Gene, Volume 590, Issue 2, 30 September 2016, pp210 – 219

Accepted Manuscript

Study of microRNAs-21/221 as potential breast cancer biomarkers in Egyptian women

Tarek Mohamed Kamal Motawi, Nermin Abdel Hamid Sadik, Olfat Gamil Shaker, Maha Rafik El Masry, Fady Mohareb

PII: S0378-1119(16)00104-9
DOI: doi: 10.1016/j.gene.2016.01.042
Reference: GENE 41145

To appear in: *Gene*

Received date: 25 July 2015
Revised date: 11 January 2016
Accepted date: 20 January 2016

Study of microRNAs-21/221 as potential breast cancer biomarkers in Egyptian women

Tarek Mohamed Kamal Motawi[1], Nermin Abdel Hamid Sadik[1]*, Olfat Gamil Shaker[2], Maha Rafik El Masry[3], Fady Mohareb[4]

[1]Biochemistry Department, Faculty of Pharmacy, Cairo University, Kasr El-Einy, Cairo, Egypt
[2]Medical Biochemistry and Molecular Biology Department, Faculty of Medicine, Cairo University
[3]Biochemistry Department, Faculty of Dentistry, October University for Modern Sciences & Arts (MSA), Giza, Egypt
[4]The Bioinformatics Group, School of Energy, Environment and AgriFood, Cranfield University, Bedford, MK43 0AL,UK
* Corresponding author. Address: Faculty of Pharmacy, Cairo University, Kasr El-Einy Street, Cairo 11562, Egypt. Tel.: +002 0103076776; fax: +20 2 3635140. E-mail address: nerminsadik@yahoo.com.

Abstract

microRNAs (miRNAs) play an important role in cancer prognosis. They are small molecules, approximately 17-25 nucleotides in length, and their high stability in human serum supports their use as novel diagnostic biomarkers of cancer and other pathological conditions. In this study, we analyzed the expression patterns of miR-21 and miR-221 in the serum from a total of 100 Egyptian female subjects with breast cancer, fibroadenoma, and healthy control subjects. Using microarray-based expression profiling followed by real-time polymerase chain reaction validation, we compared the levels of the two circulating miRNAs in the serum of patients with breast cancer (n= 50), fibroadenoma (n= 25), and healthy controls (n= 25). The miRNA SNORD68 was chosen as the housekeeping endogenous control. We found that the serum levels of miR-21 and miR-221 were significantly overexpressed in breast cancer patients compared to normal controls and fibroadenoma patients. Receiver Operating Characteristic (ROC) curve analysis revealed that miR-21 has greater potential in discriminating between breast cancer patients and the control group, while miR-221 has greater potential in discriminating between breast cancer and fibroadenoma patients. Classification models using k-Nearest Neighbor (kNN), Naïve Bayes (NB), and Random Forests (RF) were developed using expression levels of both miR-21 and miR-221. Best classification performance was achieved by NB Classification models, reaching 91% of correct classification. Furthermore, relative miR-221 expression was associated with histological tumor grades. Therefore, it may be concluded that both miR-21 and miR-221 can be used to differentiate between breast cancer patients and healthy controls, but that the diagnostic accuracy of serum miR-21 is superior to miR-221 for breast cancer prediction. miR-221 has more diagnostic power in discriminating between breast cancer and fibroadenoma patients. The overexpression of miR-221 has been associated with the breast cancer grade. We also demonstrated that the combined expression of miR-21 and miR-221can be successfully applied as breast cancer biomarkers.

Keywords
MiRNA; miR-21; miR-221; Breast cancer; Fibroadenoma.

Abbreviations

ANOVA: Analysis of variance. AUC: Area under the curve.BLBC: Basal-like breast cancer.CDK: Cyclin-dependent kinases. CI: Confidence interval.Ct: Cycle threshold. EMT: Epithelial–mesenchymal transition. ER: Estrogen receptor.HCC: Hepatocellular carcinoma. kNN: k-Nearest Neighbor. maspin: Mammary serine protease inhibitor. miR-21: microRNA-21. miR-221: microRNA-221. miRNAs: microRNAs. mRNA: messenger RNA. NB: Naïve Bayes. PBS: Phosphate buffer saline. PCA: Principle component analysis. PDCD4: Programmed cell death 4**.** RF:Random Forest.ROC: Receiver Operating Characteristic.RT-PCR: Real-time Polymerase chain reaction. RT: Reverse transcription. SD: Standard deviation.TNBC: Triple negative breast cancer.TPM1: Tropomyosin 1.

Introduction

Worldwide, breast cancer is the most diagnosed cancer affecting women (Hortobagyi et al., 2005; Gill et al., 2007). Approximately 1.67 million new cases of breast cancer were diagnosed in 2012, and by 2025, this figure is predicted to escalate to 19.3 million. Although the highest reported prevalence of breast cancer is in developed countries, an increasing incidence and lower survival rate in developing countries has been found. This trend has been attributed to the adoption of the Western lifestyle (Porter, 2008), lack of breast cancer awareness and poor access to screening and health care services (Beaglehole and Yach, 2003; Parkin and Fernández, 2006; Badar et al., 2007; Rizwan and Saadullah, 2009). Breast cancer is the most frequent cause of cancer death in women from less developed regions (324,000 deaths, 14.3% of the total), and it is a leading cause of cancer death in more developed regions (198,000 deaths, 15.4%), second only to lung cancer (Ferlay et al., 2012; Bray et al., 2012).

According to the Egyptian National Cancer Institute (NCI), breast cancer is the most common type of cancer among Egyptian women, representing 18.9% of total cancer cases (Elatar, 2002), with an age-adjusted rate of 49.6 per 100,000 people. However, this represents hospital-based data from tertiary referral centers and does not represent all breast cancer cases in Egypt. According to the population-based cancer registry of Ghrabiah, Egypt, the median age at diagnosis is one decade younger than in Europe and North America, while most patients are premenopausal (Ibrahim et al., 2002; Omar et al., 2003). Data from GLOBOCAN 2012 also reports that breast cancer is the most prevalent cancer in Egyptian women (Ferlay et al., 2012). The estimated 5-year prevalence of all cancer types occurring in females is 49.2% (Bray et al., 2012).

Fibroadenoma is the most common benign tumor occurring in female breast tissue (Dixon, 1991; Fechner 1988). It is normally diagnosed in young women, but may also occur in older women (Hunter et al., 1996). It may result from abnormal growth and hyperplasia of the breast lobular tissue.

microRNAs (miRNAs) are highly conserved noncoding RNA molecules that are approximately 17–25 nucleotides in length. They control gene expression at the posttranscriptional level by interacting with a specific target messenger RNA (mRNA) (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee et al., 2001). They also regulate a variety of cellular processes, such as proliferation, differentiation, metabolism, aging, and cell death. As such, the importance of miRNAs is increasingly recognized in almost all fields of biological and biomedical fields (Li et al., 2010). In humans, it has been estimated that there are more than 1000 miRNAs in the genome, which regulate approximately 30% of all protein-coding genes (Lewis et al., 2005). The importance of miRNAs in oncogenesis has also been recognized. Dysregulation of miRNA expression plays an important role in cancer development through various mechanisms, such as deletions, amplifications, epigenetic silencing, or mutations in miRNA loci (Kosaka et al., 2010). To date, an association between differentially expressed miRNAs and many clinicopathological features has been shown, including mRNA expression-based classification (Blenkiron et al., 2007), tumor grade, and breast cancer staging (Iorio et al., 2005).

miR-21 is one of the most important miRNAs that is deregulated and over-expressed in many malignant tumors, including breast cancers (Chan et al., 2005). Some studies have reported that miR-221 is also deregulated in breast cancer (Shah and Calin, 2011). miRNA therapy could also be a powerful tool for the treatment of poorly differentiated cancer (Lu et al., 2005).

This study aimed to evaluate the expression level and diagnostic potential of serum miR-21 and miR-221 from Egyptian female patients with breast cancer, fibroadenoma, and healthy control subjects, regardless of the age and also to identify the relationship between the clinicopathological features of breast cancer and the expression of these miRNAs.

Subjects and methods

Patients

A total of 50 female breast cancer patients (mean age ± SD: 53.5 ± 7.5) were assigned to the study. Patients were selected from the Kasr El-Einy Hospital, Faculty of Medicine in Cairo, Egypt. The serum samples were obtained from breast cancer patients who were recently diagnosed by mammogram and from untreated patients. All patients were subjected to a complete clinical examination, and a full clinical history was taken. Patients who had received chemotherapy/radiotherapy or who had an acute infection were excluded from the study, as well as patients who had cancer at any other site at the time of the selection. Chest radiology, liver ultrasound scanning, and bone scanning were used to exclude those with metastatic cancer. The clinicopathological characteristics of breast cancer patients, including the histological grade and hormone receptor status are shown in Table 1.Demographic and clinical features of study groups are shown in Table 2.

