



OPEN

Characterization of miR-200 family members as blood biomarkers for human and laying hen ovarian cancer

Pui-Wah Choi^{1,2,3}, Abbas Bahrapour^{4,5}, Shu-Kay Ng⁴, Sze Kei Liu³, Wei Qiu⁶, Fang Xie⁶, Winston Patrick Kuo⁶, Joseph Kwong⁷, Karen H. Hales^{8,9}, Dale B. Hales^{8,9}, Kwong-Kwok Wong¹⁰, Errol R. Norwitz¹¹, Chun Kin Chow^{2✉}, Ross S. Berkowitz¹ & Shu-Wing Ng^{1,12✉}

MicroRNA-200 (miR-200) family is highly expressed in ovarian cancer. We evaluated the levels of family members relative to the internal control miR-103a in ovarian cancer and control blood specimens collected from American and Hong Kong Chinese institutions, as well as from a laying hen spontaneous ovarian cancer model. The levels of miR-200a, miR-200b and miR-200c were significantly elevated in all human cancer versus all control blood samples. Further analyses showed significantly higher miR-200 levels in Chinese control (except miR-429) and cancer (except miR-200a and miR141) samples than their respective American counterparts. Subtype-specific analysis showed that miR-200b had an overall elevated level in serous cancer compared with controls, whereas miR-429 was significantly elevated in clear cell and endometrioid cancer versus controls. MiR-429 was also significantly elevated in cancer versus control in laying hen plasma samples, consistent with the fact that endometrioid tumor is the prevalent type in this species. A neural network model consisting of miR-200a/200b/429/141 showed an area under the curve (AUC) value of 0.904 for American ovarian cancer prediction, whereas a model consisting of miR-200b/200c/429/141 showed an AUC value of 0.901 for Chinese women. Hence, miR-200 is informative as blood biomarkers for both human and laying hen ovarian cancer.

According to a report published in 2018 by World Ovarian Cancer Coalition, the worldwide number of ovarian cancer cases by 2035 are predicted to increase to 371,000 per year, and deaths will increase by 67% to 254,000¹. In developed countries, ovarian cancer remains as the deadliest gynecologic malignancy^{2,3}. About 20% of women with epithelial ovarian cancer (EOC) are diagnosed at the early stage and the 5-year survival rates can be as high as 90%. In contrast, the 5-year survival rate is around 17–39% when the disease is diagnosed at the advanced stages⁴. Among all the biomarkers, cancer antigen-125 (CA-125) has been the biomarker of choice for managing ovarian cancer patients. However, serum CA125 level has low sensitivity for early diagnosis and the level is affected by conditions such as menstruation, pelvic inflammations, endometriosis, pregnancy, smoking, and intake of caffeine^{5,6}. Other biomarkers including glycoprotein human epididymis protein 4, transthyretin,

¹Department of Obstetrics/Gynecology and Reproductive Biology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. ²MedTimes Molecular Laboratory Limited, Medtimes Medical Group Limited, Unit B, 7/F Roxy Industrial Centre, Kwai Chung, Hong Kong, China. ³WomenX Biotech Limited, Kowloon, Hong Kong, China. ⁴School of Medicine and Menzies Health Institute Queensland, Griffith University, Nathan, QLD 4111, Australia. ⁵Faculty of Health, Kerman University of Medical Sciences, Kerman, Iran. ⁶CloudHealth Genomics Limited, Shanghai, China. ⁷Department of Obstetrics and Gynecology, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, Hong Kong, China. ⁸Department of Obstetrics/Gynecology, Southern Illinois School of Medicine, Carbondale, IL 62901, USA. ⁹Department of Physiology, Biochemistry and Molecular Biology, Southern Illinois School of Medicine, Carbondale, IL 62901, USA. ¹⁰Department of Gynecologic Oncology and Reproductive Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ¹¹Newton-Wellesley Hospital, Boston, MA 02462, USA. ¹²Present address: Tufts Medical Center, Tufts University School of Medicine, 800 Washington Street, Boston, MA 02111, USA. ✉email: isaac@medtimes.com.hk; sng1@tuftsmedicalcenter.org

apolipoprotein A1, beta2-microglobulin, and transferrin showed unsatisfactory predictive value^{5,7,8}. Therefore, novel biomarkers with high sensitivity and specificity are urgently needed for ovarian cancer diagnosis.

Biomarkers are more informative and may be more specific when their expression alterations are directly caused by genetic and molecular mechanisms underlying the pathophysiological development of the diseases. Mechanism-based biomarkers have utility in the identification of the origin of the disease, differentiation of disease types, and the application of more effective therapies^{9,10}. MicroRNAs (miRNAs) play an active role in physio- and pathological processes and can reflect physiological alterations more directly¹¹. Indeed, the microRNA profiles in blood exhibited a strong correlation with that in the tumor¹². In addition, miRNAs are generally more stable than mRNAs in blood¹¹ and are good candidates for mechanism-based biomarkers.

