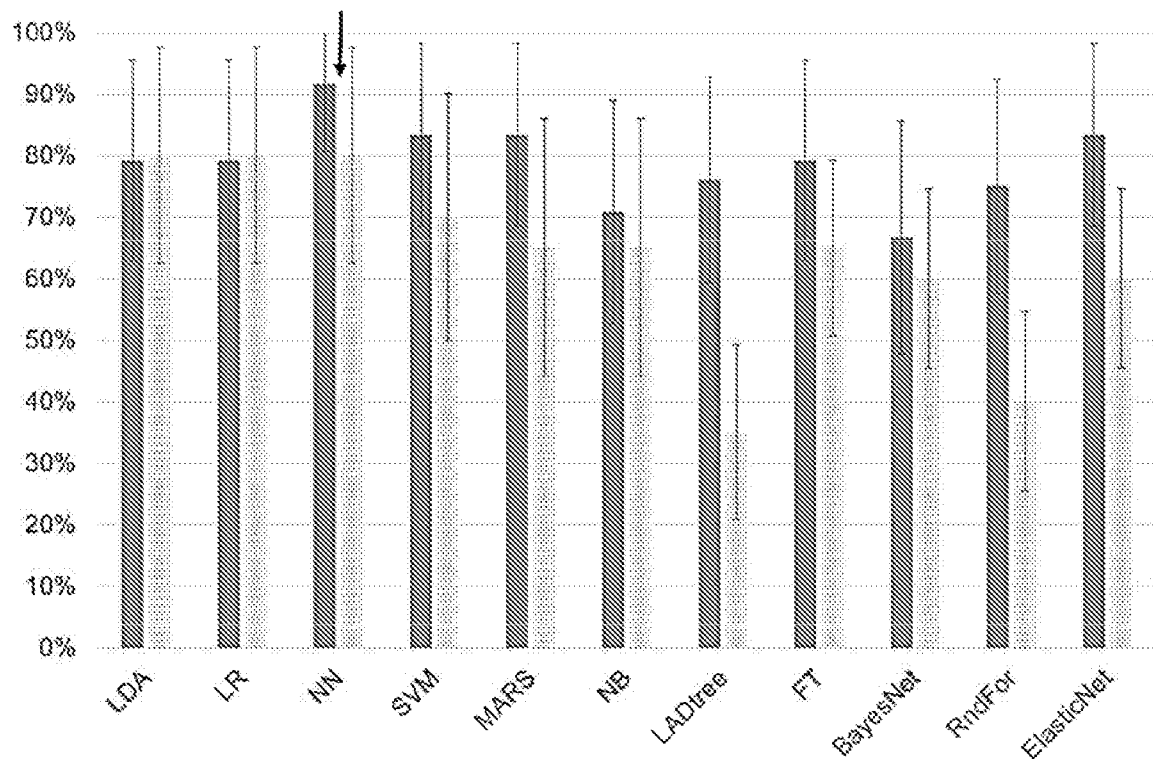


FIG. 2C

5/20

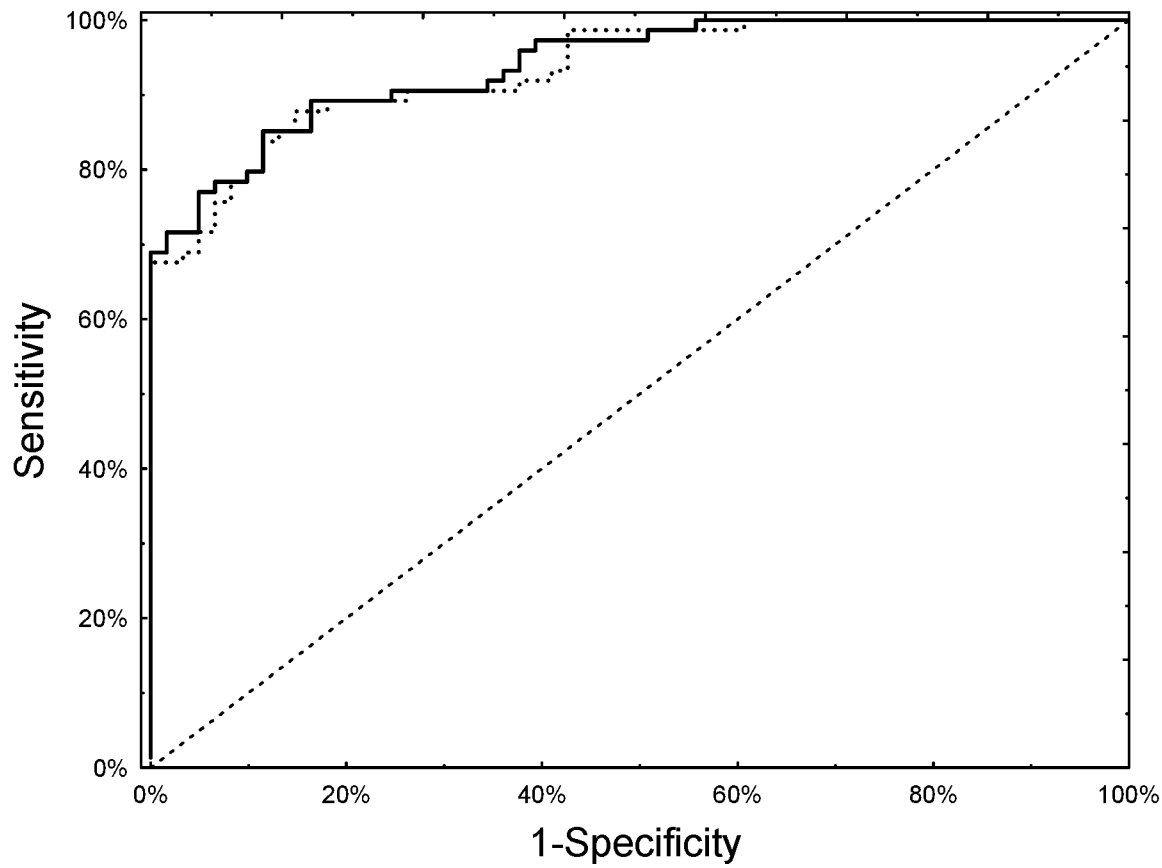


FIG. 3A

6/20

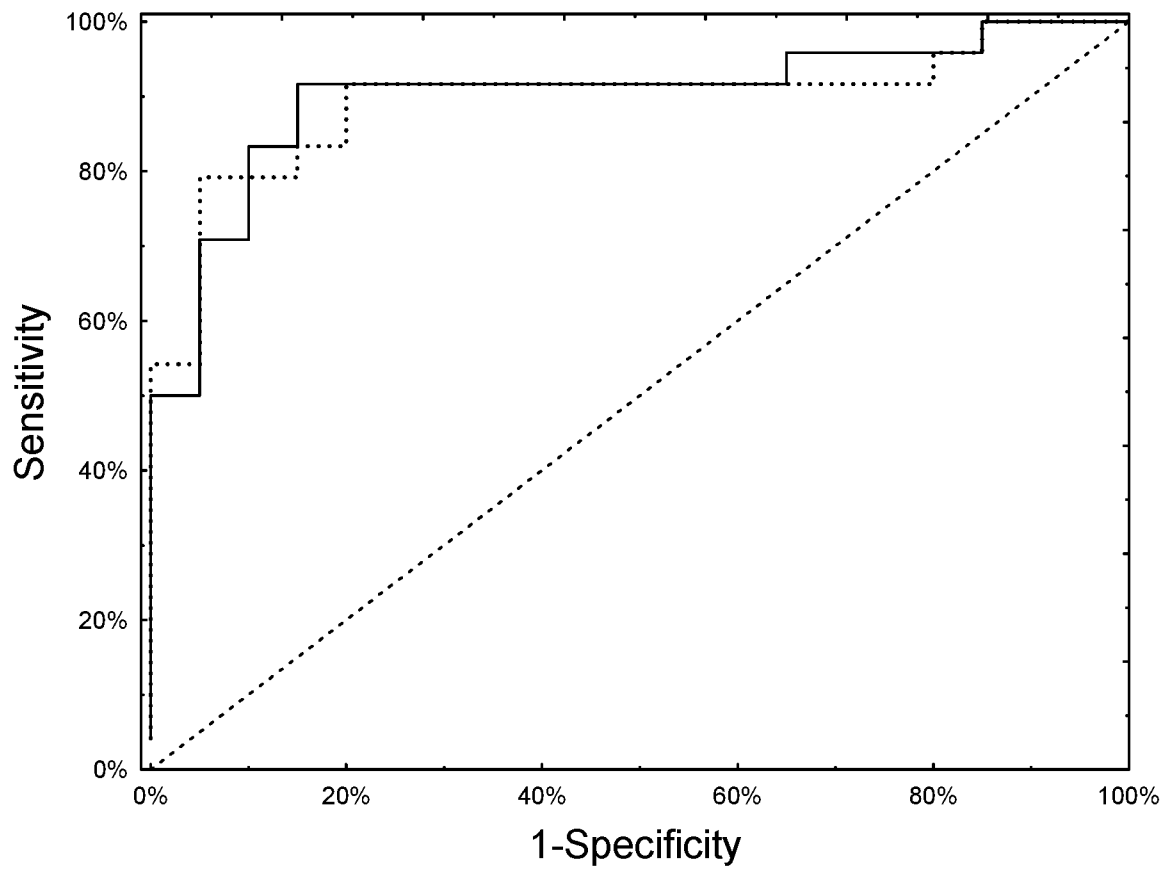