In addition, blood samples were collected from 25 female fibroadenoma patients (mean age ± SD: 32.1 ± 14.4) who were diagnosed by mammogram and breast ultrasound. In these patients, the fibroadenoma masses were solid, smooth, painless, and mobile, while aspiration cytology confirmed that the masses were benign in each patient. Additionally, a set of 25 blood samples from healthy female subjects (mean age ± SD: 28.6 ± 5.9) was collected from the outpatient clinic at El-Kasr El-Einy Hospital. None of these individuals had been previously diagnosed with malignancies, hypertension, diabetes, or any other diseases. All procedures involving blood samples collection were performed by trained technicians at the outpatient clinic at El-Kasr El-Einy, Faculty of Medicine, Cairo University, Cairo, Egypt. All participants had Egyptian ethnic origin. The study was performed with the approval of the Faculty of Pharmacy, Cairo University local ethics committee and carried out in compliance with the Helsinki Declaration (2008). Informed consent was obtained from all of the subjects enrolled in this study.

Sample collection and handling

Peripheral blood (~10 ml) was collected from every patient by trained technicians. Cellular components were removed by centrifugation in two consecutive steps (1,500× g for 10 min at 4 °C and 2,000× g for 3 min at 4 °C, respectively). Sera were stored at – 80 °C until use.

Methods

Serum miRNA assays

RNA extraction

Total RNA, including preserved miRNAs, were extracted from 200 µl of frozen serum in 200 µl of Phosphate buffer saline (PBS) using a TRIzol extraction kit (Qiagen, Valencia, CA). QIAGEN Protease (20 µl) was added; then, 4 µl of an RNase A stock solution (100 mg/ml) and 200 µl of Buffer AL were added to the sample. The mixture was pulse-vortexed for 15 s, incubated at 56°C for 10 min in a water bath, and then centrifuged at room temperature at 15-25 °C for 1 min. Ethanol (200 µl, 96%) was added to the sample, mixed again by pulse-vortexing for 15 s, and briefly centrifuged. The mixture was placed in a QIAamp Mini spin column in a 2-ml collection tube and centrifuged at 6000x g (8000 rpm) for 1 min at room temperature. The QIAamp Mini spin column was then placed in a clean 2-ml collection tube, and the tube containing the filtrate was discarded. Buffer AW1 (500 µl) was added to the QIAamp Mini spin column and centrifuged at 6000x g (8000 rpm) for 1 min, and then, the column was placed in another clean 2-ml collection tube and the filtrate was again discarded. Buffer AW2 (500 µl) was added and centrifuged at 20,000x g (14,000 rpm) for 3 min. Then, the column was placed in a new 2-ml collection tube, the filtrate was discarded and the column was centrifuged at full speed for 1 min. Finally, the column was transferred to a new 1.5-ml collection tube and 200 µl of Buffer AE was added; the tube was incubated at room temperature for 1 min and then centrifuged at 6000x g (8000 rpm) for 1 min to elute the RNA. The RNA purity was assessed using NanoDrop ND-1000 (Nanodrop, USA).

Reverse transcription (RT)

The RT kit that was used was made specifically for accurate analysis of the miRNAs from the serum samples. RT was carried out on 5 ng of total RNA in a final volume of 20 µl. The RT reactions (incubated for 10 min at 25 °C, 60 min at 37 °C, 5 min at 95 °C, and then maintained at -15 °C) were performed using the RT kit (Qiagen, Germany) according to the manufacturer's directions.

Microarray and quantitative PCR (qPCR)

The expression of mature miRNAs (miR-21; miR-221) was evaluated by qRT-PCR analysis, according to the manufacturer's directions. The housekeeping miRNA SNORD68 was used as an endogenous control. For RT-PCR, 5 µl of diluted RT products (cDNA template) was mixed with 12.5 µl of SYBR Green Master Mix (Qiagen, Germany), and nuclease free water was added to a final volume of 25 µl and dispensed into a 96-well miScript miRNA PCR array plate, which was enriched with forward and reverse miRNA specific primers supplied by (Qiagen, Germany). The plate was sealed with MicroAmp® Optical 8-Cap strips. Real-time PCR was performed using an Applied Biosystems 7500 Real-time PCR System (Applied Biosystems; Foster City, CA, USA) under the following conditions: 95°C for 15 min, followed by 40 cycles at 95°C for 5 s and 60°C for 34 s. The data obtained from the miRNA expression levels were calculated and evaluated by the cycle threshold (Ct) method, which is the number of cycles required for the fluorescent signal to cross the threshold in RT-PCR. The level of miRNA expression was reported as ΔCt value. The ΔCt was calculated by subtracting the Ct value of miRNA SNORD68 from the Ct values of the target miRNAs [mean value Ct (miR-21, miR-221) - mean value Ct (housekeeping gene)]. Because there is an inverse relationship between ΔCt and the miRNA expression level, lower ΔCt values are associated with increased miRNA. The resulting normalized ΔCt values were subtracted from an arbitrary reference value of 50 to transform the data to a scale of inverted normalized Ct, where a high number indicates a high expression level (Barshck et al., 2010). The relative expression level of the miRNA of interest corresponded to the $2^{-\Delta Ct}$ value. ΔΔCt was then determined by subtracting the average ΔCt of the control from the ΔCt of cases. The fold change in the miRNA expression level was calculated (fold change = $2^{-\Delta\Delta Ct}$) to determine the relative quantitative levels of individual miRNA (Livak and Schmittgen, 2001).

Statistical analyses

The statistical analyses were performed using IBM SPSS advanced software, version 20 (SPSS Inc., Chicago, IL). The numerical data were expressed as the mean, standard deviation, range, or frequency. The qualitative data were expressed as frequency and percentage. For nonparametric comparison analysis between two groups, the Mann Whitney-U test was used, and the Kruskal-Wallis test was used for more than two independent variables. The Chi-square test was used to examine the relationship between qualitative variables; Fisher's exact test was used instead when the expected frequency was less than 5. For quantitative data, the comparison between the 3 groups was performed by analysis of variance (ANOVA); then, the post-Hoc Scheffe test was used for pair-wise comparison. Power analysis was done to determine the statistical power and the appropriate sample size for our primary comparison of miRNA among the study groups. The Receiver Operating Characteristic (ROC) curve was used to determine the cut-off values of miRNAs and to analyze the diagnostic utility of different markers. A p-value of less than 0.05 was considered statistically significant. All P-values are two- sided.

Multivariate statistical analysis

The aim of multivariate analysis is to cluster samples according to the captured variance, which in this study should be according to the expression levels of miR-21 and miR-221. Data pre-treatment methods, namely auto-scaling and range-scaling, were initially considered prior to multivariate analysis. However, they show no further improvement in the PCA clustering; and therefore, raw expression was used to perform Principal Component Analysis (PCA). PCA was applied using the open-source R statistical environment and the "prcomp" function. PCA plots were generated using the "ggbiplot" package. Since multivariate analysis does not allow missing values, only expression levels were included in the analysis. Patients' meta-data, such as age, family history, diabetes, hypertension, and menopausal status, as well as other measurements (i.e. hormonal control, parity, number of pregnancies) were not available for the three groups and hence were not included in the multivariate analysis. Similarly, hierarchical clustering was performed based on the expression levels. Prior to clustering, miRNA profiles were standardized to have mean zero and standard deviation one. Clustering was performed using the R "gplots" and "d3Heatmap" libraries with average linkage and Pearson correlation.

Power analysis

The null hypothesis with > 99% power is rejected if true mean difference among the study groups was similar to our calculated differences. Power analysis is accepted if it is 80%. The omnibus one way analysis of variance test was used in the analysis, with type I error probability of 0.05. Calculations were performed using G*Power software version 3.1.2 for MS Windows, Franz Faul, Kiel University, Germany.

Classification modeling

The *k*-Nearest Neighbors (*k*NN) is a machine learning method which applies samples distance to perform classification. Briefly, the *k*-closest points to the sample are considered, before a majority vote is applied to classify it or predict its value (Harrington, 2012). Naïve Bayes (NB) is a probabilistic method based on conditional probability. Conditional probability is the probability of an event knowing that other event has taken place (StatSoft, 2013). The algorithm is based on the posterior probability of the sample belonging to each of the classes by combining (multiplying) the prior probability of belonging to one class by the likelihood of the new sample belonging to such class. Random Forest (RF) is an ensemble method based on bootstrap aggregation. This method constructs multiple versions of the training data by sampling with replacement (bootstrapping), creates a model and makes predictions for all of them and combines the predictions. Boosting is quite a similar approach to bagging but uses weak learners –a simple algorithm that performs slightly better than classifying by chance– and samples are re-weighted through several iterations in order to use the weights to calculate the final predictions (Kantardzic, 2005). A comprehensive outline and examples of the machine learning methods are available at (Harrington, 2012).

Random forests algorithm uses bootstrap samples, creates tree models for a certain number of random features for each one of the bootstrap samples and predictions of the tree models are combined to obtain the final prediction (Fig.1).