MicroRNA-200 (miR-200) family is a master mesenchymal–epithelial transition regulator controlling the expression of epithelial marker E-cadherin¹³. Previously we have shown that the expression of miR-200 and E-cadherin is elevated in human ovarian cancer^{14,15}, as well as in laying hen spontaneous ovarian cancer model¹⁶, the only non-primate animal that develops ovarian adenocarcinoma in nature. Human miR-200 family consists of five members (miR-200a, miR-200b, miR-200c, miR-429, and miR-141) scattered in two separate chromosomal clusters¹⁷. Interestingly, only the miR-200a, miR-200b, and miR-429 cluster is present in the chicken genome. We have completed a functional study of miR-200, demonstrating that this miRNA family governs inclusion cyst formation in endosalpingiosis, a benign lesion in association with serous ovarian neoplasm, and collective movement involved in ovarian cancer spread¹⁸. Intriguingly, ectopic expression of miR-200 in normal human ovarian surface epithelial cells induced upregulation of steroid hormone pathway genes¹⁸, which, in conjunction with frequent ovulation of the laying hens, are consistent with the incessant ovulation hypothesis for ovarian cancer development¹⁹, and highlight the critical role miR-200 plays in ovarian carcinogenesis²⁰. Taken together, miR-200 can be an excellent mechanism-based biomarker for ovarian cancer diagnosis and prognosis.

There have been multiple studies that investigated miR-200 in various diseases such as endometriosis²¹, renal fibrosis²², pleural effusion²³ and colon cancer²⁴. There were also a few studies that evaluate miR-200 as a diagnostic or prognostic marker for ovarian cancer (Table S1). Taylor et al. used a microarray platform to compare the miRNA profiles of tumor-derived EpCAM⁺ exosomes isolated from the sera of American women and found that the levels of miR-141, miR-200a, miR-200c, miR-200b, miR-203, miR-205 were elevated in patients with serous papillary adenocarcinoma compared to that isolated from the benign and healthy samples¹². The second study conducted by Kan et al. in Australia found that miR200a, miR200b and miR200c were highly expressed in sera derived from serous cancer patients when compared to healthy control, with miR-200c the most significantly different between the two groups²⁵. In that study, they selected miR-103 from four tested small RNAs as the internal control based on its detectability in sera and no significant differences between the case and control groups. The third study by Zheng et al. was conducted in China and they found nine miRNAs including miR-200a and miR-141 that were detected at higher levels in the plasma from the ovarian cases than in the healthy controls²⁶. The fourth study by Zuberi et al. was conducted in India. The levels of miR200a, miR200b and miR200c were revealed higher in the 70 EOC sera samples than 70 sera sample from the healthy controls. The expression of miR-200a was found to be significantly up-regulated in the mucinous and serous subtype, while the level of miR200a and miR-200c were positively correlated to the stages of EOC and metastasis²⁷. The fifth study by Gao et al. was conducted in China. MiR-200c and miR-141 were found significantly elevated in the EOC patient sera when compared to healthy controls⁷. Racial disparity, differences in the samples size, forms of blood, methods of blood collection and detection are factors that could contribute to the results found in the above studies.

In this study, we compared the levels of all five miR-200 family members in the blood specimens from patients with EOC and the healthy controls collected in the USA, Hong Kong and China. We also evaluated if there are racial disparities in the blood miR-200 levels, as racial differences in microRNA expression were frequently reported for various diseases^{28–30}. As laying hen ovarian cancers express high levels of miR-200 similar to human ovarian cancers^{15,16,18}, the miR-200 levels in the plasma samples collected from the healthy and cancerous laying hens were also compared to those in the human blood specimens.

Results

Elevated levels of miR-200 members in human ovarian cancer blood specimens and ethnic group difference analysis.

Peripheral blood samples were collected initially from a total of 230 female subjects including 112 healthy controls with a median age of 55 and 118 ovarian cancer patients with median age of 57 (Table 1). Clinical characteristics of the ovarian cancer patients are presented in Table S2. The samples were contributed by two US institutions and three sources in Hong Kong/China. All blood samples were processed and reverse transcription-quantitative real-time polymerase chain reactions (RT-qPCR) were performed as described in Materials and Methods. To choose the endogenous normalizer for the miRNA expression, we tested two miRNAs recommended by the manufacturer and, similar to Ken et al.²⁵, chose the same miR-103a-3p as the endogenous reference miRNA to normalize all our RT-qPCR results, as it did not show significant variations in a preliminary test of 40 normal and cancer plasma samples (Figure S1). We noted that serum and whole blood samples were provided by Hong Kong MedTimes and CloudHealth Genomics Ltd, respectively, while the blood samples provided by other institutions were plasmas. We compared the miRNA levels in 10 MedTimes specimens, which were available in all three forms. Statistical analysis of qPCR data showed that the miRNA levels in most serum and plasma specimens corroborated each other, whereas the whole blood samples were significantly different from the other two forms (Figure S2). Glinge et al. have shown that long-term storage of whole blood samples might change miRNA stability³¹. We therefore excluded the 16 CloudHealth whole blood samples from final analysis. miR-103a-3p normalized $-\Delta\text{Ct}$ values³² were used to represent miR-200 levels for all the plasma samples as well as the 60 serum samples from Hong Kong MedTimes. In total there were 96 normal and 118 cancer human specimens for the final analysis.

	Sample type	Normal	Cancer	Subtype distribution of cancer			
				Serous	Endometrioid	Clear cell	Mucinous
Human samples							
American institutions							
Brigham and Women's Hospital	Plasma	10	24	22	2	0	0
MD Anderson Cancer Center	Plasma	25	25	25	0	0	0
Total		35	49	47	2	0	0
Chinese institutions							
Chinese University of Hong Kong	Plasma	1	69	21	16	17	15
CloudHealth genomics	Whole blood	16	0				
Hong Kong MedTimes	50 Serum and 10 (Serum + Plasma + Whole blood)	60	0				
Total		77	69	21	16	17	15
All human samples		112	118	68	18	17	15
Laying hen samples	Plasma	11	15				

Table 1. Sample distribution in this study.