FIG. 3B

7/20

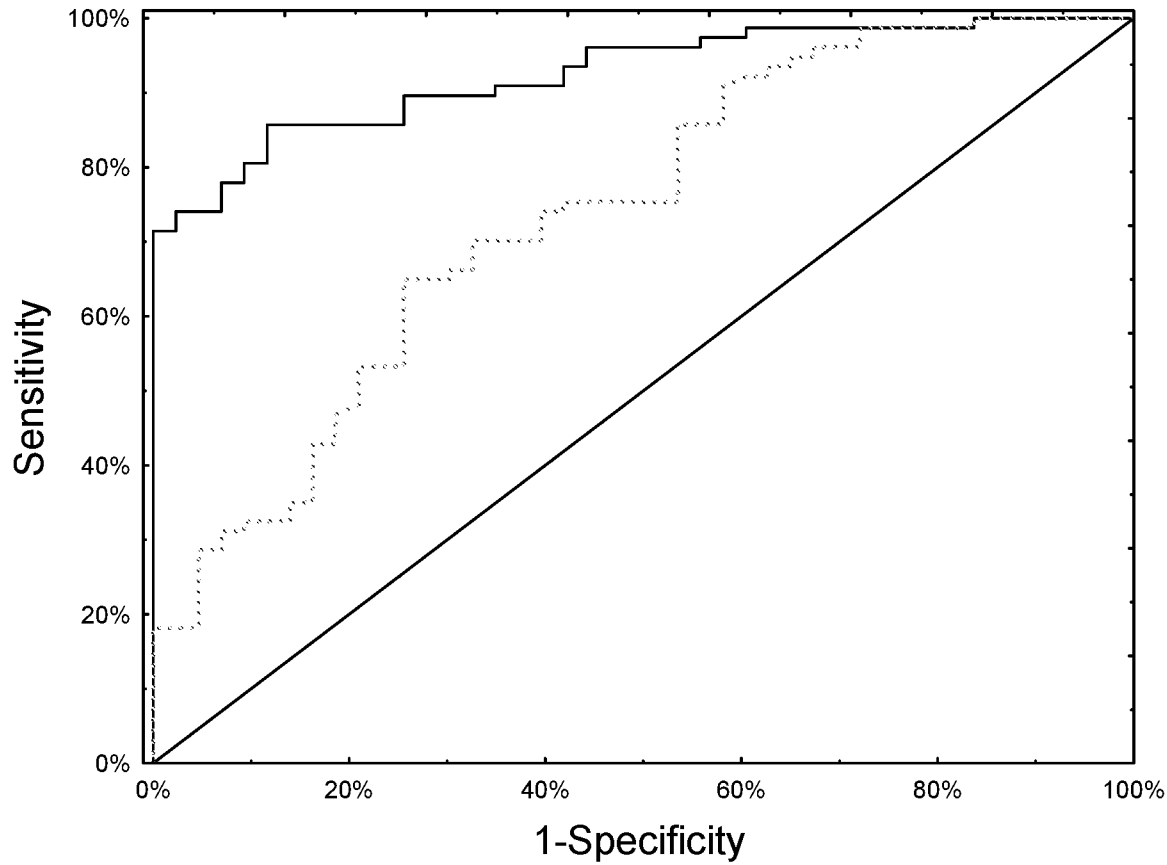


FIG. 4A

8/20

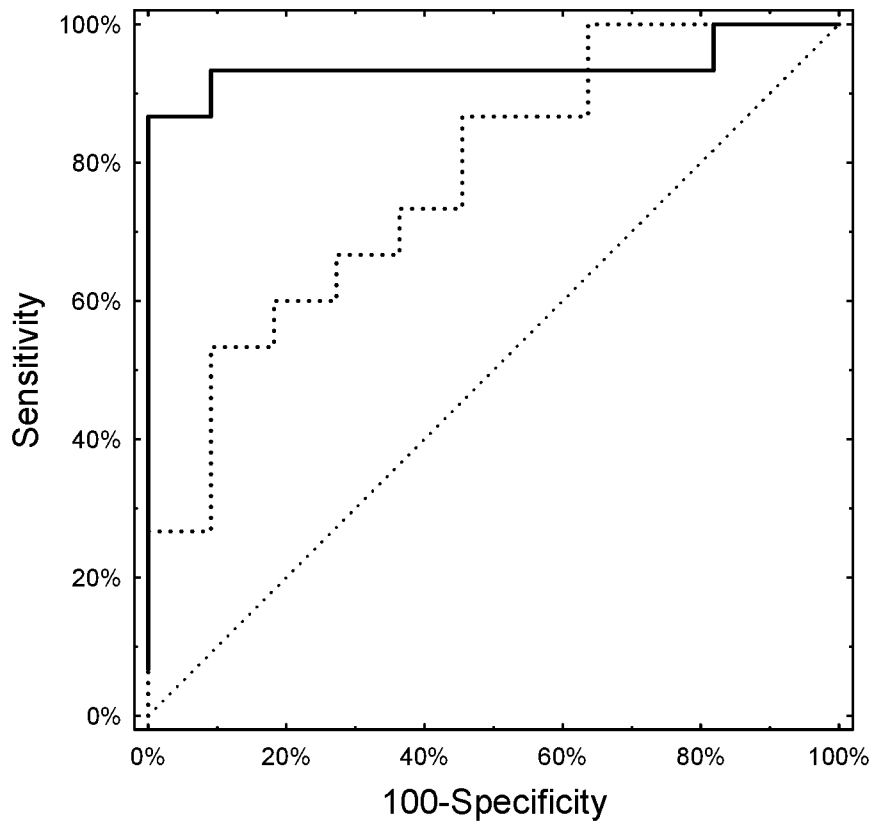


FIG. 4B

9/20

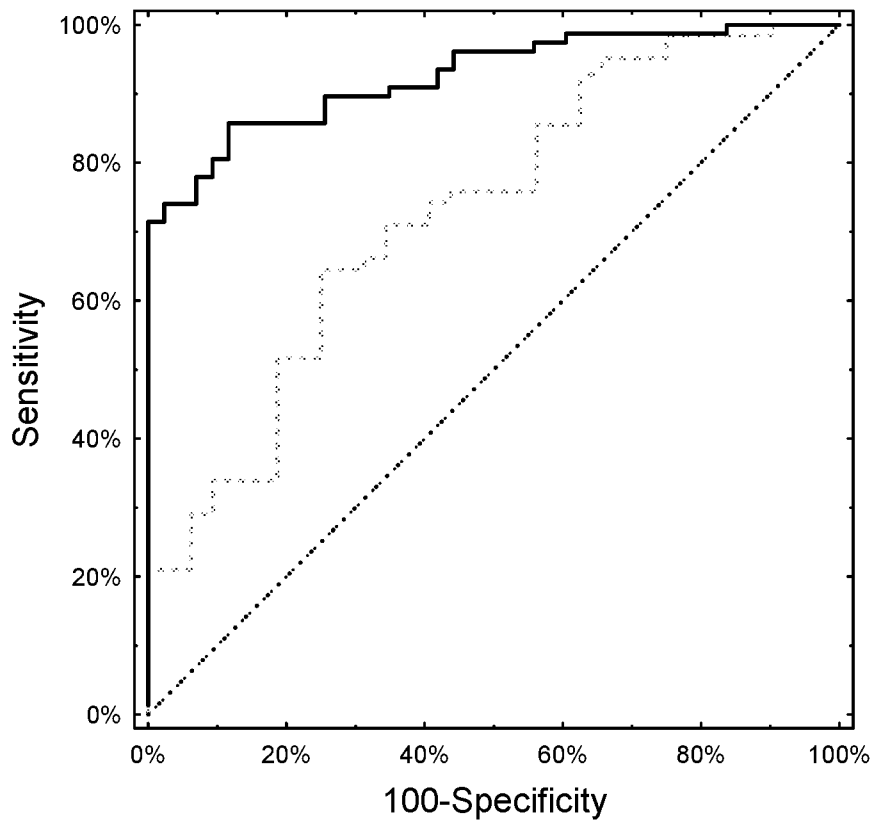