Classification models using kNN, NB, and RF were developed based on the expression levels of miR-21 and miR-221 of the total 100 subjects considered in this study. It should be noted that the patients' metadata and other prior information were not incorporated within the models input to prevent the model output from being dependent on any prior knowledge about the samples, apart from the expression levels of the miR-21 and miR-221. Steps involved in developing the classification models are shown in Fig. 2. Firstly, input samples were randomly divided into a training and a testing subset consisting of 75 and 25 samples, respectively. Testing the models accuracy using a testing subset completely unknown to the developed models is far more indicative than the conventional leave-one-out cross-validation method. In order to ensure the balance among the three classes (Cancer, Control, and Fibroadenoma), we included a representative number of samples of each class in each subset. The training subset was then used to develop the classification models using the kNN, NB, and RF. For each classification approach, a grid search was performed to identify the optimum parameter by examining the confusion matrix of the training dataset. The optimized kNN, NB, and RF models were then used in order to predict the classes for the testing (unknown) subset created earlier. The overall model performance for each classifier was assessed as a percentage value based on the total number of correct classification divided by the total number of samples within the testing subset. The kNN, NB, and RF classifiers were developed using the "kNN", "e1071", and "RandomForest" R packages respectively.

Results

Demographic and clinical features of study groups

Age in breast cancer patients was significantly different from the controls (P <0.001) and fibroadenoma group (P <0.001). No significant difference was revealed between the fibroadenoma group and the controls. A significant difference was observed between breast cancer and fibroadenoma patients in pre and post-menopausal patients (P <0.0001), family history (P <0.0001), diabetes (P <0.001) and hypertension (P < 0.007) (Table 2).

Serum expression levels of miR-21 and miR-221

The expression levels of miR-21 and miR-221 were evaluated by qRT-PCR. The serum levels of miR-21 andmiR-221 were significantly higher in cancer patients than in healthy control subjects, corresponding to an average fold change of 2.2 and 2.09, respectively. They were also significantly higher in cancer patients than fibroadenoma patients, with an average fold change of 1.6 and 1.9, respectively at P< 0.001 (Table 3).On the other hand, no significant increase was observed in the serum levels of miR-21 and miR-221 in the fibroadenoma patients compared to the control group. Power analysis of serum of miR-21 and miR221 revealed power of 95%. As shown in Table 4, there was a significant increase in the serum expression level of miR-221 of tumor grade III (GIII) patients compared to GI and II patients (P< 0.05). The statistical analysis between the miRNA concentration and clinical and histopathological data did not reveal any statistical significance.

Evaluation of the diagnostic accuracy of miR-21 and miR-221

The diagnostic accuracy of miR-21 and miR-221 were evaluated using ROC curve analysis. ROC curve analysis showed that the two miRNAs can significantly differentiate between breast cancer and healthy controls, showing an area under the curve (AUC) of 0.98 for miR-21 (95% CI 1.0-0.96, P< 0.05) and AUC 0.97 (95% CI 1.003-0.936, P< 0.05) for miR-221. The optimal sensitivity and specificity were (96% and 92%) and (94% and 88%), respectively. In addition, miR-21 and miR-221 can discriminate between breast cancer patients and fibroadenoma patients, showing an AUC 0.85 (95% CI 0.937-0.772, P<0.05) and AUC 0.93 (95% CI 0.986-0.866, P<0.05), respectively (Fig. 3 a-d). The optimal sensitivity and specificity were (82% and 76%) and (90% and 84%), respectively.

When the diagnostic significance of serum miRNAs was compared in breast cancer patients, the results of the ROC curve suggested that the diagnostic accuracy of serum miR-21 was superior tomiR-221, with AUCs of 0.98 and 0.97, respectively.

Multivariate analysis and classification models

RT-PCR data from the two miRNAs was used as input to generate PCA to visually assess the intrinsic variation in the two miRNAs profiles among the three groups. The first principal component (PC1) captured 72.1% variance while the second one (PC2) captured 27.9% of variance. Control and fibroadenoma samples were more compactly clustered showing a small variance among subjects compared to the cancer subjects (Fig. 4). The same observation was seen in the heatmap hierarchical clustering in Fig. 5, where miRNAs expression clearly separated control group from cancer group. On the other hand, we did not observe a clear separation of fibroadenoma group from the two groups. Initial separation shown using multivariate analysis indicated the suitability of the input dataset for classification modeling.

A series of three different classifiers was developed based on RT-PCR measurements, in order to predict the patient categories (cancer, control, or fibroadenoma). The optimum parameters for each model were identified using grid search, these parameters were used in order to build the final set of models using the training subset. For kNN model, it was found that the optimum k value was achieved at 5. In the case of NB, the optimum classification was achieved using ~100 trees and by calculating the proximity measure among the rows. The overall performance for each model was assessed against the testing subset. Both kNN and NB models achieved 87.5% overall classification accuracy when tested against the randomly selected testing subset, while RF achieved the best performance accuracy with 97.8%. Furthermore, RF achieved a 100% correct classification for the cancer and fibroadenoma categories, while only one control sample misclassified as fibroadenoma

(See confusion matrices for the training and testing subsets at Table 5). We also examined the robustness of the three optimized models by re-running the classification process through a series of 100 cycles. At each cycle, the training and testing subset were reshuffled and the overall performance for each model was measured. The overall average performance also indicated that RF achieved the best performance, with an average of 91%, followed by NB at 88%, and kNN at 82.3% (Fig. 6). The models stabilized after ~30 iterations, indicating a very good stability of the classifiers in terms of prediction accuracy.

Discussion

The expression patterns and levels of specific miRNAs could reflect altered physiological and pathological conditions. Due to their high stability in human serum, they represent attractive novel diagnostic biomarkers for certain health conditions, such as cancer. miRNAs are associated with the regulation of oncogenes and tumor suppressor genes. Previous research has shown that their disruption is related to many types of cancer (Lu et al., 2005; Shenouda and Alahari, 2009; Garofalo et al., 2010; Kouhkam et al., 2011), including breast cancer (Iorio et al., 2005).

In the present study, we set out to analyze the expression patterns of miR-21 and miR-221 as a single biomarker. Our results demonstrated that serum levels of miR-21 and miR-221 are significantly increased in breast cancer patients compared to those of fibroadenoma patients and healthy control subjects. We found that the expression levels of these two miRNAs can significantly discriminate between breast cancer patients and healthy subjects, with high specificity and sensitivity using ROC curve analysis and a fold change of 2.2 and 2.09, respectively. The results showed that miR-21 has considerable diagnostic power in discriminating between breast cancer patients and control subjects, yielding an AUC of 0.98 with a sensitivity of 96% and a specificity of 92%. Moreover, miR-21 andmiR-221 can discriminate between breast cancer and fibroadenoma patients. miR- 221 has more diagnostic power than miR-21, yielding an AUC of 0.93 with a sensitivity of 90% and a specificity of 84%. With regard to the clinicopathological features, miR-221 was significantly overexpressed in grade III compared to Grade I and II, with a fold change of 1.3.

We thoroughly tested the robustness and specificity of using the expression levels for miR-21 and miR-221 by developing and comparing the performance of three classification models using kNN, NB, and RF. All three models achieved high prediction accuracies, reaching a top performance at 97.8% for RF. Furthermore, the developed RF model achieved a 100% correct classification for both cancer and fibroadenoma categories, indicating a promising potential of successfully using miR-21 and miR-221 biomarkers for breast cancer.

In general, high serum miRNA levels in cancer patients are due to excessive secretion by primary cancer cells (Mitchel et al., 2008). Previous studies have shown that miRNAs can be selectively secreted into the bloodstream via small membrane vesicles, such as exosomes (Gallo et al. 2012), that are released into the extracellular environment (Mathivanan et al., 2010). A study has shown that cellular gene products, including miRNAs, are packaged inside exosomes and are delivered to the target cells, where they have a biological effect (Ohshima et al., 2010).

miR-21 is one of the most important miRNAs associated with cell migration and the invasiveness of breast cancer cells, thus contributing to tumor progression and metastasis (Han et al., 2012a; Han et al., 2012b). Chan et al. (2005) reported the aberrant expression of miR-21 in glioblastoma. Previous research has shown that miR-21, along with other miRNAs, is over-expressed in human breast cancer (Iorio et al., 2005).