	Normal	Cancer	Fold-change (cancer/normal)	P value [^]
(a) All human samples (N = 214)				
miR-200a	-21.1 (-23.5, -18.8)	-16.2 (-18.3, -14.2)	29.9 (2.3, 390.2)	0.010
miR-200b	-21.3 (-23.7, -18.8)	-15.5 (-17.6, -13.3)	55.3 (3.8, 802.9)	0.003
miR-200c	-17.5 (-19.9, -15.2)	-12.2 (-14.3, -10.1)	40.0 (3.0, 526.8)	0.005
miR-429	-19.4 (-23.6, -15.3)	-16.5 (-18.1, -14.9)	7.4 (0.3, 163.0)	0.203
miR-141	-15.2 (-19.3, -11.1)	-15.8 (-17.4, -14.2)	0.67 (0.03, 14.2)	0.794
(b) American-only samples (N = 84)				
miR-200a	-23.0 (-26.0, -20.0)	-18.8 (-21.2, -16.4)	18.6 (1.3, 267.2)	0.032
miR-200b	-22.0 (-25.0, -18.9)	-17.5 (-19.8, -15.2)	21.7 (1.5, 308.0)	0.023
miR-200c	-20.1 (-23.2, -17.1)	-15.6 (-17.9, -13.3)	23.2 (1.6, 332.0)	0.021
miR-429	-17.7 (-20.8, -14.6)	-18.0 (-20.4, -15.6)	0.82 (0.06, 12.2)	0.885
miR-141	-21.0 (-24.1, -17.8)	-16.5 (-18.9, -14.1)	22.5 (1.4, 357.1)	0.028
	Normal	Cancer	Fold-change (cancer/normal)	P value [#]
(c) Chinese-only samples (N = 130)				
miR-200a	-16.8 (-19.0, -14.7)	-16.0 (-18.1, -14.0)	1.7 (0.2, 13.8)	0.592
miR-200b	-19.4 (-21.8, -17.1)	-13.4 (-15.7, -11.2)	64.1 (7.0, 584.9)	<0.001
miR-200c	-12.3 (-14.5, -10.1)	-10.5 (-12.5, -8.4)	3.6 (0.5, 29.2)	0.222
miR-429	-18.0 (-19.9, -16.0)	-13.6 (-15.4, -11.8)	19.9 (3.2, 122.9)	0.001
miR-141	-10.2 (-12.0, -8.3)	-14.5 (-16.2, -12.7)	0.05 (0.01, 0.29)	0.001

Table 2. Normalized miR-200 family levels (estimated means of $-\Delta Ct$, 95% CI) in (a) All human, (b) American-only, or (c) Chinese-only blood samples. The normalized miRNA levels are presented as $-\Delta Ct$ values of RT-qPCR results. [^]P value using linear regression models with adjustment for institutional effects. [#]P value using linear regression models without adjustment for institutional effects due to confounding with normal and cancer group comparison.

The initial analysis of all 214 human samples using linear regression models with adjustment for institutional effects showed that miR-200a, miR-200b, and miR-200c in cancer samples have significantly higher levels than in normal samples (Table 2 and Figure S3). However, further analysis of the data showed that there were significant differences between American and Chinese samples. Within the 84 American samples, cancer samples have significantly higher miR-200a, miR-200b, miR-200c, and miR-141 levels (in average 18.6-fold, 21.7-fold, 23.2-fold, and 22.5-fold, respectively) than normal samples, whereas within the 130 Chinese samples, only miR-200b and miR-429 levels (64.1-fold and 19.9-fold, respectively) were significantly higher than normal samples, and the levels of miR-141 in cancer blood samples were actually significantly lower (0.05-fold) than in normal samples (Table 2). Within the normal samples, four miR-200 family members except miR-429 were significantly higher in Chinese samples than their American counterparts; and within the cancer samples, the miR-200b, miR-200c, and miR-429 levels were significantly higher in Chinese samples than the American samples (Table S3).

miRNA	Normal	Serous	Clear cell	Endometrioid	Mucinous	P value [^]
miR-200b	-20.2 (7.7)	-10.8* (11.3) 670.5* (10.9, 41418.9)	-17.2 (10.1) 7.9 (0.1, 678.4)	-14.1 (10.6) 66.7 (0.7, 6356.4)	-12.1* (7.0) 267.1* (2.5, 28526.2)	0.001
miR-200c	-12.6 (6.3)	-5.0* (11.3) 198.2* (4.3, 9146.5)	-14.2 (7.9) 0.34 (0.01, 21.48)	-11.0 (8.8) 3.1 (0.04, 211.0)	-13.3 (10.5) 0.63 (0.01, 48.0)	0.005
miR-429	-18.2 (5.2)	-15.1 (7.4) 8.7 (0.3, 272.7)	-12.2* (9.7) 64.9* (1.6, 2698.6)	-11.8* (8.6) 88.8* (2.0, 4020.1)	-15.2 (10.8) 8.4 (0.2, 416.7)	0.011
miR-141	-10.7 (5.7)	-9.2 (8.8) 2.9 (0.1, 71.4)	-17.1* (7.3) 0.01* (0.0004, 0.40)	-17.1* (8.2) 0.01* (0.0004, 0.41)	-16.0* (7.3) 0.03* (0.0007, 0.95)	<0.001

Table 3. Significant miR-200 family levels in human blood samples from different histological types of ovarian cancer versus normal blood samples. Data are mean (standard deviation) in the first row and mean fold-change relative to normal (95% CI) in the second row. [^]P value using ANOVA. Dunnett test relative to Normal (*P value < 0.05).

miRNA	Normal	Cancers	P value [^]
miR-200a	-8.2 (7.4)	-5.5 (9.5) 6.6 (0.6, 70.3)	0.119
miR-200b	-5.4 (4.1)	-1.6 (6.8) 13.2 (0.8, 232.2)	0.078
miR-429	-8.7 (4.7)	-4.6 (4.7) 17.3* (1.4, 208.5)	0.025

Table 4. Comparison of miR-200 family levels between ovarian cancer and normal laying hen blood samples. Data are mean (standard deviation) in the first row and mean fold-change relative to normal (95% CI) in the second row. [^]P value using Mann–Whitney test.