FIG. 4C

10/20

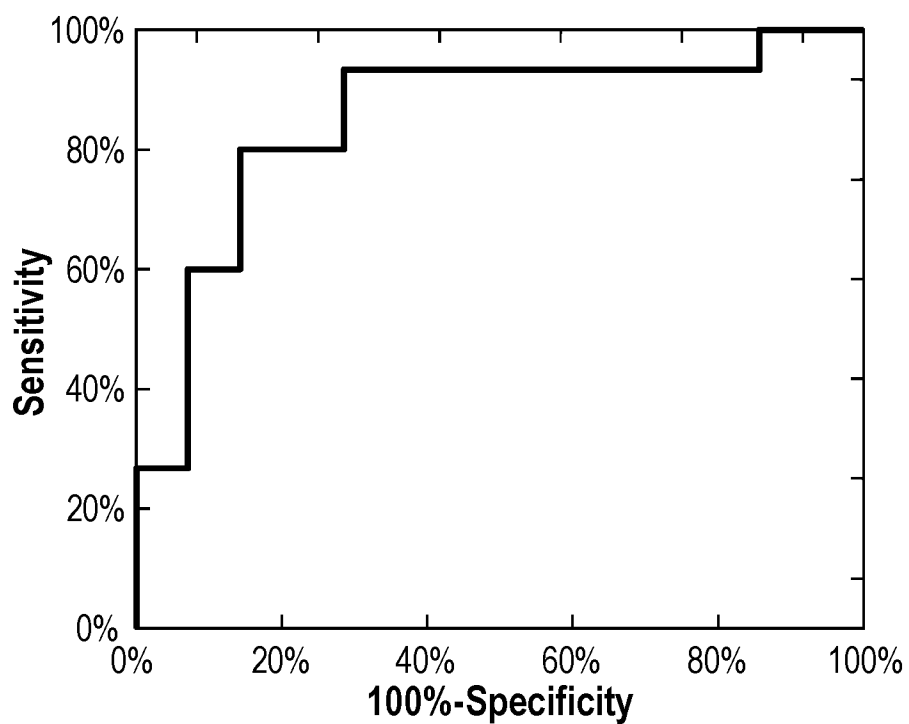


FIG. 5A

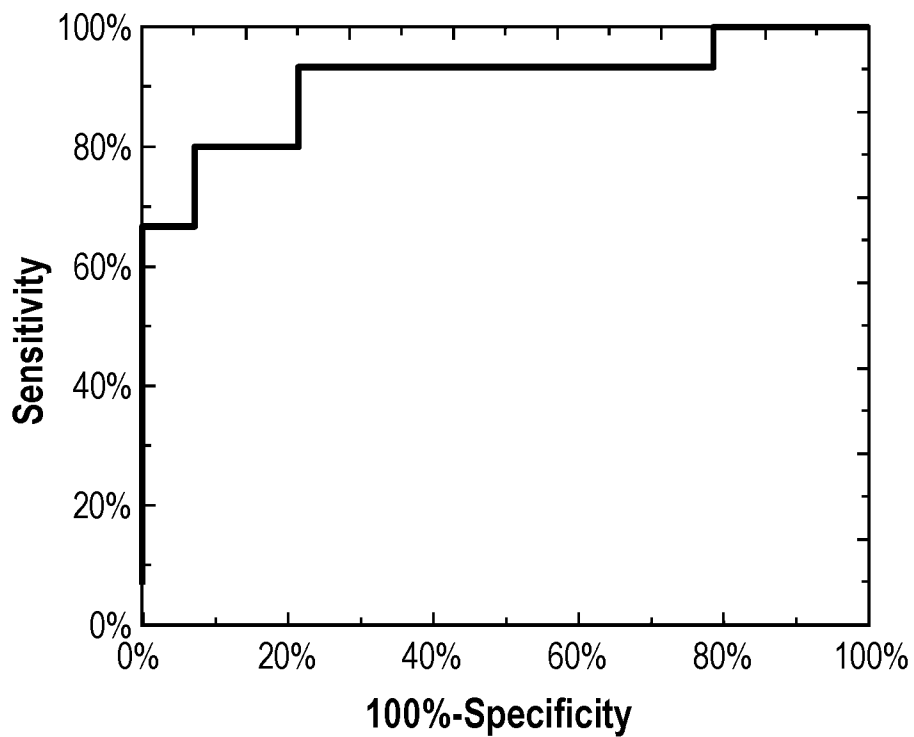


FIG. 5B

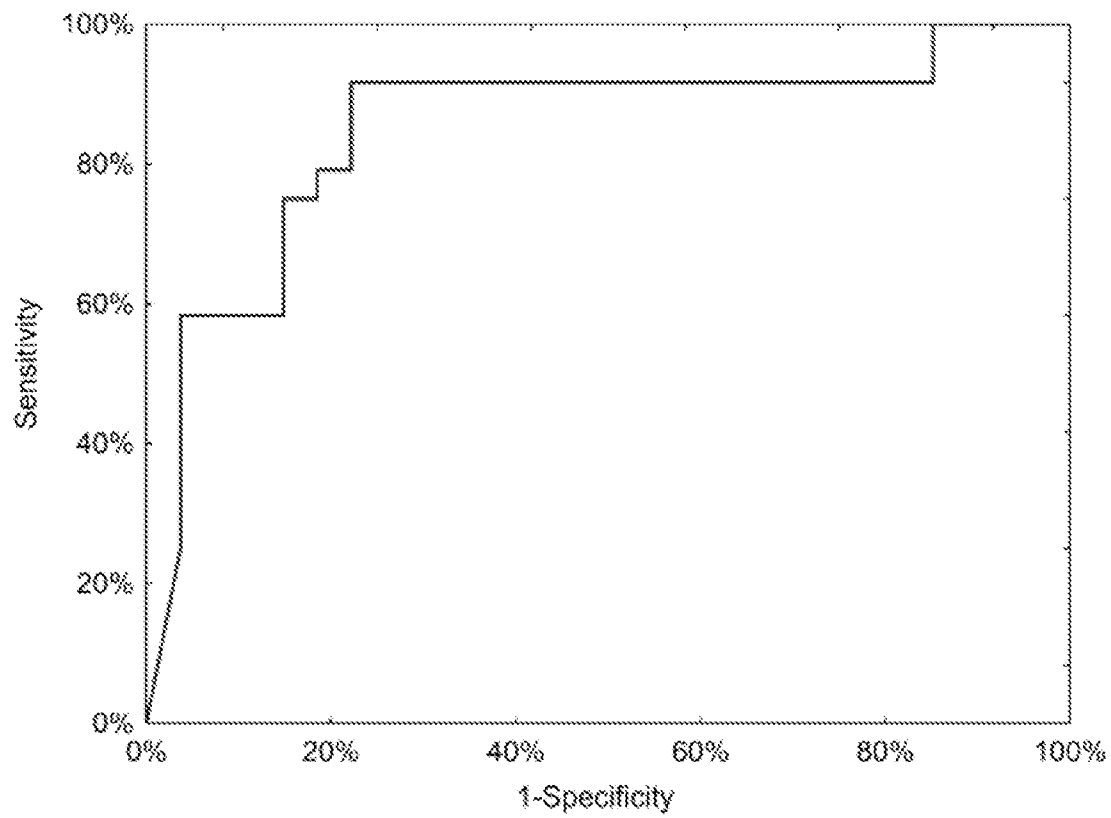