Our results from this study coincide with previous research (Iorio et al., 2005; Si et al., 2007), showing a significant overexpression of miR-21 in breast cancer patients, which suggests that it acts as an oncogene (Yan et al., 2008). Previous research has shown that the mammary serine protease inhibitor (maspin) and programmed cell death 4 (PDCD4), which are involved in invasion and metastasis, have been identified as targets for miR-21 (Zhu et al., 2008). Maspin plays an important role in breast cancer, as it suppresses invasion and metastasis with its ability to induce apoptosis, thus suppressing angiogenesis (Brew et al., 2000; Bailey et al., 2006; Song et al., 2012). PDCD4 induces the expression of p21, which acts as an inhibitor of cyclin-dependent kinases (CDK) (Frankel et al., 2008), therefore playing a role in apoptosis. Altered function of miR-21 inhibits PDCD4, encouraging uncontrolled cellular growth and cancer progression. In addition, another study confirmed that miR-21 targets the tumor suppressor tropomyosin 1 (TPM1). This is a member of the tropomyosin family of proteins, which are associated with actin and serve to stabilize microfilaments and act as tumor suppressor genes (Perry, 2001, Zhu et al., 2007). Thus, the downregulation of TPM1 by miR-21 suppression may explain the inhibition of tumor invasion. Suppression ofmiR-21 by anti-miR-21 is linked with reduced cell proliferation in vitro and tumor growth in vivo (mouse model) (Corcoran et al., 2011).

miR-21 overexpression is associated with an advanced clinical stage and lymph node metastasis in human breast cancer and is also associated with low sensitivity and a poor response to chemotherapy (Yan et al., 2008; Krichevsky et al., 2009). Our results showed that overexpression of serum miR-21 does not appear to correlate significantly with tumor grade or discriminate between different receptor statuses.

miR-221, encoded on human chromosome X, is overexpressed in many aggressive carcinomas (Galardi et al., 2007; Waters et al., 2012; Nassirpour et al., 2013; Wang et al., 2013), including breast cancer (Radojicic et al., 2011; Waters et al., 2012; Nassirpour et al., 2013). Furthermore, an elevated expression level of miR-221 in certain carcinomas facilitates invasion (Waters et al., 2012), a larger tumor size (Li et al., 2011), early metastasis (Liu et al., 2012), and a shorter time to recurrence (Kang et al., 2012). Fornari et al. (2008) established a potential oncogenic function of miR-221, which is upregulated in Hepatocellular carcinoma (HCC). It has been reported that miR-221 targets the cyclin-dependent kinase inhibitors CDKN1B/p27 and

CDKN1C/p57. Upregulation of miR-221 causes a downregulation of these inhibitors and promotes the loss of cell cycle control (le Sage et al., 2007; Pineau et al., 2010). Other studies have shown that miR-221 regulates two key mechanisms that promote the aggressive tumorigenic characteristics observed in triple negative breast cancer (TNBC): it promotes cell cycle progression by inhibiting the protein cyclin-dependent kinase (p27kip1) and promotes epithelial–mesenchymal transition (EMT) by inhibiting the expression of E-cadherin. Both of these mechanisms may account for the aggressive cellular proliferation, suppression of apoptosis, as well as higher cell migration and invasiveness associated with basal-like breast cancer (BLBCs) and TNBCs (Dudda et al., 2013; Manavalan et al., 2013). Previous research has shown that miR-221 is involved in suppressing ERα expression in luminal breast cancer cells and EMT transition in basal-like breast cancers (Miller et al., 2008; Rao et al., 2011). Moreover, it has been reported that miR-221 is involved in the promotion of an aggressive basal-like breast cancer phenotype, functioning downstream of the RAS pathway and triggering epithelial-to-mesenchymal transition (EMT) (Wang et al., 2011). Other studies have shown the predictive role of miR-221 in resistance to neoadjuvant chemotherapy (Kawaguchi et al., 2013; Wurz et al., 2010).

In conclusion, our data and statistical analysis indicate that miR-21 and miR-221 are indeed good candidates to be used as molecular diagnostic markers for breast cancer, specially when their expression levels is combined with machine learning algorithms to accurately predict the patient categories. An important finding is that the overexpression of miR-221 is associated with the breast cancer grade. In addition, miR-221 has more diagnostic power in discriminating between breast cancer and fibroadenoma patients, which makes it a potential marker that may enhance the discriminating power of this plasma quantitation test in the future.

Acknowledgements

The authors acknowledge the Medical Biochemistry and Molecular Biology Department, Faculty of Medicine, Cairo University.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Hortobagyi, G. N., de la Garza Salazar, J., Pritchard, K., Amadori, D., Haidinger, R., Hudis, C.A., Khaled, H., Liu, M.C., Martin, M., Namer, M., et al., 2005. The global breast cancer burden: Variations in epidemiology and survival. Clin. Breast Cancer. 6:391 401.

2. Gill, J. K., Maskarinec, G., Wilkens, L. R., Pike, M. C., Henderson, B. E., Kolonel, L. N., Gill, J. K., Maskarinec, G., Wilkens, L. R., Pike, M. C., et al., 2007. Nonsteroidal antiinflammatory drugs and breast cancer risk: The multiethnic cohort. Am. J. Epidemiol. 166:1150 1158.

3. Porter, P., 2008. “Westernizing” women's risks? Breast cancer in lower-income countries. N Engl. J. Med. 358:213–216.

4. Beaglehole, R., Yach, D., 2003. Globalisation and the prevention and control of non-communicable disease: The neglected chronic diseases of adults. Lancet. 362, 903 908.

5. Parkin D. M., Fernández, L. M., 2006. Use of statistics to assess the global burden of breast cancer. Breast J., 12 Suppl 1:S70-80.

6. Badar, F., Faruqui, Z. S., Ashraf, A., Uddin, N., 2007. Third world issues in breast cancer detection. J. Pak. Med. Assoc. 57, 137-140.

7. Rizwan, M. M.; Saadullah, M., 2009. Lack of awareness about breast cancer and its screening in developing countries. Indian J. Cancer. 46, 252-253.

8. Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer; 2013. http://globocan.iarc.fr, accessed on 13/07/2015.

9. Bray, F., Ren, J. S., Masuyer, E., Ferlay, J., 2008. Estimates of global cancer prevalence for 27 sites in the adult population in. Int. J. Cancer. 132(5): 1133-45.

10. Elatar I. Cancer Registration, NCI Egypt 2001; National Cancer Institute: Cairo, Egypt; 2002

11. Ibrahim, A.S.; Komodiki, C.; Najjar, K.; Rahamimoff, R.; Tuncer, M., 2002. Cancer Profile in Gharbiah, Egypt. Methodology and Results; Ministry of Health and Population Egypt and Middle East Cancer Consortium: Cairo, Egypt.

12. Omar, S., Khaled, H., Gaafar, R., Zekry, A. R., Eissa, S., El-Khatib, O., 2003. Breast cancer in Egypt: A review of disease presentation and detection strategies. East Mediter. Health J. 9: 448–463.

13. Dixon, J. M., 1991. Cystic diseases and fibroadenoma of the breast: natural history and relation to breast cancer risk. Br. Med. Bull. 47(2):258±71.

14. Fechner, R. E. 1988. Fibroadenoma and related lesions. In: Page DL, Anderson TJ, eds. Diagnostic Histopathology of the Breast. Edinburgh, Scotland: Churchill Livingstone. 72±85.

15. Hunter, B. T., Roberts, C. C., Hunt, K. R., Fajardo, L. L., 1996. Occurrence of fibroadenomas in postmenopausal women referred for breast biopsy. J. Ageing Geriatr. 44:61±4.

16. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T., 2001. Identification of novel genes coding for small expressed RNAs. Science 294: 853-858.

17. Lau, N. C., Lim, L. P., Weinstein, E. G., Bartel, D. P., 2001. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 294: 858-862.

18. Lee, R. C., Ambros, V., 2001. An extensive class of small RNAs in Caenorhabditis elegans. Science 294: 862-864.

19. Li, M., Li, J., Ding, X., He, M., Cheng, S.Y., 2010. MicroRNA and cancer. AAPS J. 12:309–317.

20. Lewis, B. P., Burge, C. B., Bartel, D. P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 120:15–20.

21. Kosaka, N., Iguchi, H., Ochiya, T., 2010. Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. Cancer Sci. 101:2087–2092.

22. Blenkiron, C., et al., 2007. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. Genome Biology, 8, R214.

23. Iorio, M. V., Ferracin, M., Liu, C., et al., 2005. MicroRNA Gene Expression Deregulation in Human Breast. Cancer Res. 65:7065-7070.

24. Chan, J. A., Krichevsky, A. M., Kosik, K. S., 2005. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. Cancer Res. 65(14):6029-33.

25. Shah, M. Y. and Calin, G. A., 2011. MicroRNAs miR-221 and miR-222: a new level of regulation in aggressive breast cancer. Genome Med. 3(8): 56.

26. Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebet, B. L., Mak, R. H., Ferrando, A. A., Downing, J. R., Jacks, T., Horvitz, H. R., Golub, T. R., 2005. MicroRNA expression profiles classify human cancers. Nature, 435, pp. 834–838.

27. Barshack, I., Lithwick-Yanai, G., Afek, A., Rosenblatt, K., Tabibian-Keissar, H., Zepeniuk, M., et al. 2010. MicroRNA expression differentiates between primary lung tumors and metastases to the lung. Pathol Res Pract. 206(8):578-84

28. https://www.manning.com/books/machine-learning-in-action

29. http://www.statsoft.com/Textbook

30. http://www.amazon.co.uk/Data-Mining-Concepts-Methods-Algorithms/dp/0470890452

31. Livak, K. J., Schmittgen T. D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. (4):402-8.

32. Shenouda, S. K., Alahari, S. K., 2009. MicroRNA function in cancer: oncogene or a tumor suppressor? Cancer Metastasis Rev 28:369– 378.

33. Garofalo, M., Condorelli, G. L., Croce, C. M., Condorelli, G., 2010. MicroRNAs as regulators of death receptors signaling. Cell Death Differ. 17:200–208.