Correlation analysis of levels of miR-200 family members in human blood samples. To account for the observed racial differences, we compared the correlations of the levels of miR-200 family members in all samples and within different subgroups. When analyzed in all 214 samples, the levels of most miR-200 family members were significantly correlated with each other, except between miR-429 and miR-141 (Table S4). We saw similar correlation pattern in the analysis of all 118 cancer cases, with insignificant correlation between miR-429 and miR-141, and higher positive correlation coefficients among the other members (Table S5). However, the correlations were much weaker except between miR-200c and miR-141 in the analysis within all 96 normal controls, with miR-429 showing either insignificant or negative correlations with the other members (Table S6). When the analysis was performed within the American samples (N = 84), we saw significant correlations between miR-200c and miR-200a and miR-200b, as well as between miR-141 and all miR-200 members except miR-429 (Table S7). The same correlation pattern was also observed in the analysis within American cancer samples (N = 49), with miR-429 not showing any significant correlations with other members (Table S8). Within American normal samples (N = 35), correlations were only observed between miR-200b and miR-200c and miR-141, and negatively with miR-429 (Table S9). When the analysis was performed within the Chinese group (N = 130), all miR-200 family members showed significant correlations with each other except between miR-429 and miR-141 (Table S10), similar to what we observed in the analysis of all samples (Table S4). Like the American cancer counterparts, Chinese cancer samples (N = 69) showed the same pattern as the analysis for all Chinese samples, with even higher significant correlation coefficients (Table S11), suggesting that the correlation pattern observed with all samples was primarily contributed by the elevated levels of miR-200 family in cancer samples. In the analysis within Chinese normal samples (N = 61), only the levels between miR-200a and miR-200b, and the levels between miR-141 and miR-200b and miR-200c showed weak significance (Table S12).

Elevated levels of miR-429 in human endometrioid and laying hen ovarian cancers. We then proceeded to compare the miR-200 levels across different histological subtypes of ovarian cancer. We chose to analyze only the Chinese samples, which had comparable number of samples derived from different histological subtypes. By Dunnett multi-comparison, we showed that serous and mucinous ovarian cancer blood samples had significantly higher levels of miR-200b (670.5-fold and 267.1-fold, respectively) than normal blood samples; serous cancer blood samples also had significantly higher levels of miR-200c (198.2-fold) than normal; miR-429 levels in clear cell and endometrioid cancer blood samples were significantly higher (64.9-fold and 88.8-fold, respectively) than normal; whereas miR-141 levels in clear cell, endometrioid, and mucinous cancer blood samples were significantly lower (0.01-fold, 0.01-fold, and 0.03-fold, respectively) than normal blood samples (Table 3). Interestingly, when we examined miR-200 levels in the plasma samples from laying hens, only miR-429 showed significantly elevated levels (17.3-fold) in the cancer plasmas compared to control (Table 4 and Figure S4). Correlation analysis showed that chicken miR-200a and miR-200b levels had the highest correlation (Spearman correlation coefficient = 0.817, $P < 0.001$). MiR-429 also showed respectively high correlations with miR-200a (0.535, $P = 0.005$) and miR-200b (0.661, $P < 0.001$) (Table S13). As endometrioid ovarian cancer is the