FIG. 6

12/20

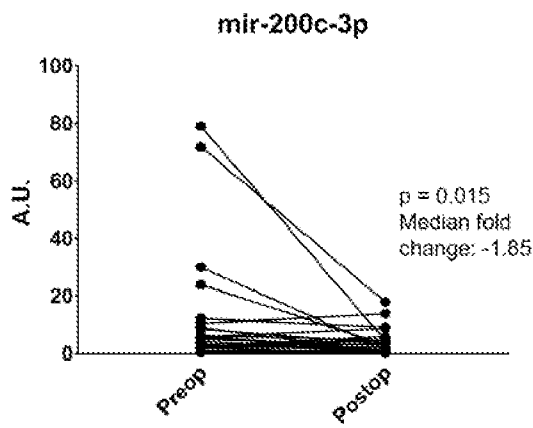


FIG. 7A

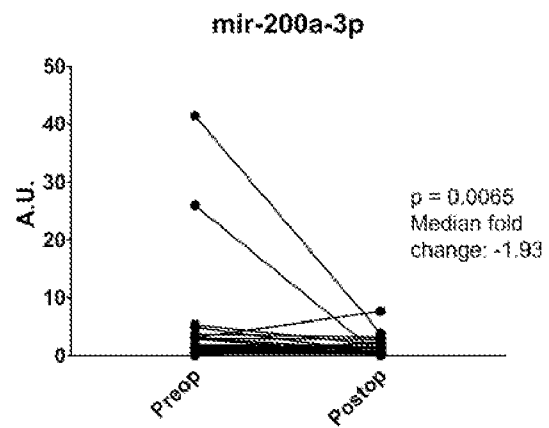


FIG. 7B

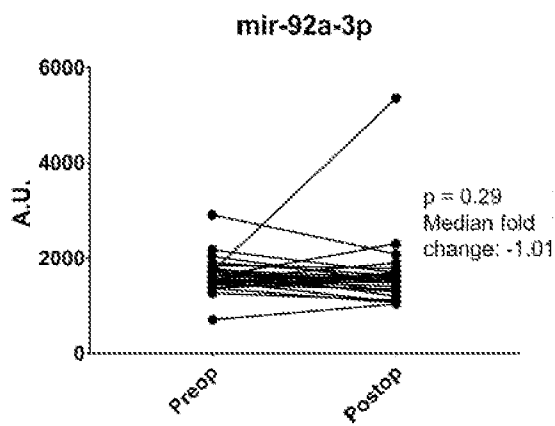