34. Kouhkan, F., Alizadeh, S., Kaviani, S., Soleimani,M. Pourfathollah AA, Amirizadeh N, Abroun S, Noruzinia M, Mohamadi S., 2011. MiR-155 Down Regulation by LNA Inhibitor can Reduce Cell Growth and Proliferation in PC12 Cell Line. Avicenna J Med Biotechnol. (2):61-6.

35. Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L., et al., 2008. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci U S A. 105(30):10513-8.

36. Gallo, A., Tandon, M., Alevizos, I., Illei, G. G., 2012. The majority of microRNAs detectable in serum and saliva is concentrated in exosomes. PLoS One. 7(3):e30679.

37. Mathivanan, S., Ji, H., Simpson, R. J., 2010. Exosomes: extracellular organelles important in intercellular communication. J Proteomics. 73:1907–1920.

38. Ohshima, K., Inoue, K., Fujiwara, A., Hatakeyama, K., Kanto, K., Watanabe, Y., Muramatsu, K., Fukuda, Y., Ogura, S., Yamaguchi, K., et al., 2010. Let-7 microRNA family is selectively secreted into the extracellular environment via exosomes in a metastatic gastric cancer cell line. PLoS One 5, e13247.

39. Han, M., Liu, M., Wang, Y., Mo, Z., Bi, X., Liu, Z., et al., 2012. Re-expression of miR-21 contributes to migration and invasion by inducing epithelial-mesenchymal transition consistent with cancer stem cell characteristics in MCF-7 cells. Mol Cell Biochem 363:427-36.

40. Han, M., Liu, M., Wang, Y., Chen, X., Xu, J., Sun, Y., et al., 2012. Antagonism of miR-21 reverses epithelial-mesenchymal transition and cancer stem

cell phenotype through AKT/ERK1/2 inactivation by targeting PTEN. PLoS One 7:e39520.

41. Si, M. L., Zhu, S., Wu, H., Lu, Z., Wu, F., Mo, Y. Y., 2007. MiR-21-mediated tumor growth. Oncogene 26:2799- 803.

42. Yan, L. X., Huang, X. F., Shao, Q., Huang, M. Y., Deng, L., Wu, Q. L., et al., 2008. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. RNA 14:2348-60.

43. Zhu, S., Wu, H., Wu, F., Nie, D., Sheng, S., Mo, Y. Y., 2008. MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. Cell Res 18:350-9.

44. Brew, K., Dinakarpandian, D., Nagase, H., 2000. Tissue inhibitors of metalloproteinases: evolution, structure and function. Biochim Biophys Acta. 1477:267-83.

45. Bailey, C. M., Khalkhali-Ellis, Z., Seftor, E. A., Hendrix, M. J., 2006. Biological functions of maspin. J Cell Physiol. 209:617-24.

46. Song, M. S., Salmena, L., Pandolfi, P. P., 2012. The functions and regulation of the PTEN tumour suppressor. Nat. Rev. Mol. Cell Biol. 13:283-96.

47. Frankel, L. B., Christoffersen, N. R., Jacobsen, A., et al., 2008. Programmed Cell Death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer Cells. J. Biol. Chem. 283, 1026-33.

48. Zhu, S., Si, M. L., Wu, H., Mo, Y. Y., 2007. MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). J. Biol. Chem. 282:14328-36.

49. Perry, S. V., 2001. Vertebrate tropomyosin: distribution, properties and function. J. Muscle Res Cell Motil. 22, 5-49.

50. Corcoran, C., Friel, A. M., Duffy, M. J., Crown, J., O’Driscoll, L., 2011. Intracellular and extracellular microRNAs in breast cancer. Clin. Chem. 57, 18–32.

51. Krichevsky, A. M., Gabriely, G., 2009. miR-21: a small multi-faceted RNA. J. Cell. Mol. Med. 13(1): 39-53.

52. Galardi, S., Mercatelli, N., Giorda, E., Massalini, S., Frajese, G. V., et al., 2007. miR- 221 and miR-222 expression affects the proliferation potential of human prostate carcinoma cell lines by targeting p27Kip1. J. Biol. Chem. 282: 23716–23724.

53. Waters, P. S., McDermott, A. M., Wall, D., Heneghan, H. M., Miller, N., et al., 2012. Relationship between circulating and tissue microRNAs in a murine model of breast cancer. PLoS One 7 (11): e50459.

54. Nassirpour, R., Mehta, P. P., Baxi, S. M., Yin, M. J., 2013. miR-221 Promotes Tumorigenesis in Human Triple Negative Breast Cancer Cells. PLoS One 8 (4): e62170.

55. Radojicic J, Zaravinos A, Vrekoussis T, Kafousi M, Spandidos DA, et al., 2011. MicroRNA expression analysis in triple-negative (ER, PR and Her2/neu) breast cancer. Cell Cycle 10 (3): 507–17

56. Wang, Z., Zhang, H., He, L., Dong, W., Li, J., et al., 2013. Association between the expression of four upregulated miRNAs and extrathyroidal invasion in papillary thyroid carcinoma. Onco. Targets Ther 6: 281–7.

57. Li, J., Wang, Y., Yu, W., Chen, J., Luo, J., 2011. Expression of serum miR-221 in human hepatocellular carcinoma and its prognostic significance. Biochem Biophys Res Commun 406 (1): 70–3.

58. Liu, K., Li, G., Fan, C., Diao, Y., Wu, B., et al., 2012. Increased Expression of MicroRNA-221 in gastric cancer and its clinical significance. J. Int. Med. Res 40 (2): 467–74

59. Kang, S. G., Ha, Y. R., Kim, S. J., Kang, S. H., Park, H. S., et al., 2012. Do microRNA 96, 145 and 221 expressions really aid in the prognosis of prostate carcinoma? Asian J. Androl 14 (5): 752–7.

60. Fornari, F., Gramantieri, L., Ferracin, M., Veronese, A., Sabbioni, S., Calin, G. A., et al., 2008. miR 221 controls CDKN1C/p57 and CDKN1B/p27 expression in human hepatocellular carcinoma. Oncogene, 27:5651– 61.

61. Le Sage, C., Nagel, R., Egan, D. A., Schrier, M., Mesman, E., et al., 2007. Regulation of the p27(Kip1) tumor suppressor by miR-221 and miR-222 promotes cancer cell proliferation. EMBO J 26 (15): 3699–708.

62. Pineau, P., Volinia, S., McJunkin, K., Marchio, A., Battiston, C., et al., 2010. miR-221 overexpression contributes to liver tumorigenesis. Proc. Natl. Acad. Sci. U S A 107 (1): 264–9.

63. Dudda, J. C., Salaun, B., Ji, Y., Palmer, D. C., Monnot, G. C., et al., 2013. MicroRNA-155 Is Required for Effector CD8(+) T Cell Responses to Virus Infection and Cancer. Immunity 38 (4): 742–53.

64. Manavalan, T. T., Teng, Y., Litchfield, L. M., Muluhngwi, P., Al-Rayyan, N., et al., 2013. Reduced Expression of miR-200 Family Members

Contributes to Antiestrogen Resistance in LY2 Human Breast Cancer Cells. PLoS One 8 (4): e62334.

65. Miller, T. E., Ghoshal, K., Ramaswamy, B., Roy, S., Datta, J., et al., 2008. MicroRNA-221/222 confers tamoxifen resistance in breast cancer by targeting p27Kip1. J. Biol. Chem. 283: 29897–29903.

66. Rao, X., Di Leva, G., Li, M., Fang, F., Devlin, C., et al., 2011. MicroRNA-221/222 confers breast cancer fulvestrant resistance by regulating multiple signaling pathways. Oncogene 30: 1082–1097.

67. Wang, Y., Zhou, B. P., 2011. Epithelial-mesenchymal transition in breast cancer progression and metastasis. Chinese Journal of Cancer. 30(9):603–611.

68. Kawaguchi, T., Komatsu, S., Ichikawa, D., Morimura, R., Tsujiura, M., et al., 2013. Clinical impact of circulating miR-221 in plasma of patients with pancreatic cancer. Br. J. Cancer 108 (2): 361–9.

69. Wurz, K., Garcia, R. L., Goff, B. A., Mitchell, P. S., Lee, J. H., et al., 2010. MiR-221 and MiR-222 alterations in sporadic ovarian carcinoma: Relationship to CDKN1B, CDKNIC and overall survival. Genes Chromosomes Cancer 49 (7): 577–84.