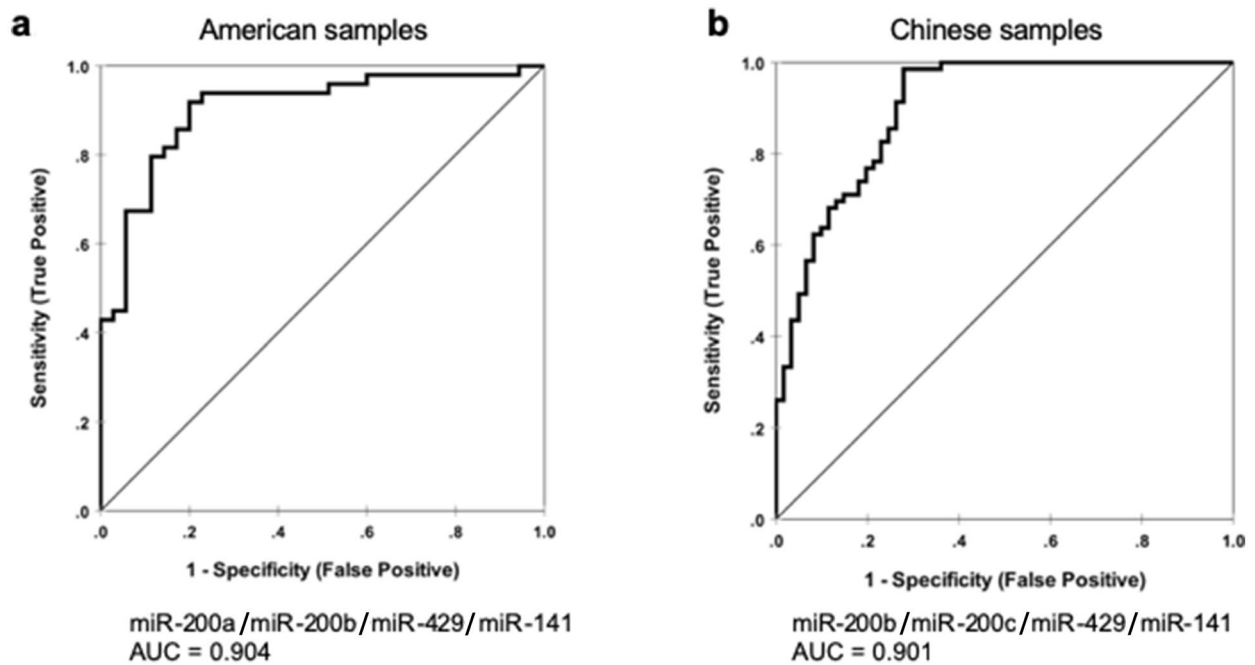


Figure 1. Receiver operating characteristic (ROC) curves for the best multilayer perceptron neural network predictive models to distinguish cancers from non-cancer controls for (a) American patients; and (b) Chinese patients.

major histological type in the chicken³³, the laying hen finding is also consistent with the significant increase of miR-429 levels in human endometrioid blood samples.

Multilayer perceptron neural network models to predict different miR-200 combinations for American and Chinese ovarian cancer patients. Multilayer perceptron (MLP) neural network predictive models³⁴ were used to estimate separately individual and combinations of miR-200 members in differentiating cancer and control blood samples as described in Materials and Methods. The best neural network model consisting of miR-200a/200b/429/141 showed an AUC value of 0.904 for American women in distinguishing between cancer patients and controls, whereas a model consisting of miR-200b/200c/429/141 showed an AUC value of 0.901 for Chinese ovarian cancer prediction (Fig. 1).

Discussion

We examined levels of miR-200 family members in normal and cancer blood specimens collected from American and Chinese institutions. Statistical analyses have revealed that miR-200a, miR-200b, and miR-200c were significantly elevated in cancer specimens relative to controls in the analysis of all specimens, which are similar to the findings by Taylor et al.¹², Kan et al.²⁵, and Zuberi et al.²⁷. Correlation analysis has shown that the levels of these three miR-200 family members were consistently highly correlated with each other than miR-429 and miR-141, with the latter two showing discordant correlation with each other.

Further analysis of the American plasma samples showed that all miR-200 family members except miR-429 were elevated in cancer relative to control plasma samples (Table 2), which is also consistent with the finding of another American miRNA biomarker discovery study using tumor serum-derived exosomes reported by Taylor et al.¹², further confirming that plasma and serum specimens have similar miRNA levels and consistent elevation of these four miR-200 family members in American ovarian cancer patients. Similar findings of elevated miR-200a, miR-200b, miR-200c levels in cancer serum samples were also found in the Australian study²⁵ and the Indian study²⁷. However, our analysis of Chinese specimens revealed that Chinese cancer blood samples had significantly higher levels of miR-200b and miR-429 and lower miR-141 levels than in normal samples (Table 2), which is contradictory to two Chinese studies, which showed either increased levels of plasma miR-200a and miR-141²⁶, or serum miR-200c and miR-141⁷ in cancer than in normal blood samples. The differences might be due to the use of different endogenous normalizer, the form of the blood samples, and the widely variable miR-141 levels observed in our study.

Correlation analysis of the miR-200 family levels show that the levels of miR-200 members were highly correlated with each other in the cancer samples with the exception of miR-429 in American cancer samples (Tables S5, S8, and S11), which also strongly represented the patterns observed in the analysis of combined cancer and normal samples (Tables S4, S7, and S10) than the normal samples (Tables S6, S9, and S12). This observation implies that normal blood samples usually have low and variable miR-200 levels, whereas cancer samples have consistently elevated levels of miR-200 members and dominate the patterns revealed from combined samples.

The significant positive association of clear cell and endometrioid ovarian cancers with miR-429 and negative association with miR-141 in our histological subtype analysis (Table 3) may account for the differences between American and Chinese cancer samples (Table 2) and differences in correlation analysis (Tables S8 and S11). Since our American cancer samples are primarily serous cancer type and have higher levels of all miR-200 family members except miR-429, and our Chinese samples have significant contribution from clear cell and endometrioid cancer samples, hence we find significant positive contribution by miR-429 in Chinese cancer cases (Table 2 and Table S11), but not in the American cases (Table 2 and Table S8). Future studies with more balanced numbers of histological subtypes in both ethnic groups will confirm the interesting finding of racial disparity of miR-200 blood levels, as there are reports about racial differences in microRNA expression in other diseases^{28–30}. More importantly, the human endometrioid cancer-specific miR-429 association is also observed in the cancer plasma samples of laying hen (Table 4), which is prevalently associated with endometrioid ovarian tumors. This finding not only supports the incessant ovulation hypothesis, but also highlights the importance of miR-200-mediated ovarian carcinogenesis in both species^{15,16,18}.

In conclusion, we show that miR-200 family is a highly relevant and informative mechanistic biomarker for ovarian cancer, demonstrating outstanding AUC values in our neural network modeling. Our study also shows racial disparity of miR-200 level in the blood, which might contribute to different miR-200 combinations identified in our neural network models for American and Chinese populations (Fig. 1). More importantly, miR-200 family members provide histological subtype-dependent information in both human and laying hen blood samples and have important implication in ovarian carcinogenesis. Longitudinal laying hen blood samples collected before and after tumor development may be useful for ovarian cancer biomarker discovery.