FIG. 7C

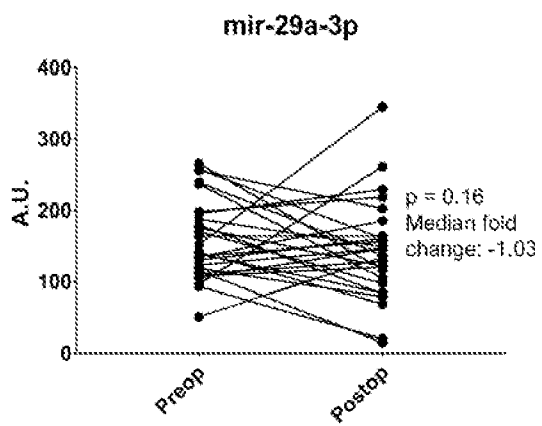


FIG. 7D

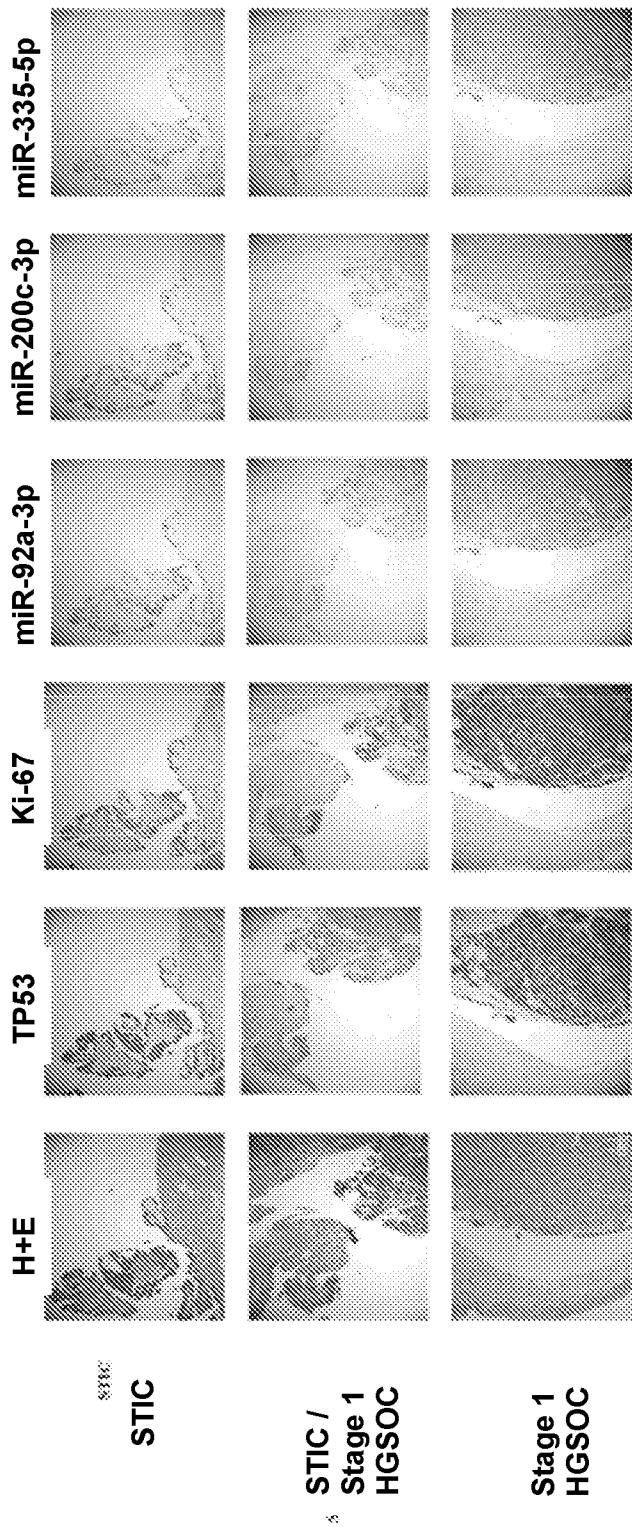
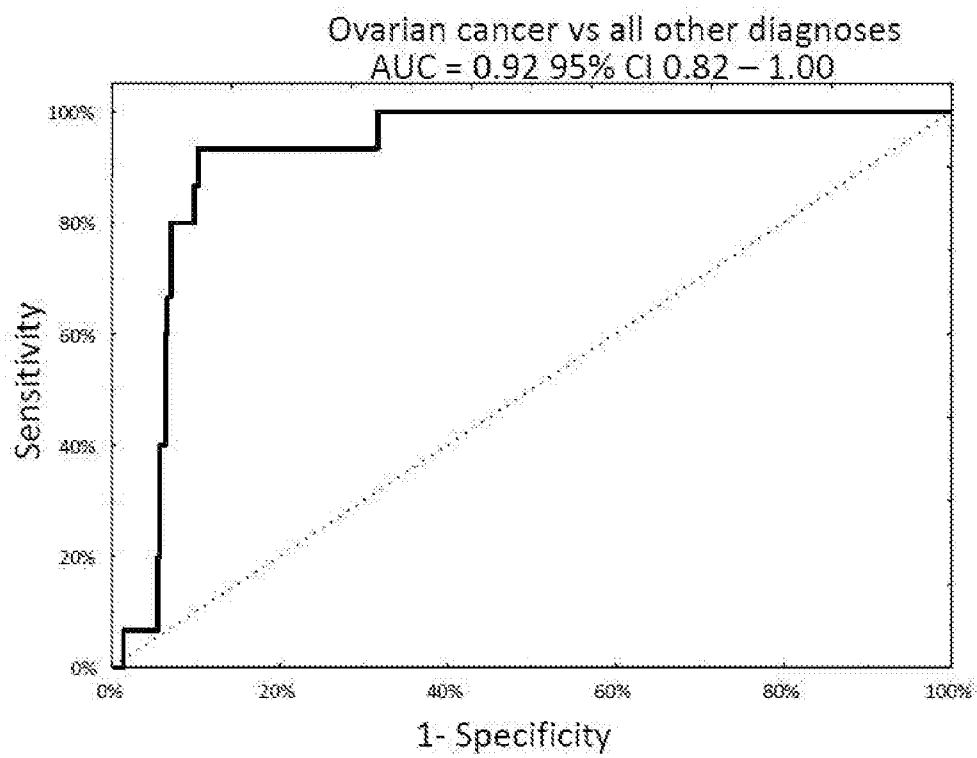