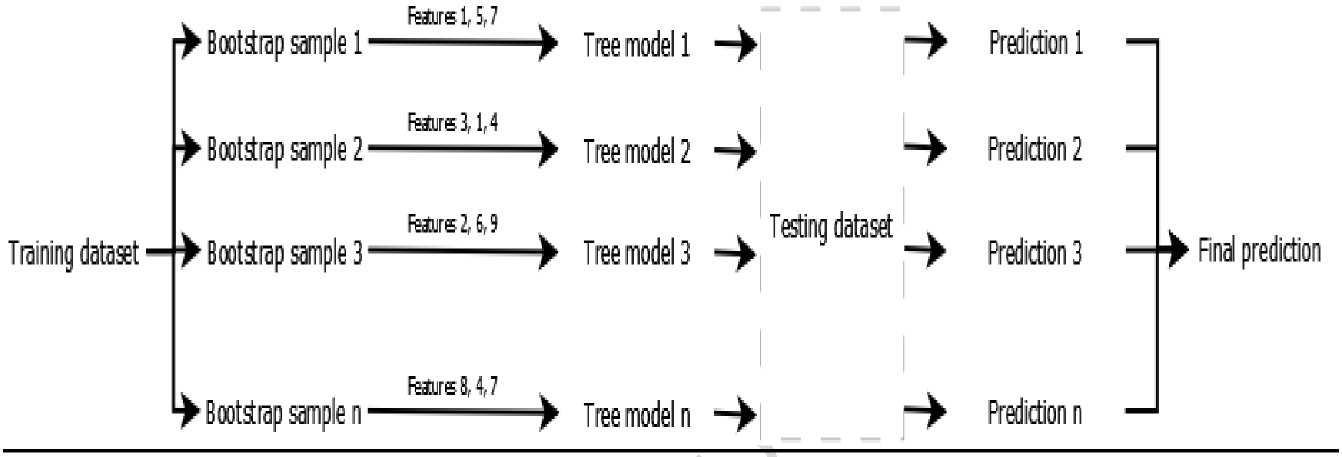


Fig. 1. Random forest algorithm summary.

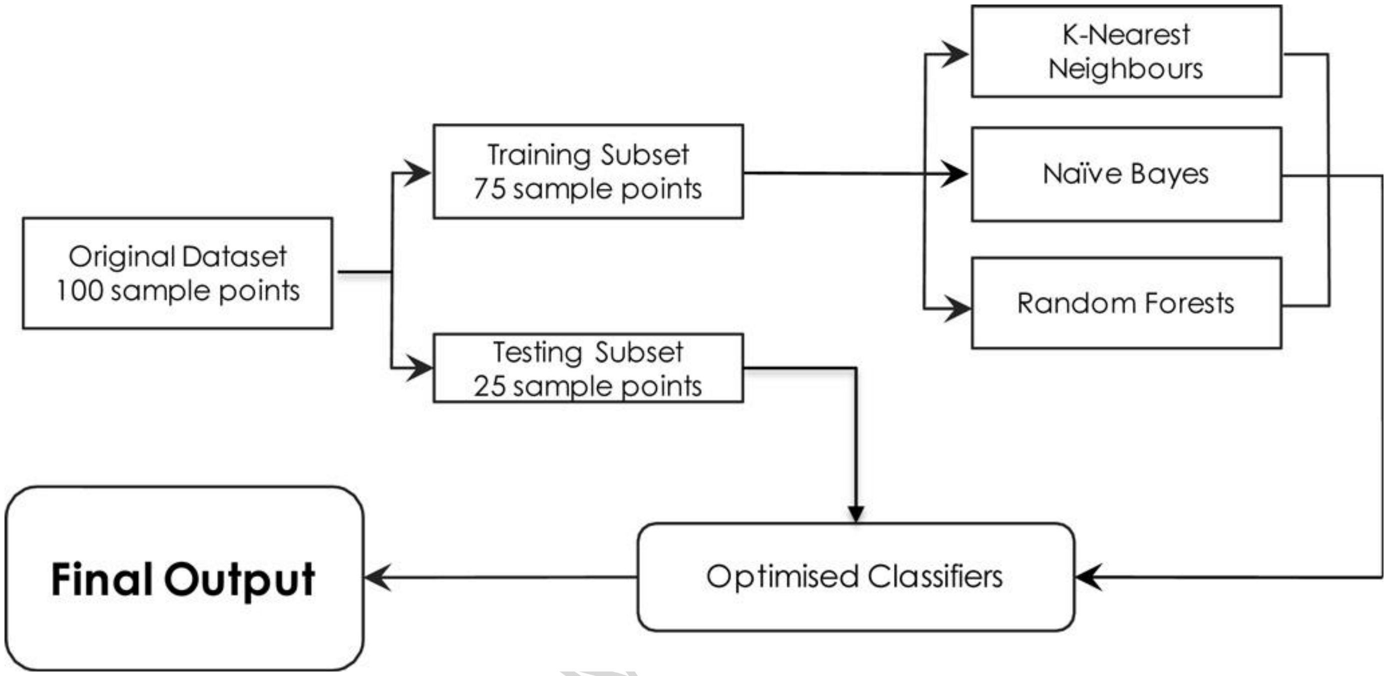
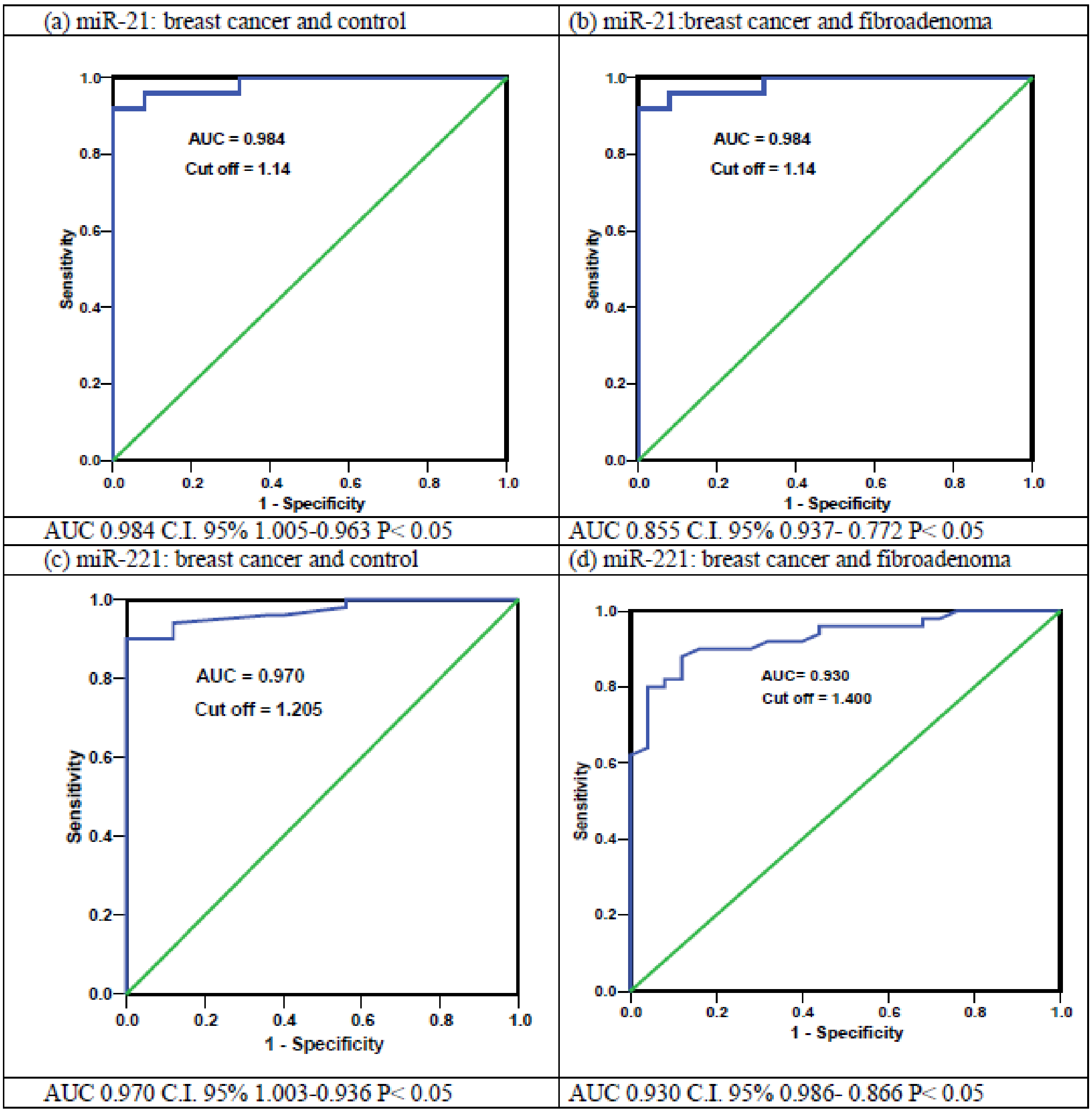

K-Nearest Neighbours
Naïve Bayes
Random Forests
Training Subset 75 sample points
Original Dataset 100 sample points
Testing Subset 25 sample points
Final Output
Optimised Classifiers

Fig. 2. Flowchart illustrating the development process for machine learning classifiers. The original dataset is split randomly into training and testing subset. The split algorithm ensures enough representable samples within each class. The optimum parameters are used to build the optimized classifiers, which is then used to assess the prediction accuracy of the testing subset; which is a representative of unknown dataset.



(a) miR-21: breast cancer and control

AUC 0.984 C.I. 95% 1.005-0.963 P< 0.05

(b) miR-21:breast cancer and fibroadenoma

AUC 0.855 C.I. 95% 0.937- 0.772 P< 0.05

(c) miR-221: breast cancer and control

AUC 0.970 C.I. 95% 1.003-0.936 P< 0.05

(d) miR-221: breast cancer and fibroadenoma

AUC 0.930 C.I. 95% 0.986- 0.866 P< 0.05

Fig.3. P of the ROC of miR-21 (a) and miR-221 (c) between breast cancer patients and control groups, and P of the ROC of miR-21 (b) and miR-221 (d) between breast cancer patients and fibroadenoma groups

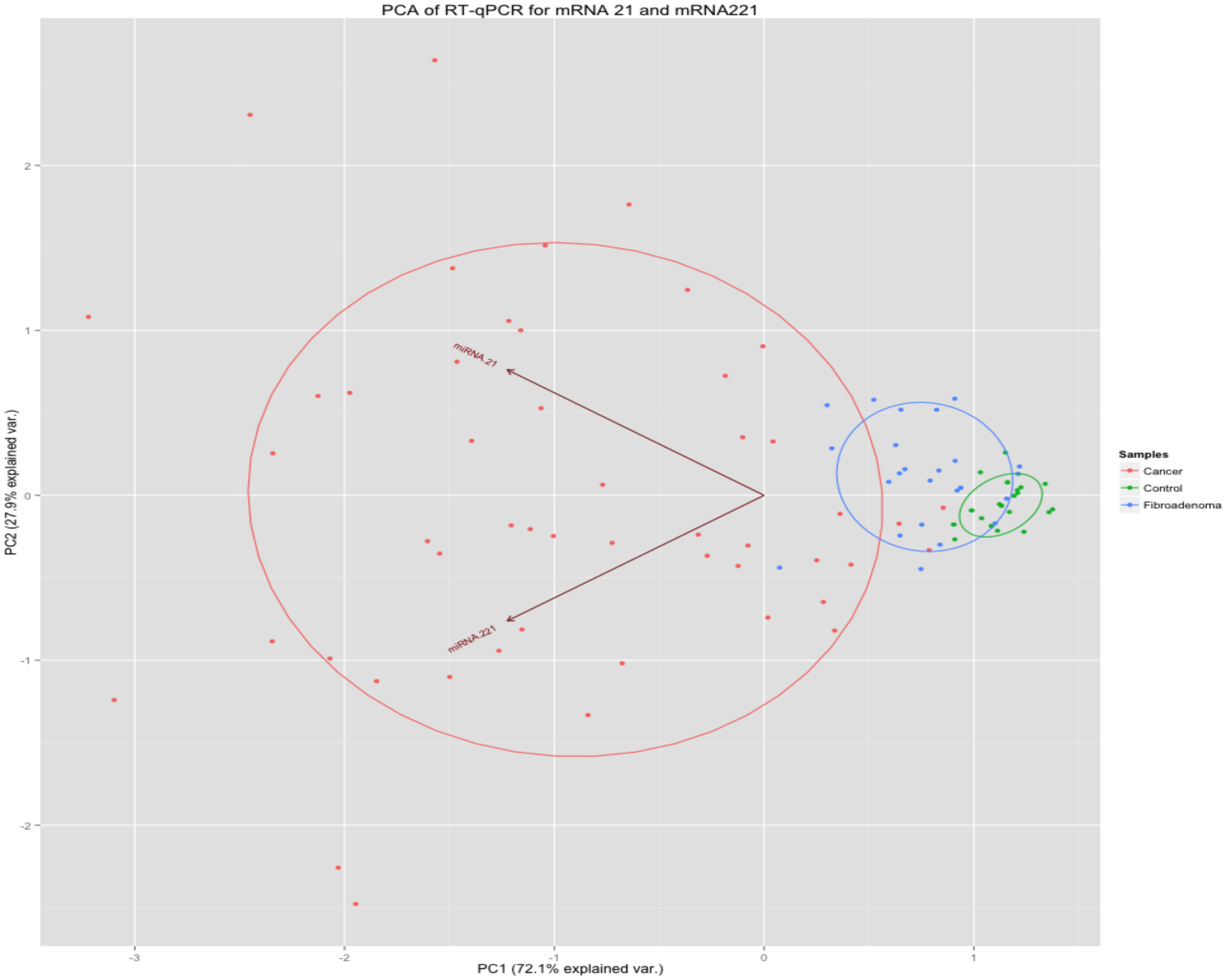


Fig. 4. Multivariate analysis using a biplot principal component analysis (PCA) using miRNA expression to differentiate between breast cancer, fibroadenoma and normal controls. The red dots indicate the cancer group, the blue dots indicate fibroadenoma patients, while the green dots indicate the control samples. The first principle components (PC1) accounted for 72.1%, while the second principle component (PC2) accounts for 27.9% of the variance (Total 100% of variance).

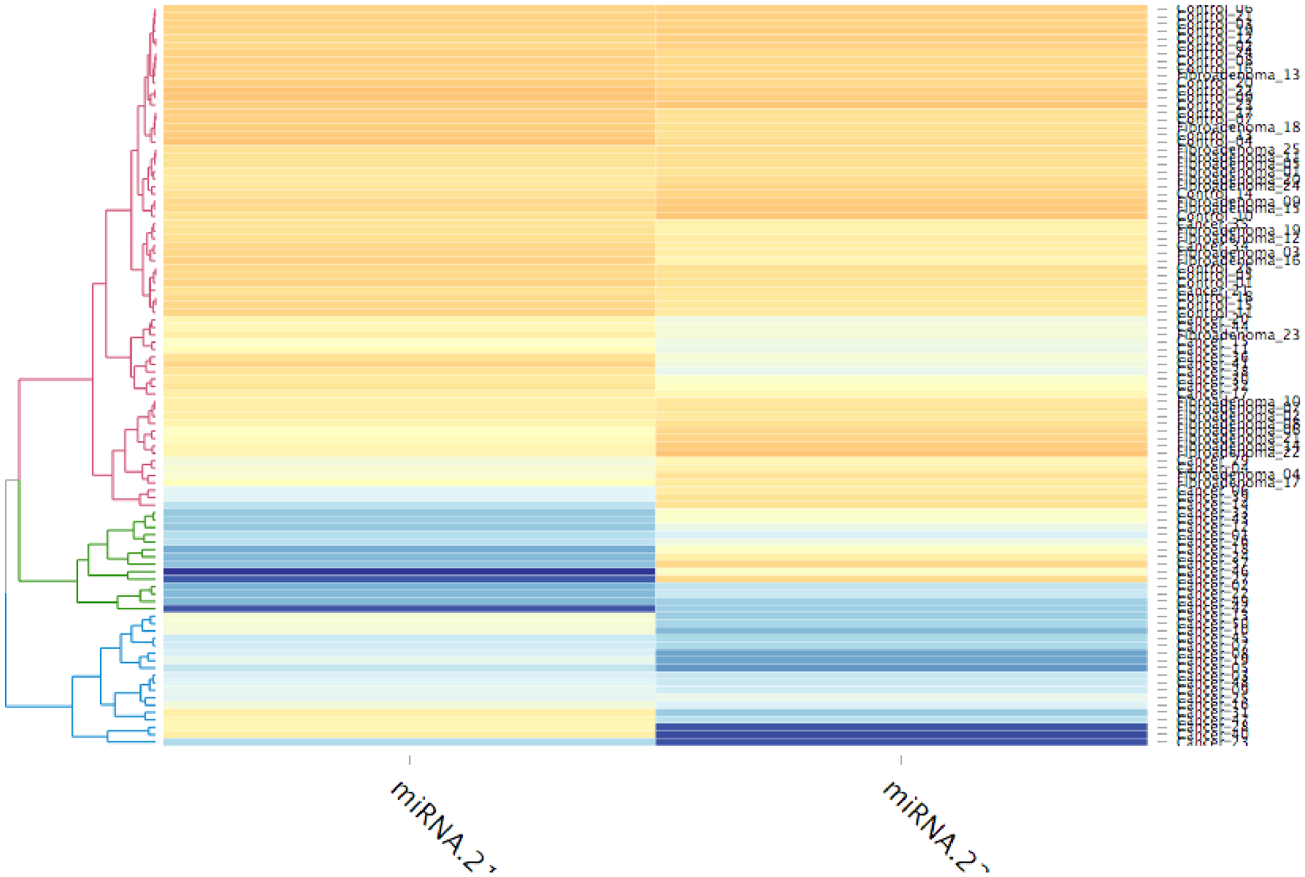


Fig. 5. Heatmap clusters showing expression levels for miR-21 and miR-221 across 100 samples. Color ramp indicate the expression level; ranging from yellow (low expression) to dark blue (high expression. Each row represents an individual, and each column represents the corresponding miRNA.