Materials and methods

Human and laying hen blood specimens. All methods were carried out in accordance with relevant guidelines and regulations for both humans and animals. Informed consent was obtained from all human subjects. All patient-derived biologic specimens were collected and archived under protocols approved by the respective Human Subjects Committee of the institutions, namely, Human Subjects Committee of the Brigham and Women's Hospital, Institutional Review Board of The MD Anderson Office of Human Subjects Protection, The Joint CUHK-NTEC Clinical Research Ethics Committee, Medtimes Medical Group Ethics Review Board, and CloudHealth Medical Group Clinical Research Review Board. For the American samples, 34 banked plasma samples (10 normals and 24 cancers) at Brigham and Women's Hospital (BWH) were selected through a Biobank Portal query tool and requested from Partners HealthCare Biobank (<https://personalizedmedicine.partners.org/Biobank/Default.aspx>). 50 plasma specimens (25 normals and 25 cancers) were from MD Anderson Cancer Center, obtained from the Blood Specimen Research Resource (BSRR) (<https://www.mdanderson.org/research/departments-labs-institutes/programs-centers/center-for-translational-and-public-health-genomics/resources.html>). A search of databases showed that not any single specimen was from Asian patients. For the Chinese samples, 70 plasma samples were collected at the Department of Obstetrics and Gynecology, Prince of Wales Hospital, Chinese University of Hong Kong. 16 whole blood of healthy Chinese female individuals with no known diseases were collected at CloudHealth Genomics Limited. In addition, 60 serum samples of healthy Chinese females were collected at MedTimes Molecular Laboratory Limited. Of the 60 samples, 10 samples had accompanied plasma and whole blood for comparison. Peripheral blood samples were collected in color-coded Vacutainer tubes (BD Biosciences, San Jose, CA) using routine phlebotomy techniques and fractionated into multiple aliquots and stored at -80°C .

Single-comb White Leghorn hens were maintained as previously described^{35,36}, with review and approval of the Institutional Animal Care and Use Committees at the University of Illinois at Urbana-Champaign and Southern Illinois University at Carbondale. The plasma samples tested were from age-matched laying hens fed with standard commercial layer feed.

Handling of blood specimens. All the blood specimens were sent to Brigham and Women's Hospital in dry ice and processed by the first author except for the whole blood and serum samples from Chinese healthy women. The whole blood samples from CloudHealth Genomics were processed by the fifth author in Shanghai because of the strict policy for exporting biological samples, while the serum samples were processed in MedTimes Molecular Laboratory by the first author. Nevertheless, the same RNA extraction and RT-qPCR procedure were followed for all plasma, sera and whole blood specimens.

RNA extraction and reverse transcription. All samples were centrifuged at 2000 rpm for 5 min to remove remaining blood cells. One hundred thirty microliters of human blood sample or 50 μl of chicken plasmas were transferred to a 1.5 ml tube and top up to 200 μl with RNase free water. RNA extraction was performed according to miRCURY™ RNA Isolation Kit-Biofluids Instruction manual v1.7 (#300112 and #300113, Exiqon, Woburn, MA). Briefly, lysis buffer was added to the samples, followed by protein precipitation solution. After centrifugation, supernatant was transferred to a new collection tube. RNA was precipitated with glycerol-isopropanol mix. The microRNAs were captured by microRNA Mini Spin Column and eluted with RNase free water. Samples were stored at -80°C until reverse transcription was performed.

Reverse transcription was performed according to miRCURY LNA™ Universal RT microRNA PCR Instruction manual v6.1 (#203301, 203351, Exiqon, Woburn, MA). Spike-in RNA control (UniSp6) was included in the reverse transcription reaction to reveal the efficiency of reverse transcription and the subsequent PCR reactions. Normalization was done by subtracting the Ct value obtained with the spike-in RNA (UniSp6). Two microliters of RNA template and 8 μl of working master mix solution containing reaction buffer, enzyme for reverse

transcription and synthetic RNA spike-in control were mixed together and incubated at 42 °C for 60 min, followed by 5 min at 95 °C.

Quantitative PCR reactions. We employed Exiqon (Woburn, MA) miRCURY LNA™ Universal RT microRNA PCR primer sets because of the exceptional sensitivity and specificity and low background. cDNA was diluted 40-fold with nuclease-free water before PCR amplification. cDNA from human samples were mixed with PCR master mix containing primers for different miR-200 members, UniSp6, or the endogenous reference miRNAs. cDNA from chicken samples were mixed with PCR master mix containing compatible human miR-200a, miR-103a-3p, or chicken (gga)-miR-200b, miR-429, or UniSp6. All PCR reactions were performed in a StepOne Plus or a 7500 real-time PCR instrument with 60 amplification cycles (95 °C for 10 s, followed by 60 °C for 1 min). Melt curve analysis was done to reveal primer dimer formation or non-specific binding. Ct value of each reaction was obtained with threshold set to 5000. Ct values larger than 60 were counted as 60. All reactions were performed in duplicates and the average Ct values were used to compute normalized miRNA level for each human and chicken sample.

Statistical analysis. Comparative- Δ Ct (negative delta Ct) method was used to calculate normalized miRNA level for each sample³². Normalized delta Ct value was calculated as [Ct(target) - Ct(UniSp6) - Ct(103a)]. Mean levels of miR-200 family members between groups were compared using independent-sample *t* test and regression analysis. Mean fold-changes of miR-200 expression in various histological types of cancer relative to normal samples were calculated using the $2^{-\Delta\Delta Ct}$ method. Levels of miR-200 members in the laying hen groups were compared using Mann-Whitney test due to small sample size. Histological subtype-specific levels of miR-200 members were analyzed using ANOVA with Dunnett post-hoc test comparing each subtype versus normal samples. Pearson correlation coefficients were calculated for correlation analysis in human samples and Spearman correlation coefficients for the laying hen samples. Multilayer perceptron (MLP) neural network predictive models³⁴, separately for each marker and for a combination of markers, were estimated to differentiate between cancer and control blood samples. Sensitivity analyses were conducted to compute the importance of each marker in determining the neural network predictive models. Validation of predictive models was performed by randomly dividing the samples into a training data set and a test data set in a ratio of 7:3. The receiver operating characteristic (ROC) curve for each predictive model was obtained and the corresponding area under curve (AUC) was reported. Statistical analyses were performed using IBM SPSS-25 (IBM, Chicago, IL). For all statistical comparisons, a level of $P < 0.05$ was accepted as statistically significant.

Data availability

All data generated of the current study are available from the corresponding authors on reasonable request.

Received: 30 May 2020; Accepted: 21 October 2020

Published online: 18 November 2020

References

1. Reid, F. The World Ovarian Cancer Coalition Atlas- Global Trends in Incidence, Mortality and Survival. (World Ovarian Cancer Coalition, 2018).
2. Vogell, A. & Evans, M. L. Cancer Screening in Women. *Obstet. Gynecol. Clin. N. Am.* **46**, 485–499. <https://doi.org/10.1016/j.ogc.2019.04.007> (2019).
3. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA A Cancer J. Clin.* **69**, 7–34. <https://doi.org/10.3322/caac.21551> (2019).
4. Lu, K. H. Screening for ovarian cancer in asymptomatic women. *JAMA* **319**, 557–558. <https://doi.org/10.1001/jama.2017.21894> (2018).
5. Dochez, V. *et al.* Biomarkers and algorithms for diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review. *J. Ovarian Res.* **12**, 28. <https://doi.org/10.1186/s13048-019-0503-7> (2019).
6. Duffy, M. J. *et al.* CA125 in ovarian cancer: European Group on tumor markers guidelines for clinical use. *Int. J. Gynecol. Cancer* **15**, 679–691. <https://doi.org/10.1111/j.1525-1438.2005.00130.x> (2005).
7. Gao, Y. C. & Wu, J. MicroRNA-200c and microRNA-141 as potential diagnostic and prognostic biomarkers for ovarian cancer. *Tumour Biol.* **36**, 4843–4850. <https://doi.org/10.1007/s13277-015-3138-3> (2015).
8. Zhang, Z. *et al.* Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* **64**, 5882–5890. <https://doi.org/10.1158/0008-5472.can-04-0746> (2004).
9. Antoranz, A., Sakellaropoulos, T., Saez-Rodriguez, J. & Alexopoulos, L. G. Mechanism-based biomarker discovery. *Drug Discov. Today* **22**, 1209–1215. <https://doi.org/10.1016/j.drudis.2017.04.013> (2017).
10. Robinson, W. H., Lindstrom, T. M., Cheung, R. K. & Sokolove, J. Mechanistic biomarkers for clinical decision making in rheumatic diseases. *Nat. Rev. Rheumatol.* **9**, 267–276. <https://doi.org/10.1038/nrrheum.2013.14> (2013).
11. Hausler, S. F. *et al.* Whole blood-derived miRNA profiles as potential new tools for ovarian cancer screening. *Br. J. Cancer* **103**, 693–700. <https://doi.org/10.1038/sj.bjc.6605833> (2010).
12. Taylor, D. D. & Gercel-Taylor, C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol. Oncol.* **110**, 13–21. <https://doi.org/10.1016/j.ygyno.2008.04.033> (2008).
13. Park, S. M., Gaur, A. B., Lengyel, E. & Peter, M. E. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev.* **22**, 894–907. <https://doi.org/10.1101/gad.1640608> (2008).
14. Choi, P. W. *et al.* Loss of E-cadherin disrupts ovarian epithelial inclusion cyst formation and collective cell movement in ovarian cancer cells. *Oncotarget* **7**, 4110–4121. [https://doi.org/10.18632/oncotarget.65886588\[pilii\]](https://doi.org/10.18632/oncotarget.65886588[pilii]) (2016).
15. Yang, J. *et al.* Characterization of MicroRNA-200 pathway in ovarian cancer and serous intraepithelial carcinoma of fallopian tube. *BMC Cancer* **17**, 422. <https://doi.org/10.1186/s12885-017-3417-z> (2017).
16. Hales, K. H., Speckman, S. C., Kurrey, N. K. & Hales, D. B. Uncovering molecular events associated with the chemosuppressive effects of flaxseed: a microarray analysis of the laying hen model of ovarian cancer. *BMC Genom.* **15**, 709. <https://doi.org/10.1186/1471-2164-15-709> (2014).

17. Gregory, P. A. *et al.* The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat. Cell Biol.* **10**, 593–601 (2008).
18. Choi, P. W. *et al.* MicroRNA-200 family governs ovarian inclusion cyst formation and mode of ovarian cancer spread. *Oncogene* <https://doi.org/10.1038/s41388-020-1264-x> (2020).
19. Fathalla, M. F. Incessant ovulation—a factor in ovarian neoplasia?. *Lancet* **2**, 163. [https://doi.org/10.1016/S0140-6736\(71\)92335-X](https://doi.org/10.1016/S0140-6736(71)92335-X) (1971).
20. Choi, P. W. & Ng, S. W. The functions of microRNA-200 family in ovarian cancer: beyond epithelial-mesenchymal transition. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms18061207> (2017).
21. Moga, M. A. *et al.* Circulating miRNAs as biomarkers for endometriosis and endometriosis-related ovarian cancer—an overview. *J. Clin. Med.* <https://doi.org/10.3390/jcm8050735> (2019).
22. Jaurreche, S., Perez, G., Antongiovanni, N., Perretta, F. & Venera, G. Variables associated with a urinary microRNAs excretion profile indicative of renal fibrosis in fabry disease patients. *Int. J. Chron. Dis.* **2019**, 4027606. <https://doi.org/10.1155/2019/4027606> (2019).
23. Hydring, P. *et al.* Exosomal RNA-profiling of pleural effusions identifies adenocarcinoma patients through elevated miR-200 and LCN2 expression. *Lung Cancer (Amsterdam, Netherlands)* **124**, 45–52. <https://doi.org/10.1016/j.lungcan.2018.07.018> (2018).
24. Cheng, H. *et al.* Circulating plasma MiR-141 is a novel biomarker for metastatic colon cancer and predicts poor prognosis. *PLoS ONE* **6**, e17745. <https://doi.org/10.1371/journal.pone.0017745> (2011).
25. Kan, C. W. *et al.* Elevated levels of circulating microRNA-200 family members correlate with serous epithelial ovarian cancer. *BMC Cancer* **12**, 627. <https://doi.org/10.1186/1471-2407-12-627> (2012).
26. Zheng, H. *et al.* Plasma miRNAs as diagnostic and prognostic biomarkers for ovarian cancer. *PLoS ONE* **8**, e77853. <https://doi.org/10.1371/journal.pone.0077853> (2013).
27. Zuberi, M. *et al.* Expression of serum miR-200a, miR-200b, and miR-200c as candidate biomarkers in epithelial ovarian cancer and their association with clinicopathological features. *Clin. Transl. Oncol.* **17**, 779–787. <https://doi.org/10.1007/s12094-015-1303-1> (2015).
28. Dluzen, D. F. *et al.* Racial differences in microRNA and gene expression in hypertensive women. *Sci. Rep.* **6**, 35815. <https://doi.org/10.1038/srep35815> (2016).
29. Bhagirath, D. *et al.* Role of a novel race-related tumor suppressor microRNA located in frequently deleted chromosomal locus 8p21 in prostate cancer progression. *Carcinogenesis* **40**, 633–642. <https://doi.org/10.1093/carcin/bgz058> (2019).
30. Telonis, A. G. & Rigoutsos, I. Race disparities in the contribution of miRNA isoforms and tRNA-derived fragments to triple-negative breast cancer. *Cancer Res.* **78**, 1140–1154. <https://doi.org/10.1158/0008-5472.can-17-1947> (2018).
31. Glinge, C. *et al.* Stability of circulating blood-based microRNAs—pre-analytic methodological considerations. *PLoS ONE* **12**, e0167969. <https://doi.org/10.1371/journal.pone.0167969> (2017).
32. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**, 1101–1108. <https://doi.org/10.1038/nprot.2008.73> (2008).
33. Ansenberger, K. *et al.* E-cadherin expression in ovarian cancer in the laying hen, *Gallus domesticus*, compared to human ovarian cancer. *Gynecol. Oncol.* **113**, 362–369. <https://doi.org/10.1016/j.ygyno.2009.02.011> (2009).
34. Rumelhart, D. E. H., G.E.; Williams, R.J. in *Parallel distributed processing: Explorations in the microstructure of cognition: Foundation Vol. 1* (ed D.E.; McClelland Rumelhart, J.L.; and the PDP research group) 318–362 (MIT Press, Cambridge, 1986).
35. Ansenberger, K. *et al.* Decreased severity of ovarian cancer and increased survival in hens fed a flaxseed-enriched diet for 1 year. *Gynecol. Oncol.* **117**, 341–347. <https://doi.org/10.1016/j.ygyno.2010.01.021> (2010).
36. Eilati, E., Pan, L., Bahr, J. M. & Hales, D. B. Age dependent increase in prostaglandin pathway coincides with onset of ovarian cancer in laying hens. *Prostaglandins Leukot. Essent Fatty Acids* **87**, 177–184. <https://doi.org/10.1016/j.plefa.2012.09.003> (2012).

Acknowledgements

We acknowledge the support of the Robert and Deborah First Fund, the Spering Family Fund Foundation, Ruth N. White Gynecologic Oncology Research Fund, Women's Cancer Program and Gillette Center for Women's Cancer from Dana-Farber Cancer Institute, Ovarian Cancer Research Foundation, Adler Foundation, Inc., and Friends of Dana Farber Cancer Institute to The Laboratory of Gynecologic Oncology at Brigham and Women's Hospital.

Author contributions

Conception and design: S-W.N. and P-W.C. Acquisition of data (provided samples, acquired and managed patients, provided facilities, etc.): C.K.C., F.X., W.P.K., J.K., K.H.H., D.B.H., K.K.W. Experiments: P-W.C., S.K.L., and W.Q. Statistical analysis: A.B. and S-K.N. Writing, review, and/or revision of manuscript: P-W.C., S-K.N., E.R.N., C.K.C., R.S.B., and S-W.N.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-77068-0>.

Correspondence and requests for materials should be addressed to C.K.C. or S.-W.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020