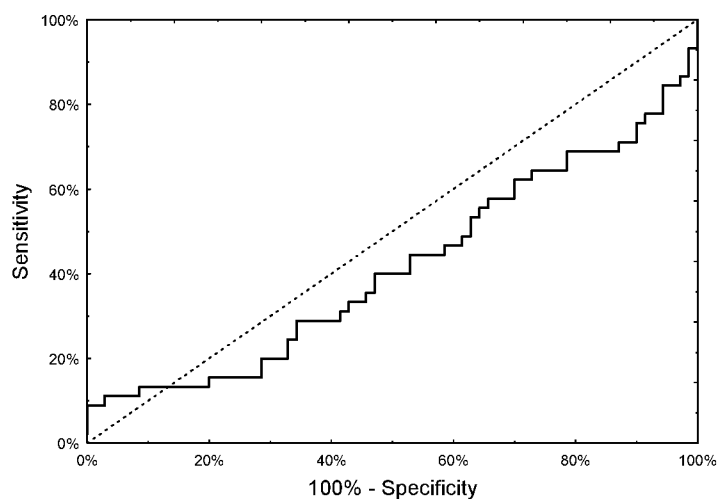
FIG. 8

14/20

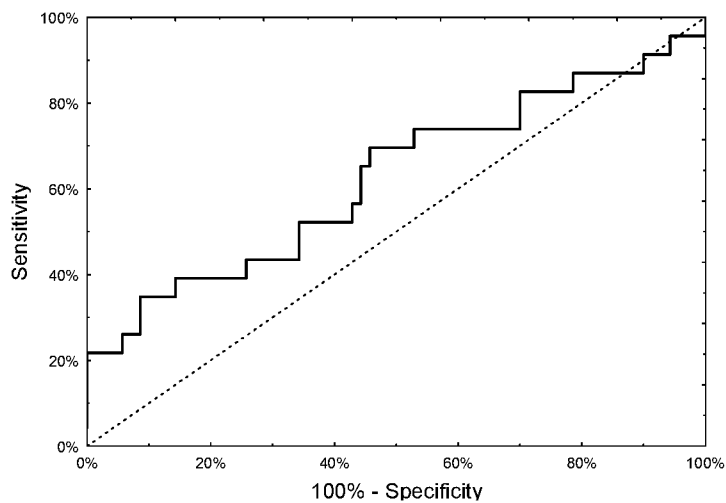
**FIG. 9**

15/20

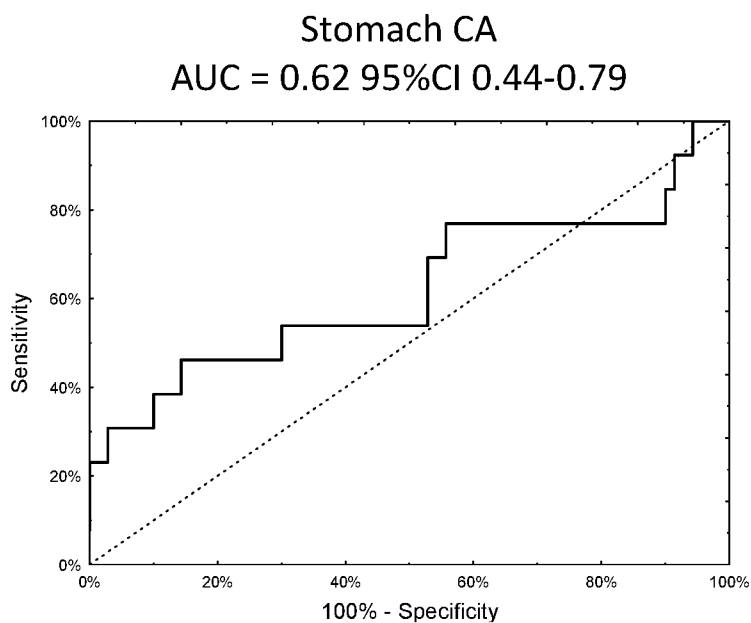
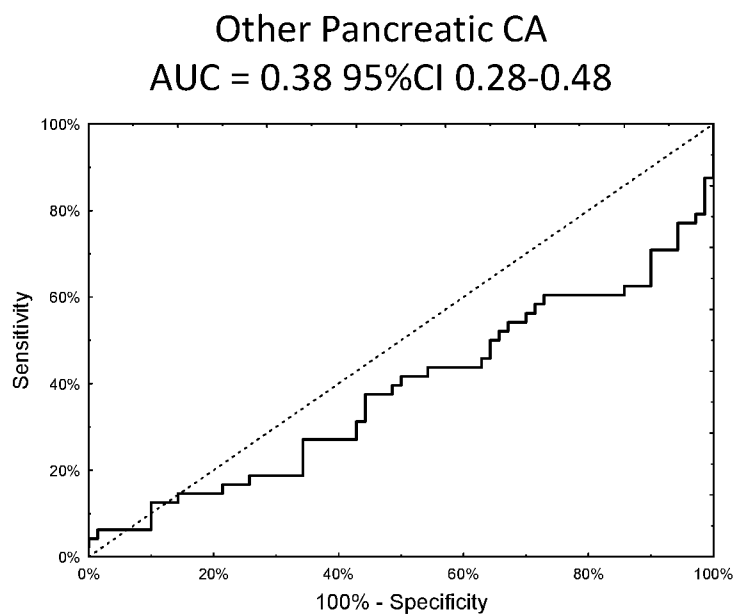
Pancreatic Ductal CA
AUC = 0.41 95%CI 0.31-0.51

**FIG. 10A**

Prostate CA
AUC = 0.62 95%CI 0.48-76

**FIG. 10B**

16/20

**FIG. 10C****FIG. 10D**

17/20

Melanoma CA
AUC = 0.63 95%CI 0.51-0.76

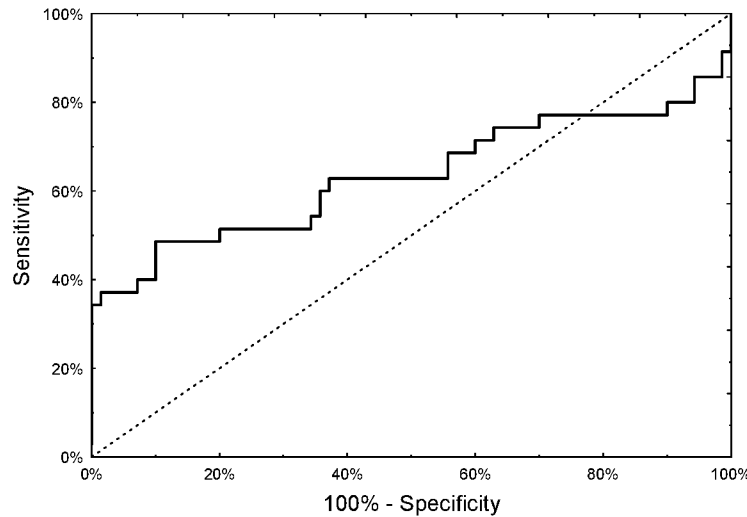


FIG. 10E

Lung CA
AUC = 0.51 95%CI 0.39-0.53

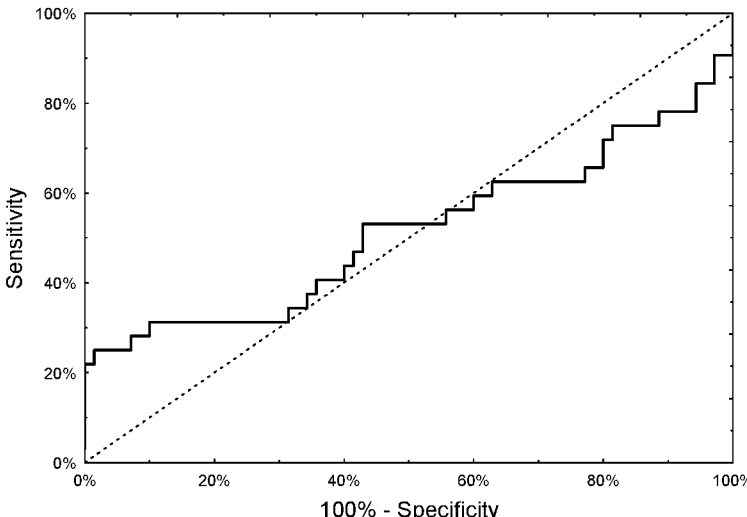
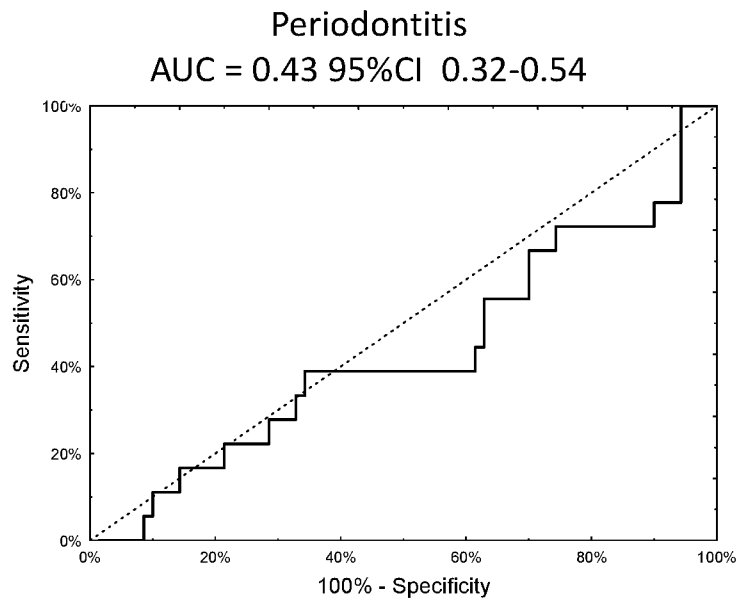
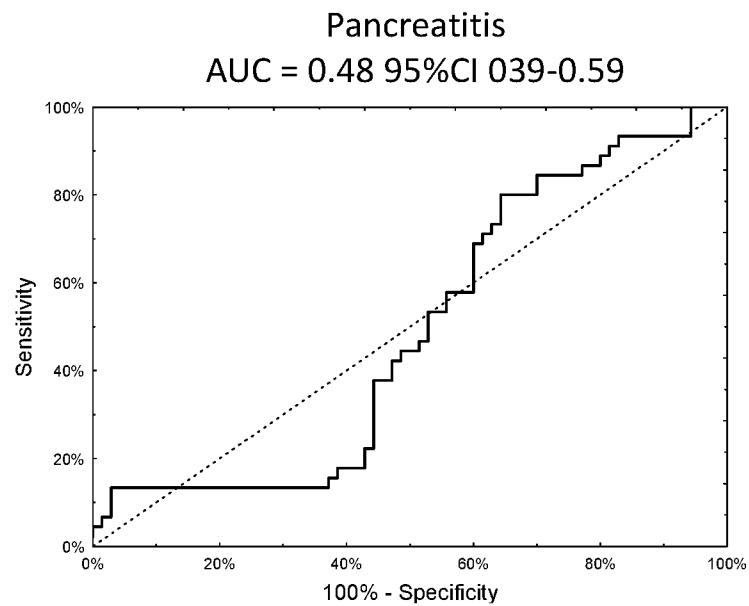