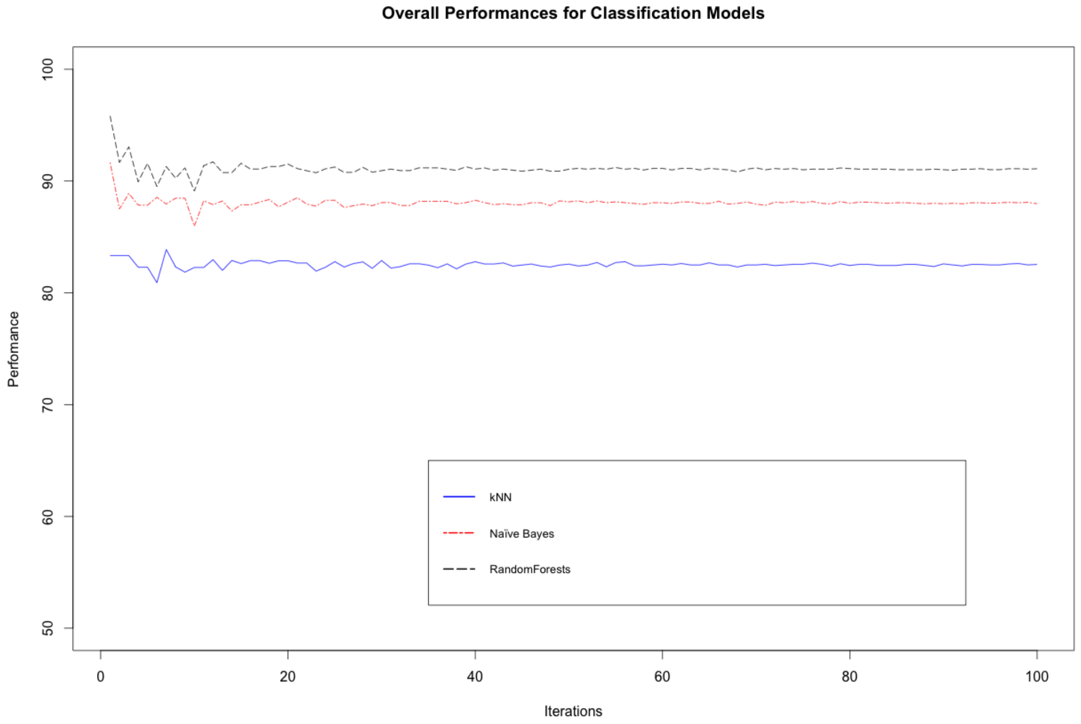


Fig. 6. Overall performances of the k-Nearest Keibours (kNN), Naïve Bayes (NB), and Random Forests (RF) models. The best prediction accuracy was achieved by RF, with an average of 91%, followed by NB at 88%, and kNN at 82.3%

Table 1. Clinicopathological characteristics of breast cancer patients

Parameters	**Patients frequency**
Total	50
Age	Range (37- 70 years)
Less than 55	*31*
More than 55	*19*
Distant metastasis	
M0	*50*
Family History	
Positive	*28*
Negative	*22*
Tumor type	
Invasive Ductal Carcinoma	*46*
Invasive Lobular Carcinoma	*4*
Grading	
G I, II	*37*
G III	*13*
Tumor Stage	
T2	*35*
T3	*15*
Lymph node metastasis	
N 1	*6*
N 2	*30*
N3	*14*
Estrogen receptor	
Positive	*8*
Negative	*42*
Progesterone receptor	
Positive	*7*
Negative	*43*

Table 2. Demographic and clinical features of study groups

Parameters		Control N=25	Fibroadenoma N=25	Breast cancer N=50	P value
Age		28.6 ± 8.5	32.1 ±14.4	53.5 ± 7.5^{ab}	**<0.001***
Menstrual history	*Pre-menopause*	23	21	13	**<0.0001***
	Post-menopause	2	4	37	
Family history	*Yes*	-	5	28	**<0.0001***
	No	25	20	22	
Diabetes	*Yes*	-	-	13	**<0.001***
	No	25	25	37	
Hypertension	*Yes*	-	2	12	**0.007***
	No	25	23	38	

Values are expressed as the means ± S.D (age) or frequency. ***** Indicates statistical significance. P values < 0.05 are considered significant.

Clinical data were analyzed by a. ANOVA and b. Chi square test and Fisher's exact test.

a Statistical significance from the control group
b Statistical significance from the fibroadenoma group

Table 3. Relative expression level of serum miR-21 and miR-221 in breast cancer patients, fibroadenoma patients, and the control group

Variables	**No.**	**miR-21**	**miR-221**
Control	25	1.0 ± 0.1	1.1 ± 0.1
Fibroadenoma	25	1.3 ± 0.2	1.2 ± 0.2
Breast Cancer	50	2.2 ± 0.8[a, b]	2.3 ± 0.8[a, b]

All values are expressed as mean ± SD
a. Statistical significance from control group
b. Statistical significance from fibroadenoma group
Significance at $P < 0.05$

Table 4. Correlation between the relative expression of serum miRNA values and patient clinicopathological characteristics at the time of primary breast cancer diagnosis

Parameters	miR- 21	P value	miR- 221	P value
Age				
Less than 55	2.2 ± 0.8	0.582	2.2 ± 0.9	0.147
More than 55	2.1 ± 0.8		2.1 ± 0.7	
Distant metastasis				
M0	2.2 ± 0.8		2.3 ± 0.8	
Family History				
Positive	2.1 ±0.6	0.868	2.2 ± 0.8	0.646
Negative	2.2 ± 1.0		2.4 ± 0.9	
Tumor type				
Invasive Ductal Carcinoma	2.2 ± 0.8	0.569	2.2 ± 0.7	0.357
Invasive Lobular Carcinoma	1.9 ± 0.5		3.0 ± 1.6	
Grading				
G I&II	2.2 ± 0.8	0.912	2.1 ± 0.7	**0.038***
G III	2.1 ± 0.6		2.8 ± 1.1	
Tumor Stage				
T2	2.3 ± 0.8	0.112	2.3 ± 0.8	0.824
T3	1.9 ± 0.6		2.3 ± 0.8	
Lymph node metastasis				
N 1	2.0 ± 0.8	0.877	1.9 ± 0.5	0.344
N 2	2.2 ± 0.8		2.3 ± 1.0	
N3	2.1 ± 0.7		2.4 ± 0.7	
Estrogen receptor				
Positive	2.4 ± 1.0	0.427	2.9 ± 1.0	0.099
Negative	2.1 ± 0.7		2.2 ± 0.7	
Progesterone receptor				
Positive	2.2 ± 0.8	0.870	3.0 ± 1.0	0.056
Negative	2.2 ± 0.8		2.2 ± 0.7	

All values are expressed as mean ± SD**.**
Mann-Whitney U and Kruskal-Wallis nonparametric test were used for comparing different groups.
P*: Indicates statistical significance at $P < 0.05$

Table 5. Confusion matrices for the training and testing subsets

KNN	Training Confusion Matrix				Testing Confusion Matrix		
predicted	**Cancer**	**Control**	**Fibroadenoma**	**predicted**	**Cancer**	**Control**	**Fibroadenoma**
Cancer	36	1	4	**Cancer**	12	0	0
Control	1	16	5	**Control**	0	5	2
Fibroadenoma	1	2	10	**Fibroadenoma**	0	1	4
Naïve Bayes	**Training Confusion Matrix**				**Testing Confusion Matrix**		
predicted	**Cancer**	**Control**	**Fibroadenoma**	**predicted**	**Cancer**	**Control**	**Fibroadenoma**
Cancer	38	0	0	**Cancer**	11	0	0
Control	0	19	0	**Control**	0	5	1
Fibroadenoma	0	0	19	**Fibroadenoma**	1	1	5
Random Forests	**Training Confusion Matrix**				**Testing Confusion Matrix**		
predicted	**Cancer**	**Control**	**Fibroadenoma**	**predicted**	**Cancer**	**Control**	**Fibroadenoma**
Cancer	38	0	0	**Cancer**	12	0	0
Control	0	19	0	**Control**	0	5	0
Fibroadenoma	0	0	19	**Fibroadenoma**	0	1	6

Abbreviations

- ANOVA: Analysis of variance.
- AUC: Area under the curve.
- BLBC: Basal-like breast cancer.
- CDK: Cyclin-dependent kinases.
- CI: Confidence interval.
- Ct: Cycle threshold.
- EMT: Epithelial–mesenchymal transition.
- ER: Estrogen receptor.
- HCC: Hepatocellular carcinoma.
- kNN: k-Nearest Neighbor.
- maspin: Mammary serine protease inhibitor.
- miR-21: microRNA-21.
- miR-221: microRNA-221.
- miRNAs: microRNAs.
- mRNA: messenger RNA.
- NB: Naïve Bayes.
- PBS: Phosphate buffer saline.
- PCA: Principle component analysis.
- PDCD4: Programmed cell death 4**.**
- RF: Random Forest.
- ROC: Receiver Operating Characteristic.
- RT-PCR: Real-time Polymerase chain reaction.
- RT: Reverse transcription.
- SD: Standard deviation.
- TNBC: Triple negative breast cancer.
- TPM1: Tropomyosin 1.

Research Highlights

- MiRNA- 21 and 221 can significantly differentiate between breast cancer and healthy controls.
- The diagnostic accuracy of serum miRNA-21 is superior than miRNA-221 for breast cancer prediction.
- MiRNA-221 has more diagnostic power than miRNA-21 in discriminating between breast cancer and fibroadenoma patients.