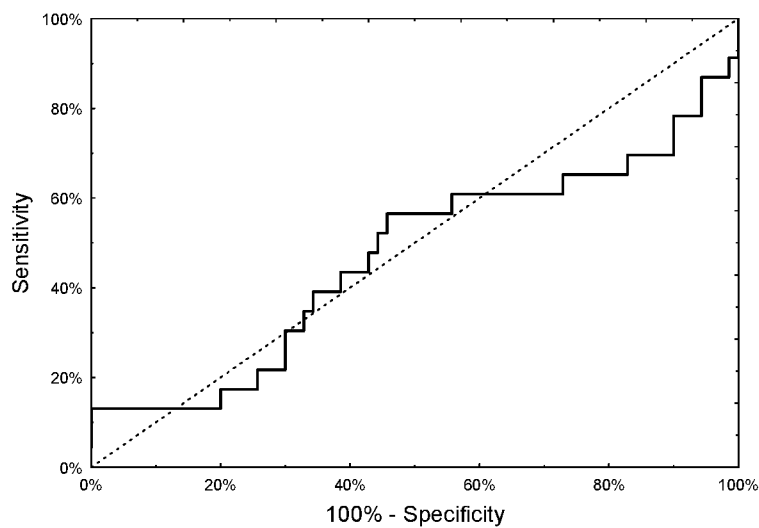
FIG. 10F

18/20

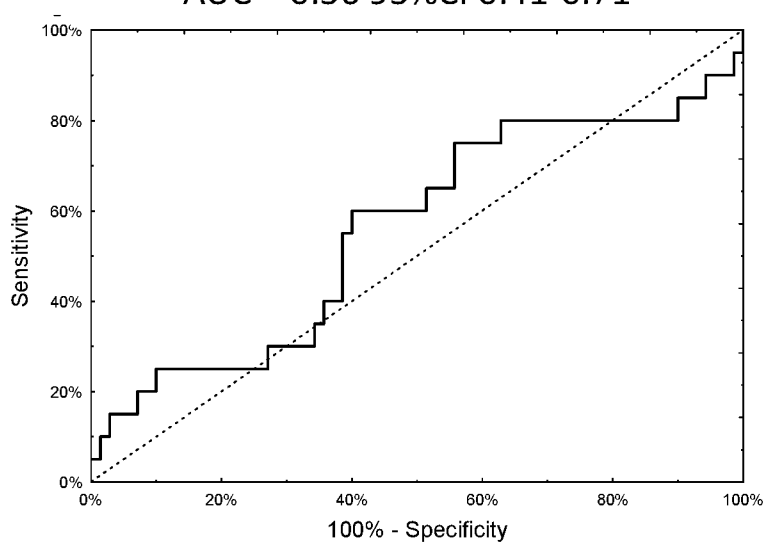
**FIG. 10G****FIG. 10H**

19/20

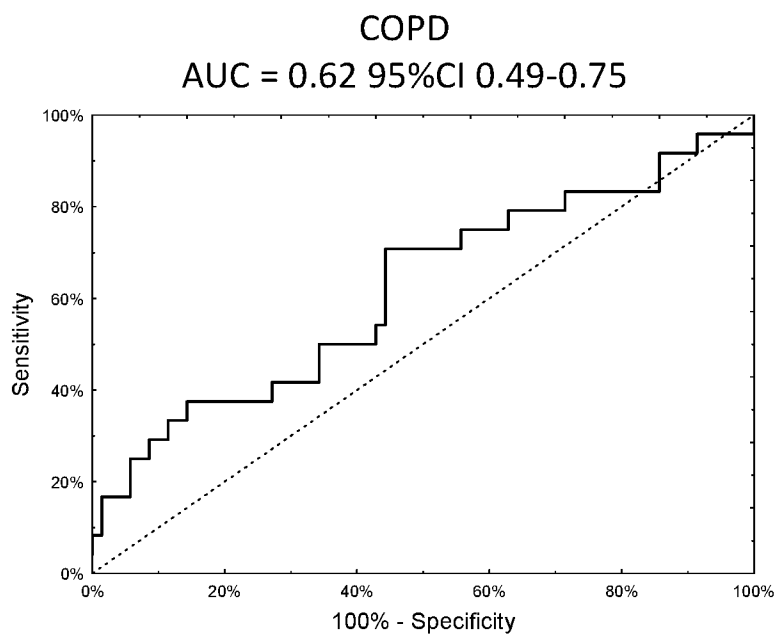
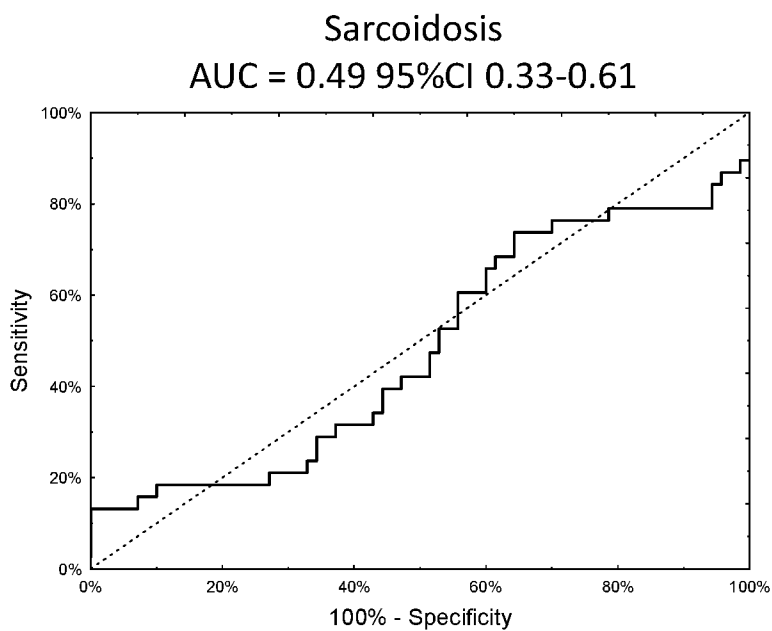
Multiple Sclerosis
AUC = 0.47 95%CI 0.29-0.58

**FIG. 10I**

Acute MI
AUC = 0.56 95%CI 0.41-0.71

**FIG. 10J**

20/20

**FIG. 10K****FIG. 10L**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 18/12982

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/68, A61P 35/00 (2018.01)

CPC - C12Q 2600/178, C12Q 2600/118, C12Q 2600/158, C12Q 2600/16, C12Q 1/6886

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)

See Search History Document

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

See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2010/0249213 A1 (CROCE) 30 September 2010 (30.09.2010) Abstract; Claim 1; Claim 2; para [0018]; para [0030]; para [0170]	1, 44
A	US 2005/0261218 A1 (ESAU et al.) 24 November 2005 (24.11.2005) entire document	1, 44

☐ 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

9 May 2018

Date of mailing of the international search report

07 JUN 2018

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 18/12982

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 7-29, 36-43, 50-76
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

---Please see continuation in first extra sheet -----

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1 and 44, limited to hsa-miR-29a-3p (SEQ ID NO: 39)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 18/12982

Continuation of Box No. III. Observations where unity of invention is lacking.

Group I+: Claims 1-6, 44-49 drawn to methods for diagnosing or treating ovarian cancer in a female subject. The methods will be searched to the extent that the methods encompasses determining the amount of one test microRNA, wherein the microRNA is hsa-miR-29a-3p (SEQ ID NO: 39). It is believed that claims 1 and 44, limited to hsa-miR-29a-3p (SEQ ID NO: 39), encompass this first named invention, and thus these claims will be searched without fee to the extent that they encompass one test microRNA, wherein the microRNA is hsa-miR-29a-3p (SEQ ID NO: 39). Additional microRNAs will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected microRNA(s). Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched. An exemplary election would be wherein the microRNA encompasses two test microRNAs, wherein the microRNAs are hsa-miR-29a-3p (SEQ ID NO: 39) and hsa-miR-92a-3p (SEQ ID NO: 47), i.e. claims 1 and 44, limited to hsa-miR-29a-3p (SEQ ID NO: 39) and hsa-miR-92a-3p (SEQ ID NO: 47).

Group II+: Claims 30-35, drawn to a kit comprising test probes. Groups II+ will be searched upon payment of additional fees. The kit may be searched for an additional fee and election as such, for example, to the extent that the test probe specifically hybridizes to one microRNA, wherein the microRNA encompasses hsa-miR-92a-3p (SEQ ID NO: 47), i.e. claim 30, limited to hsa-miR-92a-3p (SEQ ID NO: 47). Additional microRNAs will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected microRNA(s). Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched. Another exemplary election would be wherein the test probes specifically hybridize to two microRNA, wherein the microRNAs are hsa-miR-92a-3p (SEQ ID NO: 47) and hsa-miR-450b-5p (SEQ ID NO: 149), i.e. claim 30, limited to hsa-miR-92a-3p (SEQ ID NO: 47) and hsa-miR-450b-5p (SEQ ID NO: 149).

The inventions listed as Groups I+, II+ do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features

Groups I+ require method steps for diagnosing or treating ovarian cancer, not required by Groups II+. No technical features are shared between the microRNA sequences of the inventions listed as Groups I+ and, accordingly, these groups lack unity a priori. Additionally, even if the inventions listed as Groups I+ were considered to share technical features, these shared technical features are previously disclosed by the prior art, as further discussed below.

Groups II+ requires a kit composition comprising probes, not required by Groups I+. No technical features are shared between the microRNA sequences of the inventions listed as Groups II+ and, accordingly, these groups lack unity a priori. Additionally, even if the inventions listed as Groups II+ were considered to share technical features, these shared technical features are previously disclosed by the prior art, as further discussed below.

Common Technical Features

The feature shared by Groups I+ and II+ is microRNA associated with ovarian cancer.

The feature shared by the inventions listed as Groups I+ is a method for diagnosing ovarian cancer in a female subject, the method comprising:

- (a) obtaining a sample collected from the subject;
- (b) determining the amounts of one or more test microRNAs
- (c) comparing the amounts of the test microRNAs determined in step (b) to a statistical model, thereby diagnosing ovarian cancer in the subject.

Another feature shared by the inventions listed as Groups I+ is a method for treating a female subject suspected of having an ovarian cancer, the method comprising

- (a) obtaining a sample collected from the subject;
- (b) determining the amounts of one or more test microRNAs
- (c) comparing the amounts of the test microRNAs determined in step (b) to a statistical model,
- (d) treating the subject for ovarian cancer, or monitor the subject for ovarian cancer, based on the comparison of step (c).

The feature shared by the inventions listed as Groups II+ is a kit comprising at least one test probe capable of specifically hybridizing to a microRNA.

However, these shared technical features do not represent a contribution over prior art, because the shared technical features are taught by US 2016/0312301 A1 to the Feinstein Institute for Medical Research (hereinafter 'FIMR').

FIMR discloses microRNA associated with ovarian cancer (para [0015] - "The present invention relates methods of providing a diagnosis for a subject who may have ovarian cancer, or a prognosis for a subject having been diagnosed with ovarian cancer. The diagnosis or prognosis is provided by identifying the presence of particular microRNA which the inventors have associated with ovarian cancer").

----Please see continuation in next extra sheet -----

Continuation of Box No. III. Observations where unity of invention is lacking.

FIMR discloses a method for diagnosing ovarian cancer in a female subject (para [0030] - "One aspect of the invention provides a method of diagnosing ovarian cancer in a subject"; note: only females have ovaries, therefore all subjects having ovarian cancer are female subjects), the method comprising:

(a) obtaining a sample collected from the subject (para [0030] - "a biological sample from the subject"; para [0041] - "The level of the target miR gene product is measured in a biological sample obtained from a subject");
(b) determining the amounts of one or more test microRNAs (para [0030] - "The method includes identifying the expression level of at least one diagnostic miRNA . . . in a biological sample from the subject");
(c) comparing the amounts of the test microRNAs determined in step (b) to a statistical model, thereby diagnosing ovarian cancer in the subject (para [0030] - "comparing the expression level of the at least one diagnostic miRNA to control expression levels, and diagnosing the subject as having or being at an increased risk of having ovarian cancer if the subject has a changed expression level of the one or more diagnostic miRNA"; para [0082] - "Comparisons between experimental groups were performed using a moderated t-test from LIMMA and p-values were adjusted for multiple testing via the Benjamini-Hochberg method. . . . A Benjamini-Hochberg adjusted p-value<0.05 was considered statistically significant in comparisons between different groups for each miRNA.").

FIMR discloses a method for treating a female subject suspected of having an ovarian cancer (para [0057] - "In some embodiments, the method of diagnosis or prognosis can be used to indicate or guide treatment of ovarian cancer"), the method comprising

(a) obtaining a sample collected from the subject (para [0030] - "a biological sample from the subject"; para [0041] - "The level of the target miR gene product is measured in a biological sample obtained from a subject");
(b) determining the amounts of one or more test microRNAs (para [0030] - "The method includes identifying the expression level of at least one diagnostic miRNA . . . in a biological sample from the subject");
(c) comparing the amounts of the test microRNAs determined in step (b) to a statistical model (para [0030] - "comparing the expression level of the at least one diagnostic miRNA to control expression levels, and diagnosing the subject as having or being at an increased risk of having ovarian cancer if the subject has a changed expression level of the one or more diagnostic miRNA"; para [0082] - "Comparisons between experimental groups were performed using a moderated t-test from LIMMA and p-values were adjusted for multiple testing via the Benjamini-Hochberg method. . . . A Benjamini-Hochberg adjusted p-value<0.05 was considered statistically significant in comparisons between different groups for each miRNA."),
(d) treating the subject for ovarian cancer, or monitor the subject for ovarian cancer, based on the comparison of step (c) (para [0057] - "the method of diagnosis or prognosis can be used to indicate or guide treatment of ovarian cancer. For example, the method can further comprise the step of providing suitable treatment if the subject is identified as being at increased risk. Suitable treatment can include use of a variety of different methods of treating cancer, such as surgery, radiation therapy, and administration of hormonal or anticancer agents.").

FIMR discloses a kit comprising at least one test probe capable of specifically hybridizing to a microRNA (para [0063] - "A kit for diagnosing ovarian cancer in a subject includes one or more detector means specific for at least one diagnostic miRNA. The detector means can include primers or probes designed to detect one or more diagnostic miRNA").

As the technical features were known in the art at the time of the invention, they cannot be considered special technical features that would otherwise unify the groups.

Groups I+, II+ therefore lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature.