A panel of four miRNAs (miR-190b, miR-584-5p, miR-452-5p, and miR-1306-5p) is capable of classifying luminal and non-luminal breast cancers.

Faranak Farahmand[1¶], Saied Rahmani[2,6¶], Hadi Bayat[1], Adel Salimi[2], Sogol Ghanbari[1], Afsaneh Malekzadeh Shafaroudi[3], Ali Sharifi-Zarchi[2], Mohammad Vasei[4], Seyed-Javad Mowla[1,5*]

Affiliations:

1. Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran
2. Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
3. Department of Biology, Faculty of Science, Ferdowsi University of Mashhad, Mashhad, Iran
4. Department of Pathology, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran
5. TechAzma Company, Tehran, Iran
6. Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

¶ Co-First authors, these authors contributed equally to this work

* Corresponding author: Dr. Seyed Javad Mowla, Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Jalal Al Ahmad, Nasr, Tehran, Iran. Tel.: +9821-82883464; Fax: +9821-82884717; E-mail: sjmowla@modares.ac.ir

Abstract.

BACKGROUND: Identifying the molecular subtypes of breast cancer (BC) plays a crucial role in enhancing the efficacy of therapy. MiRNAs (miRs) with differential expressions in different subtypes of breast tumors can be considered as non-invasive biomarkers for diagnosing BC subtypes.

OBJECTIVE: We aimed to investigate the efficacy of miR-190b, miR-584-5p, miR-452-5p, and miR-1306-5p as novel potent diagnostic biomarkers in discriminating patients with luminal (ER+) and non-luminal (ER–) BCs.

METHODS: A group of miRs significantly associated with estrogen cell receptors (ER) in breast tumors were identified using feature selection methods analysis on miR-Seq datasets retrieved from TCGA and GSE68085. Four abovementioned miRs were selected as novel potential biomarkers, and their relative expression levels were assessed within adjacent non-tumor, ER+ and ER– tumor tissues by quantitative RT-PCR. Their impact on diagnosis was also evaluated by ROC curve analysis.

RESULTS: In ER+ BCs compared to ER– BCs, the expression of miR-190b was remarkably increased, while the expression of miR-584-5p, miR-452-5p, and miR-1306-5p were significantly decreased. This group could discriminate ER+ and ER– BCs at an AUC of 0.973.

CONCLUSIONS: According to our findings, these four miRs are promising biomarkers in discriminating BC subtypes. The candidate miRs in parallel with histologic diagnosis methods can be applied for identifying patients who are most likely responding to specific therapies based on ER status.

Keywords: Breast Cancer, Luminal tumor, non-Luminal tumor, Estrogen Receptors, miRNAs, Biomarker

1 Introduction

Breast cancer (BC) has been the leading cause of cancer death (15.5%) and the most common cancer in females (24.5%) worldwide [1]. Despite significant improvements in the understanding of cancer pathogenesis, screening programs for early diagnosis, and treatment over the past few decades, there are still about 2.2 million new cases of BCs with more than half a million BC-related deaths recorded annually worldwide [1]. BC has heterogeneous nature with various morphologic signs and clinical outcomes which can be categorized in several aspects, including clinical features, expression of tumor markers, and histologic types [2, 3]. In this regard, gene expression profile analyses have led to the classification of BCs based on hormone receptors, including estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor (HER2 or HER2/neu) status. The hormone receptors are crucial factors in therapeutic prediction and should be measured on all newly diagnosed BC tumors [4-7]. ER and PR mediate mammary cell proliferation signals and stimulate the growth of both normal and neoplastic breasts [8]. Likewise, HER2 is a transmembrane receptor tyrosine kinase in the epidermal growth factor receptor family [8]. Accordingly, the intrinsic molecular subtypes of BC were classified as luminal A (ER+/PR±/HER2–), luminal B (ER+/PR±/HER2+), HER2 enriched (ER–/PR–/HER2+), basal-like, and normal-like, which show specific biological features and clinical outcomes [6].

In the clinical aspect, the ER expression is frequently examined by immunohistochemistry of biopsy to separate BC subtypes. This examination provides information on prognosis and the possibility of response to endocrine therapy [9]. However, the main disadvantage of this assay is false-negative results, which contribute to patients being denied for hormone therapy. Furthermore, diagnosis based on tissue biopsy is also an invasive process with the risk of

spreading tumor cells to adjacent tissues [10, 11]. Therefore, considering the limitations of common diagnostic tests, developing new assays with high sensitivity and specificity for BC subtypes determination is indispensable. In this regard, biomarkers are now investigated in cancer studies for the purpose of diagnosis, prognosis, and therapy [12].

MiRNAs (miRs) are a class of small non-coding RNAs with 19–25 nucleotides length. The miRs can post-transcriptionally regulate gene expression by binding to the target mRNA and affecting many biological processes such as differentiation, proliferation, apoptosis, and metastasis [13]. Dysregulated miRs play primary roles in cancer initiation and progression [12, 13]. During recent years, an increasing number of miRs acting either as oncogenes or tumor suppressors have been investigated in BC [14]. Due to remarkable stability and easily non-invasive detection in biological fluids, such as serum and plasma, miRs are considered as promising biomarkers in cancer diagnosis and prognosis [12, 15-18]. Hence, numerous studies have reported expression patterns of miRs as an informative tool for the classification of BCs [19, 20]. Although several studies have been tried to discover biomarkers based on miRs, the exploration of novel groups of miRs with high sensitivity and specificity for the diagnosis of BC subtypes is valuable for developing diagnosis strategies, specific treatment, and disease management. The main purpose of the present study was to identify novel promising miR biomarkers which are associated with the presence of ER in BCs. For this aim, we utilized boruta [21], XGBoost [22], and limma [23] R packages for TCGA and GEO datasets analysis. Moreover, quantitative RT-PCR (qRT-PCR) method was used to validate experimentally the efficacy of our candidate miR biomarkers in discriminating luminal (ER+) from non-luminal (ER–) BCs. Furthermore, we evaluated the role of the candidate miRs as biomarker for BC diagnosis by measuring their expression in breast tumor and non-tumor samples of the TCGA dataset and those samples which were used in the

qRT-PCR analysis. Finally, we confirmed the efficacy of miR-190b (miR-190b is classified as miR-190b-3p and -5p in the latest version of miRBase sequence database (release 22.1), Here, our candidate is miR-190b-5p), miR-584-5p, miR-452-5p, and miR-1306-5p as the promising biomarkers of luminal (ER+) and non-luminal (ER–) breast tumors.

2 Materials and methods

2.1. Analysis of miR-Seq datasets

Raw read counts of miR-seq dataset of BC patients along with their clinical dataset from TCGA [24] were obtained and analyzed by TCGAbiolinks package [25]. The total number of samples in this dataset was 1175 (1072 tumor and 103 adjacent non-tumor samples). Intrinsic subtype labels were assigned to samples according to the expression level of ER, PR, and HER2 from IHC test results reported in clinical data. Information about the BC subtypes based on the cell receptors (ER, PR, and HER2) was presented in positive/negative and level of existence format. In addition to TCGA, raw read counts of the miR-seq dataset of patients with TNBC and luminal (ER+) BCs were obtained from GSE68085 [26].

Here, we applied boruta feature selection, XGBoost feature selection, and limma differential expression analysis on TCGA and GSE68085 datasets. To merge the results of different feature selection methods on a dataset in a statistically acceptable manner, the Stuart method from the 'RobustRankAggreg' package in R was used. Also, in the other cases that we need to merge multiple ordered lists, we utilize the same strategy. After investigating miRs that are significantly associated with the existence of cell receptors in breast tumors, we used the ggplot2 R package to generate box plots by which we could evaluate the differential expressions of the top ten miRs with the highest association to ER between luminal (ER+) and non-luminal (ER–) breast tumor

samples in miR-seq datasets. We also reviewed the related previous studies to identify those miRs which have not been previously reported as biomarkers for BC subtypes. A group of miRs which have not been previously reported as biomarkers for BC subtypes, were selected.

2.2. Patients and samples

This research involved using archived samples that were collected from October 2018 to June 2019 and the study were conducted from October 2020 to September 2021. BC tissue specimens were collected from Khatam-ol-Anbia and Rasule-Akram hospitals, they were immediately snap-frozen in liquid nitrogen after surgery and were stored at -80°C at Tarbiat Modares University. Tissue samples were categorized into 36 pairs of breast tumors and their adjacent non-tumor tissues, plus 13 breast tumor samples. All tumor samples were examined by pathologists and classified according to the standard histopathological parameters. Clinicopathological characteristics of patients are summarized in **Supplementary Table 1**. The research protocol for in vitro experiments on tissue samples was approved by the ethics committee of Ferdowsi University of Mashhad (code number: IR.UM.REC.1399.104). No experiments were conducted on human subjects or animals. Authors have no access to the information that could identify individual participants during or after data collection.

2.3. RNA extraction

For isolating the total RNA from breast tissues, RiboEx Total RNA reagent (GeneAll Biotechnology, South Korea) was used. The concentration of the extracted RNA was then quantified using a NanoDrop™ spectrophotometer. The purity of the RNA was validated by measuring the ratio of the absorbance at 260 and 280 nm. The quality of RNA, which is regarded

as the absence of degraded RNA, was evaluated by agarose gel electrophoresis and ethidium bromide staining. Accordingly, the 18S and 28S RNA bands were visualized under ultraviolet light.

2.4. Polyadenylation and cDNA synthesis

Following RNA isolation, 1 µg of total RNA was poly-adenylated using Poly(A) Polymerase Tailing Kit (New England Biolabs., UK., Ltd.) according to manufacture protocol. Then, the poly-adenylated RNA (10 µl) was converted to complementary DNA (cDNA) by adding Anchored Oligo(dT) and using RevertAid M-MuLV RT (Thermo Fisher Scientific., UK) as the protocol provided by corresponding manufacture. The prepared cDNA was used for quantification of miR-190b, miR-584, and miR-1306.

2.5. Stem–loop RT–PCR and cDNA synthesis

Stem-loop reverse transcription was performed for miR-452 using stem-loop primers. Then, reverse transcriptase reaction was performed using RevertAid M-MuLV RT (Thermo Fisher Scientific., UK).

2.6. qRT-PCR

Ultimately, qRT-PCR was performed to quantitatively assess the expression of our selected miRs in tissue samples using appropriate primer sets (**Supplementary Tables 2 and 3)**. Syber Green PCR Master Mix (BIOFACT Co., Ltd., Korea) and primers listed in **(Supplementary Tables 2 and 3)** were utilized for qRT-PCR, which was conducted on a StepOne Plus System. Mean delta Ct values of triplicate real-time qRT-PCR amplifications were utilized in statistical analysis. The comparative delta Ct values (deltaCt (ΔCt) = Ct_{miRs} - $Ct_{U\text{-}48}$) and Log $2^{-\Delta Ct}$ were used as the

relative quantification of miRNAs, using the U48 small RNA [27-29] for normalization. All the amplified candidate miRs which TA-cloned using pGEM-T easy vector kit (Promega; USA). Finally the accuracy of cloned sequences were confirmed by the golden standard sequencing method.

2.7 Statistical analysis

GraphPad Prism ver. 8 (GraphPad Software Inc., La Jolla, CA, USA) was applied for statistical analysis of the results obtained by qRT-PCR. The two-tailed Mann-Whitney test was used to compare the differential expression level of selected miRs within the ER+ and ER– breast tumor samples.

The Wilcoxon test was used for comparison of our selected miRs expression level between breast tumor and their adjacent non-tumor tissues. Results with p-value < 0.05 were considered as significant.

2.8 *ROC Curve analysis*

Receiver operating characteristic (ROC) curves were also plotted by the pROC package [30] to validate the capability of the candidate miRs to distinguish between ER+ and ER– tumor samples and between breast tumors and their adjacent non-tumor samples. This was performed both individually and for a combination of all selected miRs. The area under curve (AUC) was also employed for evaluation of the specificity and sensitivity of the candidate miRs in distinguishing ER+ and ER– tumor samples and their adjacent non-tumor tissues; the higher AUC shows better diagnostic performance (the AUCs closer to 1 reflect more substantial differences).

3 Results

3.1. The candidate miRs associated with ER in breast tumors

The miRs with the highest possibility of being associated with ER were identified and ordered using limma, xgboost, bruta and aggregated by the RobustRankAggreg package. Boruta is an R package utilizing a random forest model to classify data. XGBoost is also an R package that provides a regression and classification model based on tree models and the ensemble technique. Further, we combined all the results to reach the most generalizable miR biomarkers. We suppose that several miRs are significantly associated with the existence of cell receptors; therefore, this association should be found in most of the datasets and by the majority of methods without considering methods and experiment biases. The final result of a feature selection method on a dataset is an ordered list of miRs. As we aimed to investigate potential biomarkers of luminal (ER+) and non-luminal (ER–) breast tumors, we focused our study on the top ten most strongly ER associated miRs **(Table 1)**. The expression status of these top ten miRNAs in BC were investigated in UCSC genome browser (GRCH37/hg19), and their differential expression were investigated in ER+ and ER– breast tumor samples of TCGA **(Supplementary Fig. 1)**. We excluded miR-577 and miR-452-3p from the top ten list due to their extremely low expression levels in breast tumors. The first rank miR-190b with the highest score in the list was selected as the potential biomarker of ER+ breast tumors. The expression levels and the association of miR-18a-5p [31], miR-505-3p [32], miR-224-5p [33, 34], and miR-135b-5p [35] with ER have been previously investigated in breast cancer. Therefore, we excluded miR-18a-5p, miR-505-3p, miR-224-5p, and miR-135b-5p from this study, and we selected those miRs which their association with highly aggressive breast tumors (ER–) have not been previously studied. In this regard, miR-584-5p, miR-452-5p, and miR-1306-5p were selected to evaluate their potential as biomarkers of ER– breast tumors.

Table 1. Top-ranked most strongly ER associated miRNAs.

ER	Aggregate Score
miR-190b	0.000032
miR-18a-5p	0.000256
miR-505-3p	0.002304
miR-224-5p	0.00484
miR-577	0.005816
miR-135b-5p	0.007
miR-584-5p	0.01024
miR-452-5p	0.065408
miR-452-3p	0.075656
miR-1306-5p	0.114264

The results of the RobustRankAggreg package analysis of miRseq datasets are summarized in this table. MiRNAs with a lower aggregate score are associated to ER receptors with a higher probability. Our candidate miRNAs are in bold style.

3.2. Evaluation of miR-190b expression level in ER+ compared to ER– breast tumors

Our *in-silico* analysis on the miR-seq datasets in TCGA **(Fig. 1A)** and GSE68085 **(Fig. 1B)** databases, elucidated higher expression level of miR-190b in ER+ compared to ER– breast tumors (P. Value < 0.05). Moreover, as shown in **Fig. 1C**, the expression level of miR-190b correlated positively with the percentage of ER level in breast tumors, which indicates that the higher the ER level is, the more miR-190b is expressed in breast tumors. The results of qRT-PCR confirmed our *in-silico* analysis **(Fig. 1D)**. Indeed, we found a similar significant higher expression level of miR-190b in 24 ER+ compared to 23 ER– breast tumor samples (P. Value < 0.05).

Fig. 1: The expression of miR-190b in patients with ER+ compared to ER– BCs. The *in-silico* analysis of the miR-seq data of BC patients in TCGA **(A)** and GSE68085 **(B)** showed a significant upregulation of miR-190b in ER+ versus ER– breast tumor samples (P. Value < 0.05). **C)** The *in-silico* evaluation of miR-190b expression based on ER percentage level showed a positive correlation between miR-190b and ER expression level. **D)** The results of the qRT-PCR analysis also showed that the expression of miR-190b (relative to u-48) in ER+ breast samples is significantly higher than ER– breast samples (P. Value < 0.05).

3.3. Evaluation of the expression of miR-584, miR-452, and miR-1306 in ER+ compared to ER– breast tumors

We validated the significant downregulation of miR-584, miR-452, and miR-1306 in ER+ compared to ER– breast tumors by qRT-PCR **(Fig. 2D)**, which was predicted by analyzing the available miR-seq datasets in TCGA **(Fig. 2A)** and GSE68085 **(Fig. 2B)** (P. Value < 0.05).

In *in-silico* analysis, a negative correlation between the percentage of ER level and the expression of miR-584, miR-452, and miR-1306 was also observed in breast tumors **(Fig. 2C)**.

Fig. 2: The expression of miR-584, miR-452, and miR-1306 in patients with ER+ compared to ER– BCs. The results of *in-silico* analysis of the miR-seq dataset of BC patients in TCGA **(A)** and GSE68085 **(B)** showed downregulation of miR-584, miR-452, and miR-1306 (P. Value < 0.05) in ER+ versus ER– breast tumor samples. **C)** The *in-silico* evaluation of the expression of miR-584, miR-452, and miR-1306 based on the percentage of ER level indicated a significant negative correlation between these three miRs and ER expression level (P. Value < 0.05). **D)** The qRT-PCR analysis also showed that the expression of miR-584, miR-452, and miR-1306 (relative to u-48) in ER+ breast samples is significantly lower than ER– breast samples (P. Value < 0.05).

3.4. Evaluation of Differential expression of our four selected miRs between breast tumor and non-tumor samples

At first, expression profiles of the selected miRs between breast tumor and non-tumor samples (adjacent non-tumor breast tissues) in TCGA datasets were determined by *in-silico* analysis. Then, the expression of the candidate miRs were evaluated experimentally in 36 breast tumor cases and their adjacent non-tumor tissues by qRT-PCR analysis. Although *in-silico* evaluation of miR-190b expression in breast tumors compared to non-tumor breast samples of TCGA data showed a significant upregulation **(Fig. 3A)**, the qRT-PCR analysis showed an opposite results **(Fig. 3B)**. Further analysis indicated that the downregulation of miR-190b in breast tumors was restricted to ER−/PR− tumors while ER+/PR+ and ER+/PR− tumors showed the upregulation of miR-190b compared to non-tumor breast tissues **(Fig. 3C and D)**. Therefore, the upregulation or downregulation of miR-190b in tumor samples versus non-tumor samples is dependent on the ER status.

Fig. 3: The expression of miR-190b (relative to U-48) in breast tumor samples compared to non-tumor adjacent tissues. A) The results of *in-silico* analysis showed a significant increase of miR-190b expression in breast tumors compared to non-tumor breast samples of TCGA data (P. Value < 0.05). **B)** However, the results of the qRT-PCR analysis showed an overall decrease of miR-190b expression in tumor tissues versus non-tumor adjacent tissues (P. Value < 0.05). The qRT-PCR analysis also showed that miR-190b was up-regulated in ER+ tumors **(C)**, whereas it was downregulated in ER– tumors compared to non-tumor adjacent tissues **(D)**.

MiR-584-5p, miR-452-5p, miR-1306 showed lower expression in breast tumor samples (both ER+ and ER–) compared to non-tumor breast tissues based on both miR-seq **(Fig. 4A, B, and C)** and qRT-PCR analysis **(Fig 4D, E, and F)**. Although, in *in-silico* analysis of miR-seq datasets, miR-1306 didn't show significant differential expression between breast tumor and non-tumor

adjacent samples **(Fig. 4C)**, in the qRT-PCR analysis, it was significantly down-regulated in breast tumor tissues compared to their adjacent non-tumor tissues (p-value < 0.05) **(Fig. 4F)**.

Fig. 4: The expression of miR-584, miR-452, and miR-1306 (relative to U-48) in breast tumor samples compared to non-tumor adjacent tissues. The results of in-silico analysis showed that the expression of miR-584 **(A)**, and miR-452 **(B)** are significantly lower in breast tumors, as compared with non-tumor breast samples of TCGA (P. Value < 0.05), while the expression level of miR-1306 **(C)** showed no significant difference between breast tumors and non-tumor breast samples. The results of the qRT-PCR analysis showed significant downregulation of miR-584 **(D)**, miR-452 **(E)**, and miR-1306 **(F)** in tumor tissues versus non-tumor adjacent tissues (P. Value < 0.05).

3.5. Roc Curve analysis

The four candidate miRs, miR-190b, miR-584, miR-452 and miR-1306, showed high AUCs with values of 0.951 (specificity of 86%, and sensitivity of 94%), 0.851 (specificity of 82%, and sensitivity of 75%), 0.846 (specificity of 79%, and sensitivity of 81%), and 0.80 (specificity of 72%, and sensitivity of 78%), respectively **(Fig. 5A, B, C, and D)**. Interestingly, the AUC of the combination of miR-190b, miR-584, miR-452, and miR-1306 increased to 0.973 with the specificity of 92% and sensitivity of 96% in discriminating ER+ and ER– samples of TCGA dataset **(Fig. 5E)**. A ROC curve was also plotted for the combination of miR-190b, miR-18a-5p, miR-505-3p, miR-224-5p, miR-135b-5p, miR-584-5p, miR-452-5p, and miR-1306-5p which created the similar AUC value of 0.977 **(Supplementary Fig. 2)**. Therefore, the addition of miR-18a-5p, miR-505-3p, miR-224-5p, and miR-135b-5p did not add values to the AUC achieved by the combination of our selected miRs (Both groups represented the AUC of 0.97). In addition to the ROC curves obtained by the in silico analysis of TCGA datasets, the pROC package was used to evaluate the diagnostic value of our selected miRs in discriminating ER+ from ER–

tissue samples used in the qRT-PCR analysis. In concordance with AUCs obtained for miR levels in TCGA datasets, significant AUCs of 0.961 (Specificity of 71%, and sensitivity of 63%), 0.674 (Specificity of 69%, and sensitivity of 61%), 0.777 (Specificity of 87%, and sensitivity of 69%), and 0.666 (Specificity of 77%, and sensitivity of 52%) have been obtained for miR-190b **(Fig. 5F)**, miR-584 **(Fig. 5G)**, miR-452 **(Fig. 5H)**, and miR-1306 **(Fig. 5I)** respectively (P. Value < 0.05).

Fig. 5: The AUCs of ROC curves for miR levels in ER+ and ER– samples. The capabilities of miR-190b **(A** and **F)**, miR-584 **(B** and **G)**, miR-452 **(C** and **H)**, miR-1306 **(D** and **I)**, and the combination of these four miRs **(E)** to discriminate ER+ and ER– samples of TCGA datasets **(A**, **B**, **C**, **D**, and **E)** and qRT-PCR analysis **(F**, **G**, **H**, and **I)** are shown in ROC curves.

As shown in the qRT-PCR results, miR-190b was up-regulated in ER+ breast tumors, whereas it was down-regulated in ER– tumors compared to non-tumor tissues. Therefore, we excluded this miR from the group of biomarkers which efficiency was evaluated in the detection of breast tumors. Accordingly, ROC curves were plotted for miR-584, miR-452, and miR-1306 expression in TCGA datasets, which represented the AUCs of 0.961 (specificity of 90%, and sensitivity of 91%), 0.926 (specificity of 93%, and sensitivity of 82%), and 0.725 (specificity of 63%, and sensitivity of 74%) (**Fig. 6A, B,** and **C**). The AUCs of 0.939 (specificity of 97%, and sensitivity of 86%), and 0.982 (specificity of 91%, and sensitivity of 95%) were also generated for the combination of miR-584 and miR-452, and the combination of miR-584, miR-452, and miR-1306 (**Fig. 6D** and **E**).

Furthermore, the ROC curves generated for the expression of miR-584, miR-452, and miR-1306 in tumor and adjacent non-tumor samples of qRT-PCR analysis represented the AUC values of

274 0.766 (specificity of 64%, and sensitivity of 78%), 0.666 (specificity of 68%, and sensitivity of
275 60%), and 0.705 (specificity of 61%, and sensitivity of 75%) **(Fig. 6F, G, and H)**.
276
277 **Fig. 6: The AUCs of ROC curves for mi levels in tumor and non-tumor samples.** The potential of miR-584 **(A**
278 and **F)**, miR-452 **(B** and **G)**, miR-1306 **(C** and **H)**, the combination of miR-584 and miR-452 **(D)**, and the
279 combination of all these three miRs **(E)** to discriminate tumor and non-tumor samples of TCGA **(A, B, C, D, E)** and
280 qRT-PCR analysis **(F, G, H)** are shown in ROC curves.
281
282 ## 4 Discussion
283 In the current study, we aimed to identify ER associated miRs to reach promising biomarkers.
284 Therefore, we used the available data from TCGA and applied boruta and XGBoost feature
285 selection and limma differentially expression analysis to obtain miRs that are significantly
286 associated with the existence of cell receptors in breast tumors. As a result, we reported the
287 discovery of four miRNAs, miR-190b, miR-584-5p, miR-452-5p, and miR-1306-5p, which are
288 significantly associated with ER status in breast tumors, as predicted by *in silico* analysis. We
289 then confirmed the efficacy of this signature in discriminating luminal BCs (ER+) and non-
290 luminal BCs (ER–) by qRT-PCR.
291 Currently, Mammography is the golden standard tool for screening and diagnosis of breast
292 cancer (BC). However, invasive histological evaluation of breast biopsy is required for accurate
293 diagnosis of the BC subtype. The identification of novel, reliable, and minimally invasive BC
294 biomarkers with high sensitivity and specificity would lead to a significant improvement in the
295 clinical management of this complex disease [36]. Numerous studies have reported the
296 significant role of miRs in the initiation and progression of BC and revealed that certain miRs are
297 differentially expressed between different breast tumor subtypes. [37]. Hence, the differentially

298 expressed miRs can be considered as a promising molecular biomarkers for distinguishing BC
299 subtypes.

300 It has been reported that miR-190b is associated with ER+ breast tumors [38] and resistance to
301 hormone therapy [39]. Furthermore, Cizeron-Clairac et al. reported that miR-190b is
302 significantly upregulated in ER+ compared to ER– tumors. In consistent with previous studies,
303 our results also demonstrated the significant higher expression level of miR-190b in ER+
304 compared to ER– breast tumors (P. Value < 0.05). Moreover, ROC curve analysis showed the
305 potential diagnostic value of miR-190b as a very remarkable biomarker in distinguishing ER+
306 from ER– breast tumors with the AUC of 0.951 (based on *in-silico* analysis). In another study,
307 the expression of miR-190b was examined in seven BC cell lines and suggested that the
308 biological and clinical implication of miR-190b may differ among BC subtypes [40].
309 Furthermore, de Anda-Jáuregui et al. showed the upregulation of miR-190b in ER+ breast
310 tumors compared to normal breast tissues, whereas the downregulation of miR-190b in ER–
311 tumors compared to healthy controls [38]. In the present study, our *in-silico* analysis represented
312 an overall upregulation of miR-190b in breast tumors versus adjacent non-tumor tissue. While in
313 the practical experiment, we observed the upregulation of miR-190b in ER+ but not in ER–
314 breast tumors compared to non-tumor tissues. Hence, it should be noted that the upregulation of
315 miR-190b in breast tumor samples versus non-tumor adjacent tissues was restricted to ER+
316 tumors. This discrepancy between the results of *in-silico* and qRT-PCR analysis may be
317 attributed to the higher number of ER+ tumors compared to ER– tumors in the TCGA dataset.

318 Previous studies mostly have focused on the role of miR-584 in lung cancer progression and
319 metastasis [41-43]. The chromosomal region where miR-584 is located, 5q32, has been
320 highlighted to be deleted in myelodysplastic syndromes that lead to malignant transformation

[44, 45]. In addition, a study indicated that miR-584 may act as a tumor suppressor in renal carcinoma cells [46]. Furthermore, it was revealed that miR-584 is significantly down-regulated in human HER2+ breast tumors compared to non-tumor adjacent tissues [47]. Our results also showed significant downregulation of miR-584 in breast tumors compared to adjacent non-tumor tissues, which may indicate the tumor suppressive role of miR-584. According to ROC curve analysis, miR-584 also can act as a strong biomarker for BC diagnosis with high sensitivity and specificity (AUC of 0.961 and 0.766 in *in-silico* and qRT-PCR analysis, respectively). Additionally, our findings validated the downregulation of miR-584 in ER+ versus ER– breast tumors predicted by *in silico* analysis, and the AUC values of 0.851 and 0.674 obtained by *in-silico* and qRT-PCR analysis that represented the notable capability of miR-584 in distinguishing ER– and ER+ BC subtypes.

It is reported that miR-452 is expressed aberrantly in different types of human cancer [48-50]. The downregulation of miR-452 in BC tissues compared with paired normal breast tissues was also identified previously [51]. Moreover, miR-452 has been predicted to have an important role in regulation of pathways specific to luminal-A by TCGA data analysis [52]. In the present study, we also demonstrated the downregulation of miR-452 in breast tumors versus adjacent non-tumor tissues. More importantly, for the first time to our knowledge, we showed that miR-452 could serve as a potential biomarker for distinguishing ER– from ER+ breast tumors with high sensitivity and specificity (AUC values of 0.846 and 0.777 by *in-silico* and qRT-PCR analysis, respectively).

The aberrant expression of miR-1306 was observed in plasma samples of patients with different diseases such as heart failure, glaucoma, or epilepsy [53]. Moreover, the dysregulation of miR-1306 has been reported in human breast and colorectal cancers [53, 54]. Likewise, the

upregulation of miR-1306-5p was observed in subjects with malignant breast lesions compared to benign tumors [53]. Our study represented the upregulation of miR-1306 in ER– breast tumors which typically are more aggressive than ER+ tumors. However, the downregulation of miR-1306 in breast tumors compared to non-tumor adjacent tissues may suggest its tumor suppressive role. This discrepancy needs to be elucidated in future studies. Likewise, no previous study has been performed to investigate the expression level and biological roles of miR-1306 in BC subtypes. Hence, this is the first study to our knowledge, that reveals the efficacy of miR-1306-5p in distinguishing ER+ and ER– breast tumors with AUC values of 0.799 and 0.666 based on *in-silico* and qRT-PCR analysis, respectively.

Overall, the AUCs acquired by qRT-PCR were not as high as those of *in-silico* analysis. This discrepancy can be attributed to the much higher number of breast tumor samples of TCGA datasets compared to the samples used in qRT-PCR. For identifying the best biomarker panel, we compared the AUCs produced from ROC curve analysis for each individual miR and their combination profiles in distinguishing ER+ and ER– BC samples of TCGA. It was revealed that the best AUC of 0.973 can be achieved from the combination of miR-190b, miR-584, miR-452, and miR-1306, providing specificity of 92% and sensitivity of 96%.

The ROC curve for the combination of the miR-584-5p, miR-452, and miR-1306 in breast tumors and non-tumor samples of TCGA showed an exceptionally high diagnostic accuracy with an AUC value of 0.982 with the specificity of 91% and sensitivity of 95%. It should be noted that miR-190b was not included in this biomarker set because its differential expression in breast tumors compared to non-tumor tissues depends on the tumor subtype. Although the signature of miR-584, miR-452, and miR-1306 is an effective diagnostic test alone with a strong performance

(specificity of 100% and sensitivity of 87%), it may show promise as a valuable assessment tool in BC diagnosis in combination with screening mammography.

A particular strength of this study is that the reported mi panel is probably less prone to biological differences than a single miR. Therefore, the miR panels are more reliable for clinical use. A second strength is that the measured expression level of the four candidate miRs in breast tumor samples revealed a specific algorithm that helps to detect the ER status of each sample. To clarify, in ER+ tumors, the expression level of miR-190b was higher than that of miR-584, miR-452, and miR-1306, while, in ER– tumors the results were vice versa **(Supplementary Fig. 3)**. However, since the miR-452 expression level was not detectable in a number of tumor samples, the role of miR-452 in this algorithm needs further investigations.

5 Conclusion

Overall, our study signified the discovery of a panel of four miRs (miR-190b, miR-584-5p, miR-452-5p, and miR-1306-5p), which expression have association with ER in breast tumors, for diagnosis of luminal (ER+) and non-luminal (ER–) BC subtypes. According to this miR panel, we have represented a classification model with high discrimination ability in classifying ER+ and ER– BCs at an AUC of 0.973. Interestingly, a straightforward classification was introduced based on the seesaw expression pattern of miR-190b compared to the other three miRs to distinguish ER– and ER+ samples. In addition, the panel containing miR-584, miR-452, and miR-1306 was shown to have the efficacy to be developed as a parallel test to examine breast samples of patients with abnormal screening mammograms, with the aim of reducing false-positive results. For future investigations, these two miR signatures should be evaluated in blood serum samples to confirm the clinical utility of these miR signatures.

Acknowledgments

Not applicable.

Author contributions

Conception: Faranak Farahmand

Interpretation or analysis of data: Faranak Farahmand, Saeid Rahmani, Hadi Bayat

Preparation of the manuscript: Faranak Farahmand, Saeid Rahmani

Revision for important intellectual content: Faranak Farahmand, Saeid Rahmani, Seyed-Javad Mowla, Hadi Bayat, Adel Salimi, Sogol Ghanbari, Ali Sharifi-Zarchi, Mohammad Vasei

Supervision: Seyed-Javad Mowla

Uncategorized References

1. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: a cancer journal for clinicians, 2021. **71**(3): p. 209-249.
2. Coleman, W.B. and C.K. Anders, *Discerning clinical responses in breast Cancer based on molecular signatures.* The American journal of pathology, 2017. **187**(10): p. 2199-2207.
3. Li, C., D. Uribe, and J. Daling, *Clinical characteristics of different histologic types of breast cancer.* British journal of cancer, 2005. **93**(9): p. 1046-1052.
4. Hu, Z., et al., *The molecular portraits of breast tumors are conserved across microarray platforms.* BMC genomics, 2006. **7**(1): p. 1-12.
5. Chen, Y.-J., et al., *Molecular subtyping of breast cancer intrinsic taxonomy with oligonucleotide microarray and NanoString nCounter.* Bioscience reports, 2021. **41**(8).
6. Perou, C.M., et al., *Molecular portraits of human breast tumours.* nature, 2000. **406**(6797): p. 747-752.
7. Song, N., et al., *Prediction of breast cancer survival using clinical and genetic markers by tumor subtypes.* PLoS One, 2015. **10**(4): p. e0122413.
8. Joshi, H. and M.F. Press, *Molecular oncology of breast cancer*, in *The Breast*. 2018, Elsevier. p. 282-307. e5.
9. Hironaka-Mitsuhashi, A., et al., *A tissue microRNA signature that predicts the prognosis of breast cancer in young women.* PLoS One, 2017. **12**(11): p. e0187638.
10. Jorns, J.M., *Breast cancer biomarkers: challenges in routine estrogen receptor, progesterone receptor, and HER2/neu evaluation.* Archives of Pathology & Laboratory Medicine, 2019. **143**(12): p. 1444-1449.
11. Alba-Bernal, A., et al., *Challenges and achievements of liquid biopsy technologies employed in early breast cancer.* EBioMedicine, 2020. **62**: p. 103100.
12. Adhami, M., et al., *Candidate miRNAs in human breast cancer biomarkers: a systematic review.* Breast Cancer, 2018. **25**(2): p. 198-205.
13. Stahlhut, C. and F.J. Slack, *MicroRNAs and the cancer phenotype: profiling, signatures and clinical implications.* Genome medicine, 2013. **5**(12): p. 1-12.
14. Mu, H., et al., *miRNAs as potential markers for breast cancer and regulators of tumorigenesis and progression.* International Journal of Oncology, 2021. **58**(5): p. 1-12.
15. Arun, R.P., H.F. Cahill, and P. Marcato, *Breast Cancer Subtype-Specific miRNAs: Networks, Impacts, and the Potential for Intervention.* Biomedicines, 2022. **10**(3): p. 651.
16. Gupta, I., et al., *Circulating miRNAs in HER2-positive and triple negative breast cancers: potential biomarkers and therapeutic targets.* International Journal of Molecular Sciences, 2020. **21**(18): p. 6750.
17. Zhang, Y., et al., *Multilayer network analysis of miRNA and protein expression profiles in breast cancer patients.* PloS one, 2019. **14**(4): p. e0202311.
18. Schrauder, M.G., et al., *Circulating micro-RNAs as potential blood-based markers for early stage breast cancer detection.* PloS one, 2012. **7**(1): p. e29770.
19. Kudela, E., et al., *miRNA expression profiles in luminal A breast cancer—implications in biology, prognosis, and prediction of response to hormonal treatment.* International Journal of Molecular Sciences, 2020. **21**(20): p. 7691.

20. Dvinge, H., et al., *The shaping and functional consequences of the microRNA landscape in breast cancer*. Nature, 2013. **497**(7449): p. 378-382.
21. Kursa, M.B. and W.R. Rudnicki, *Feature selection with the Boruta package*. J Stat Softw, 2010. **36**(11): p. 1-13.
22. Chen, T., et al., *xgboost: Extreme Gradient Boosting (2017)*. R package version 0.6-4, 2015.
23. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic acids research, 2015. **43**(7): p. e47-e47.
24. Chang, K., et al., *The cancer genome atlas pan-cancer analysis project.* Nat Genet, 2013. **45**(10): p. 1113-1120.
25. Colaprico, A., et al., *TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data*. Nucleic acids research, 2016. **44**(8): p. e71-e71.
26. Krishnan, P., et al., *Next generation sequencing profiling identifies miR-574-3p and miR-660-5p as potential novel prognostic markers for breast cancer.* BMC genomics, 2015. **16**(1): p. 1-17.
27. Xiang, M., et al., *U6 is not a suitable endogenous control for the quantification of circulating microRNAs.* Biochemical and biophysical research communications, 2014. **454**(1): p. 210-214.
28. Hirschberger, S., et al., *Identification of suitable controls for miRNA quantification in T-cells and whole blood cells in sepsis.* Scientific reports, 2019. **9**(1): p. 1-13.
29. Carvalho, T.M., et al., *MicroRNAs miR-142-5p, miR-150-5p, miR-320a-3p, and miR-4433b-5p in Serum and Tissue: Potential Biomarkers in Sporadic Breast Cancer*. Frontiers in Genetics, 2022: p. 1414.
30. Robin, X., et al., *pROC: an open-source package for R and S+ to analyze and compare ROC curves.* BMC bioinformatics, 2011. **12**(1): p. 1-8.
31. Calvano Filho, C.M.C., et al., *Triple-negative and luminal A breast tumors: differential expression of miR-18a-5p, miR-17-5p, and miR-20a-5p.* Tumor Biology, 2014. **35**(8): p. 7733-7741.
32. Jonsdottir, K., et al., *Validation of expression patterns for nine miRNAs in 204 lymph-node negative breast cancers.* PloS one, 2012. **7**(11): p. e48692.
33. Zhang, L., et al., *MicroRNA-224 promotes tumorigenesis through downregulation of caspase-9 in triple-negative breast cancer.* Disease markers, 2019. **2019**.
34. Huang, L., et al., *MicroRNA-224 targets RKIP to control cell invasion and expression of metastasis genes in human breast cancer cells.* Biochemical and biophysical research communications, 2012. **425**(2): p. 127-133.
35. Aakula, A., et al., *MicroRNA-135b regulates ERα, AR and HIF1AN and affects breast and prostate cancer cell growth.* Molecular oncology, 2015. **9**(7): p. 1287-1300.
36. Fisher, B., C.K. Redmond, and E.R. Fisher, *Evolution of knowledge related to breast cancer heterogeneity: a 25-year retrospective*. Journal of Clinical Oncology, 2008. **26**(13): p. 2068-2071.
37. Blenkiron, C., et al., *MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype.* Genome biology, 2007. **8**(10): p. 1-16.
38. Cizeron-Clairac, G., et al., *MiR-190b, the highest up-regulated miRNA in ERα-positive compared to ERα-negative breast tumors, a new biomarker in breast cancers?* BMC cancer, 2015. **15**(1): p. 1-14.

39. Joshi, T., et al., *Integrative analysis of miRNA and gene expression reveals regulatory networks in tamoxifen-resistant breast cancer.* Oncotarget, 2016. **7**(35): p. 57239.
40. Dai, W., et al., *miR-148b-3p, miR-190b, and miR-429 regulate cell progression and act as potential biomarkers for breast cancer.* Journal of breast cancer, 2019. **22**(2): p. 219-236.
41. Ma, D., et al., *Circular RNA ABCB10 promotes non-small cell lung cancer progression by increasing E2F5 expression through sponging miR-584-5p.* Cell Cycle, 2020. **19**(13): p. 1611-1620.
42. Lee, S.B., et al., *Tumor suppressor miR-584-5p inhibits migration and invasion in smoking related non-small cell lung cancer cells by targeting YKT6.* Cancers, 2021. **13**(5): p. 1159.
43. Guo, T., et al., *miR□584□5p regulates migration and invasion in non□small cell lung cancer cell lines through regulation of MMP□14.* Molecular Medicine Reports, 2019. **19**(3): p. 1747-1752.
44. Eisenmann, K., et al., *5q–myelodysplastic syndromes: chromosome 5q genes direct a tumor-suppression network sensing actin dynamics.* Oncogene, 2009. **28**(39): p. 3429-3441.
45. Lehmann, S., et al., *Common deleted genes in the 5q− syndrome: thrombocytopenia and reduced erythroid colony formation in SPARC null mice.* Leukemia, 2007. **21**(9): p. 1931-1936.
46. Ueno, K., et al., *Tumour suppressor microRNA-584 directly targets oncogene Rock-1 and decreases invasion ability in human clear cell renal cell carcinoma.* British journal of cancer, 2011. **104**(2): p. 308-315.
47. Cava, C., et al., *Theranostic application of miR-429 in HER2+ breast cancer.* Theranostics, 2020. **10**(1): p. 50.
48. He, Z., et al., *Up-regulation of MiR-452 inhibits metastasis of non-small cell lung cancer by regulating BMI1.* Cellular Physiology and Biochemistry, 2015. **37**(1): p. 387-398.
49. Kristensen, H., et al., *Hypermethylation of the GABRE∼miR-452∼miR-224 promoter in prostate cancer predicts biochemical recurrence after radical prostatectomy.* Clinical Cancer Research, 2014. **20**(8): p. 2169-2181.
50. Liu, L., et al., *Downregulation of miR-452 promotes stem-like traits and tumorigenicity of gliomas.* Clinical Cancer Research, 2013. **19**(13): p. 3429-3438.
51. Li, W., et al., *Tumor-suppressive microRNA-452 inhibits migration and invasion of breast cancer cells by directly targeting RAB11A.* Oncology Letters, 2017. **14**(2): p. 2559-2565.
52. Cava, C., et al., *How interacting pathways are regulated by miRNAs in breast cancer subtypes.* BMC bioinformatics, 2016. **17**(12): p. 111-133.
53. Loke, S.Y., et al., *A circulating miRNA signature for stratification of breast lesions among women with abnormal screening mammograms.* Cancers, 2019. **11**(12): p. 1872.
54. Chang, N., et al., *Long Noncoding RNA LINC00857 Promotes Proliferation, Migration, and Invasion of Colorectal Cancer Cell through miR-1306/Vimentin Axis.* Computational and Mathematical Methods in Medicine, 2021. **2021**.

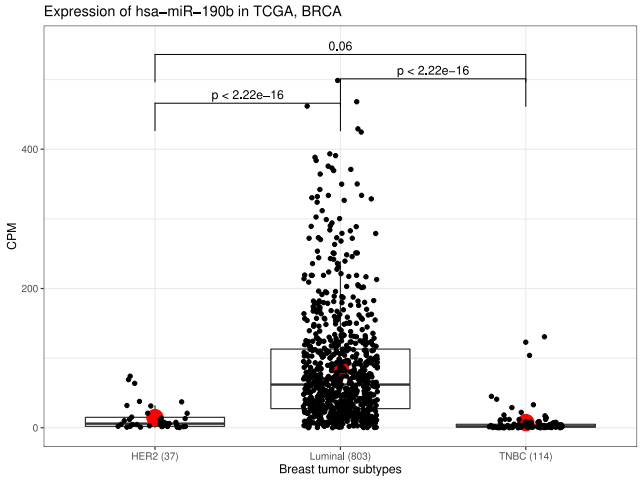
Supporting information captions:

Supplementary Fig. 1: The differential expression of miR-190b **(A)**, miR-18a-5p **(B)**, miR-505-3p **(C)**, miR-224-5p **(D)**, miR-577 **(E)**, miR-135b-5p **(F)**, miR-584-5p **(G)**, miR-452-5p **(H)**, miR-452-3p **(I)**, and miR-1306-5p **(J)** in ER+ versus ER– breast tumor samples of TCGA.

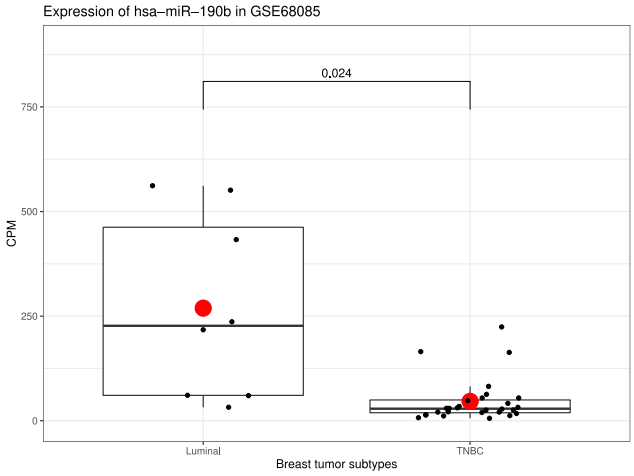
Supplementary Fig. 2: A ROC curve for the combination of miR-190b, miR-18a-5p, miR-505-3p, miR-224-5p, miR-135b-5p, miR-584-5p, miR-452-5p, and miR-1306-5p in ER+ and ER– breast tumor samples of TCGA which created the AUC value of 0.977.

Supplementary Fig. 3: The expression of miR-190b, miR-584-5p, miR-452-5p, and miR-1306-5p (relative to u-48) in ER–/PR– and ER+/PR+ breast tumor samples. The results of the qRT-PCR analysis revealed that the measured expression level of miR-190b is lower in each 9 ER–/PR–samples **(A)** and is higher in each 8 ER+/PR+ samples **(B)**, in comparison with the expression levels of the other three miRs in those samples (The expression of miR-452-5p was not detectable in T1, T2, T4, T8, T11, T12, T13, T14, T15, T16, and T17 samples).

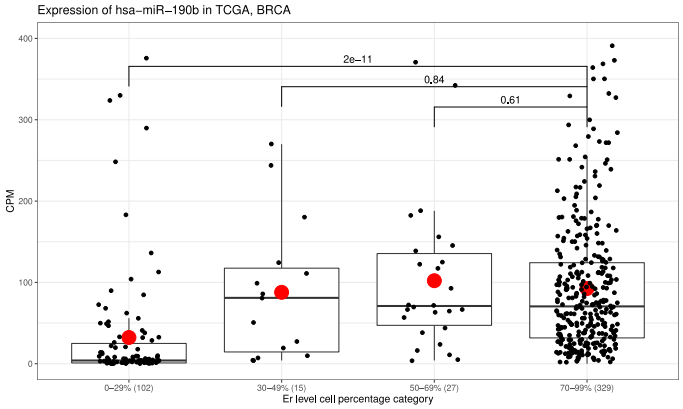
A)



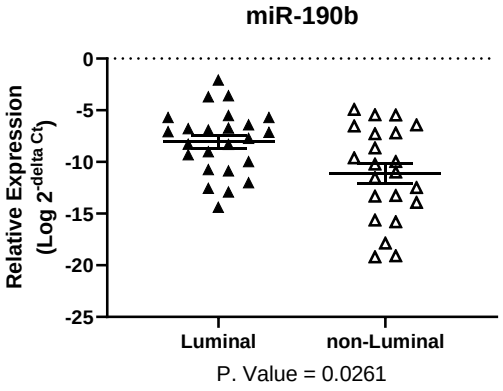
B)



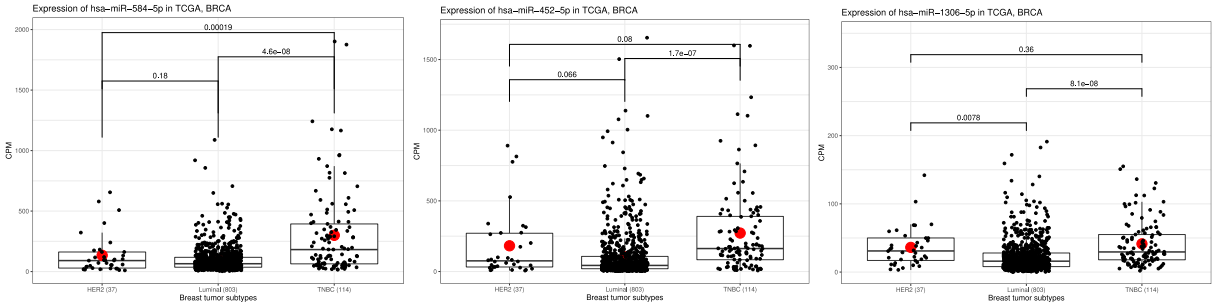
C)



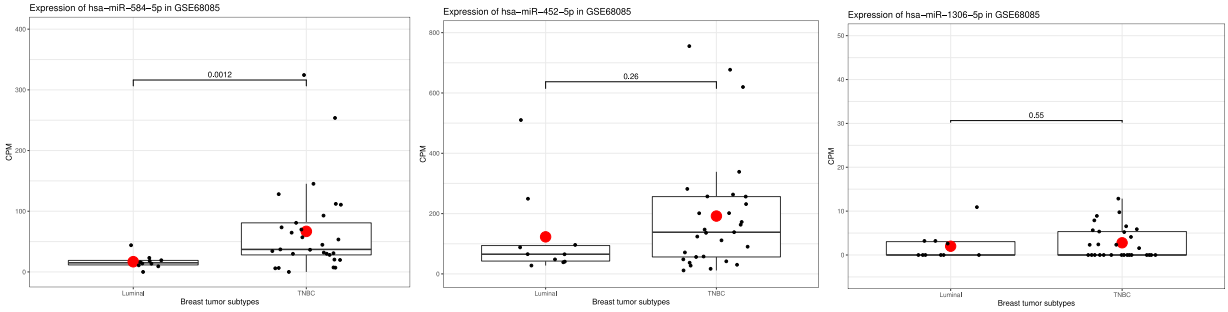
D)



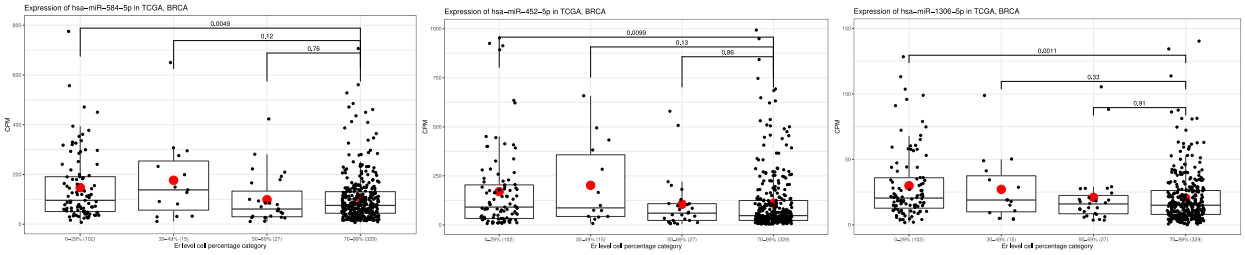
A)



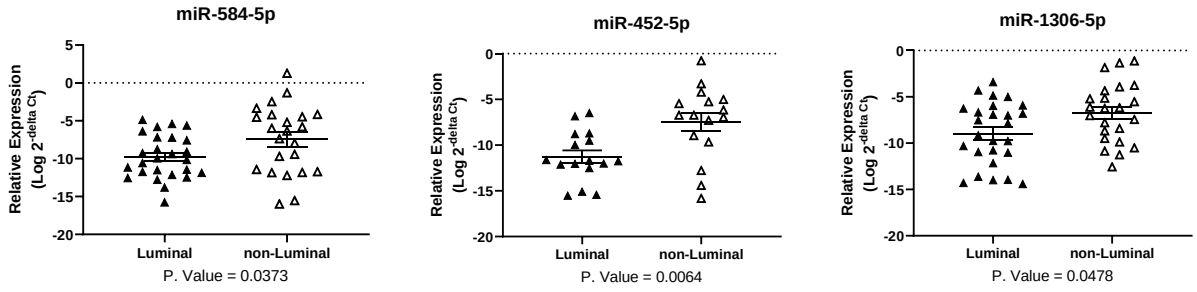
B)



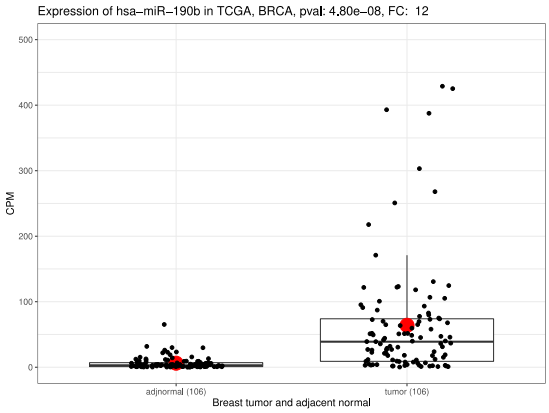
C)



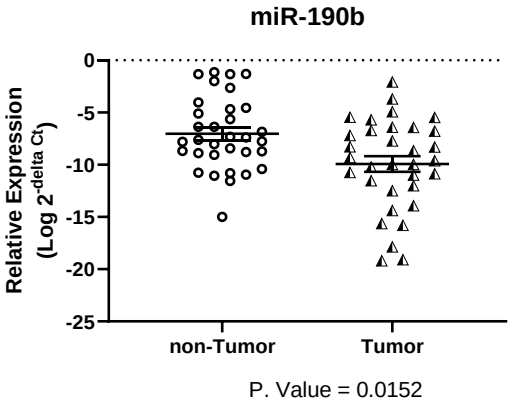
D)



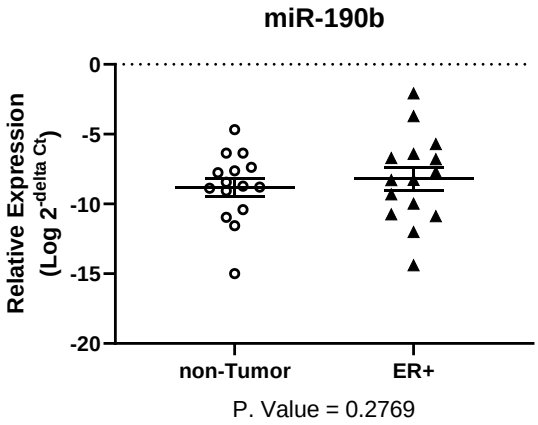
A)



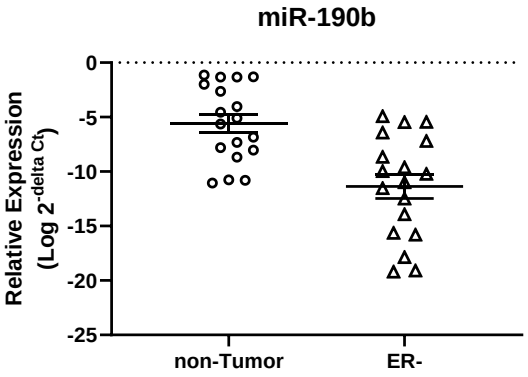
B)



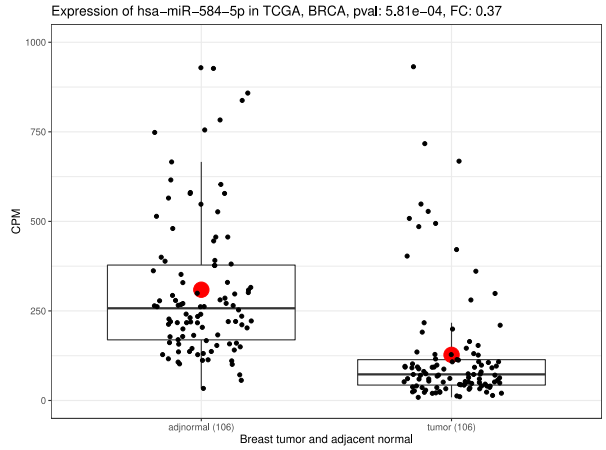
C)



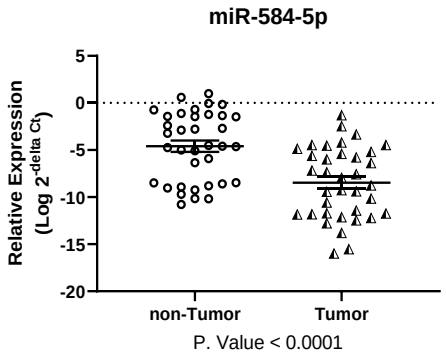
D)



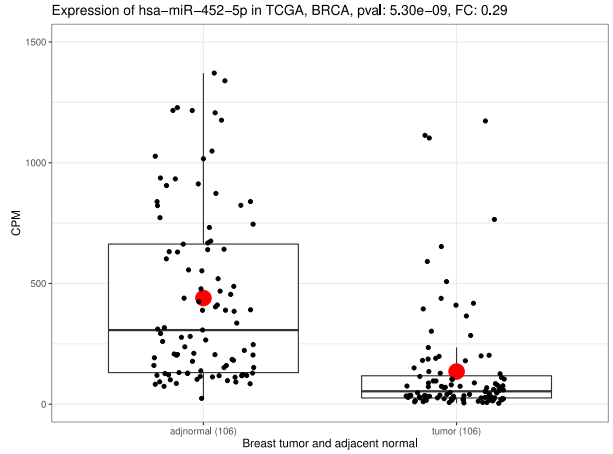
A)



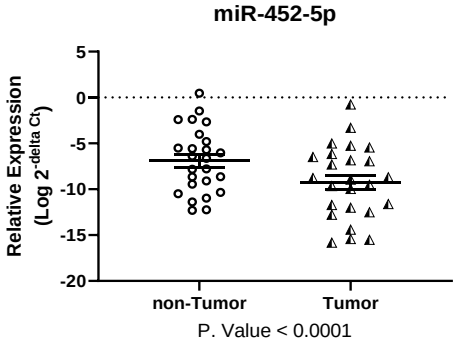
D)



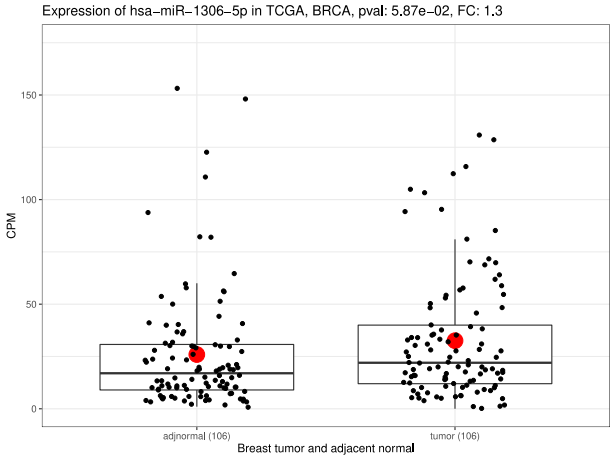
B)



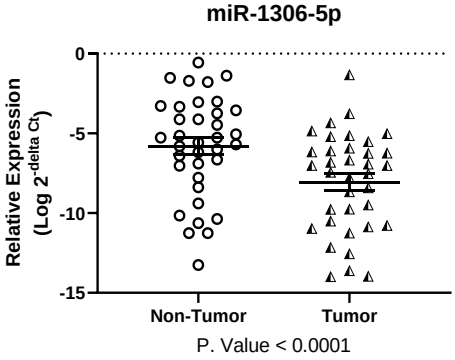
E)



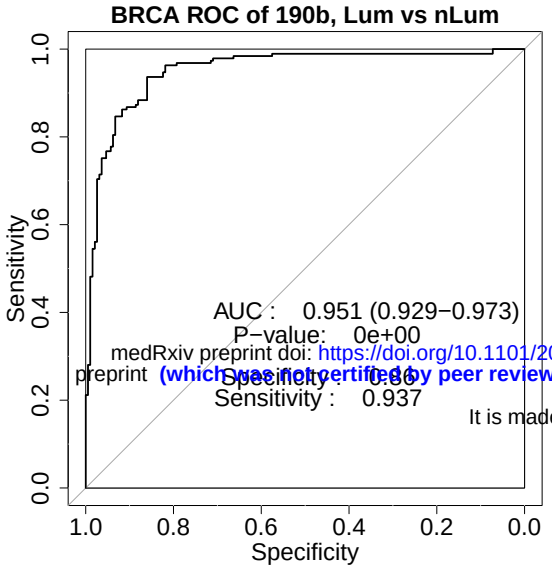
C)



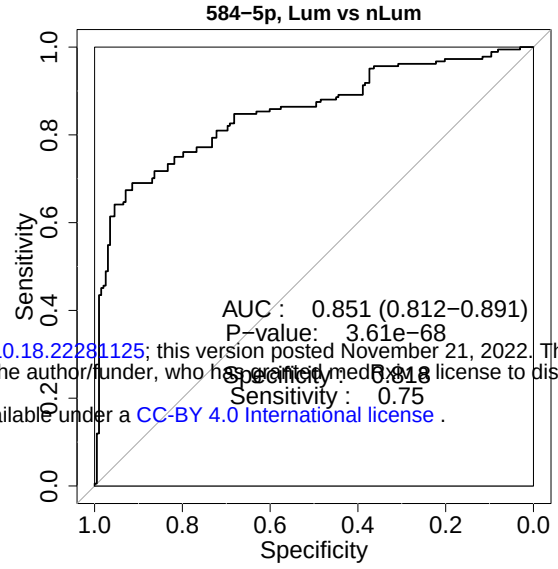
F)



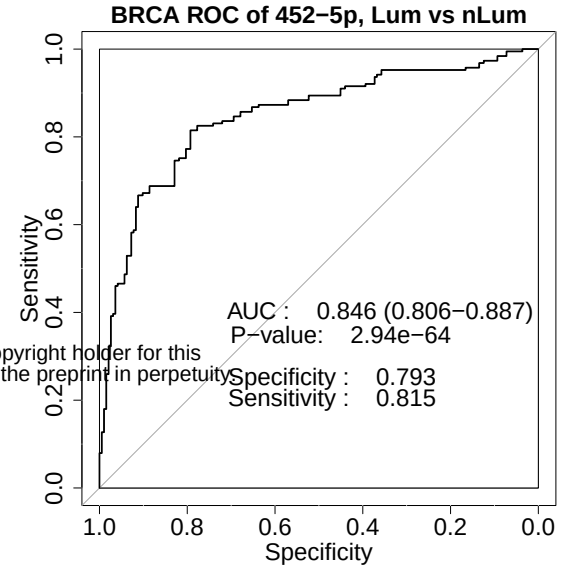
A)



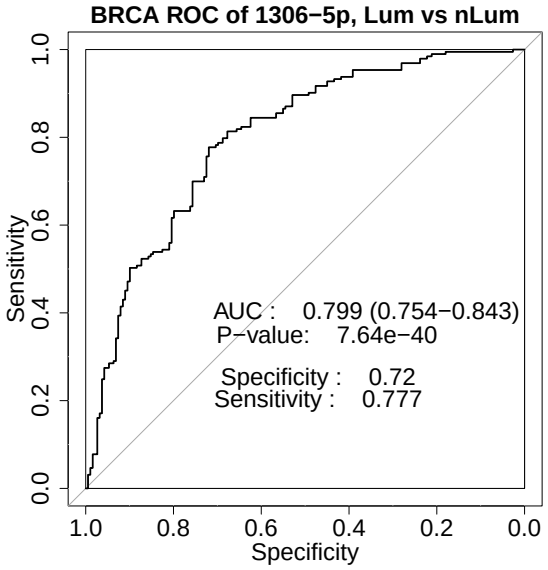
B)



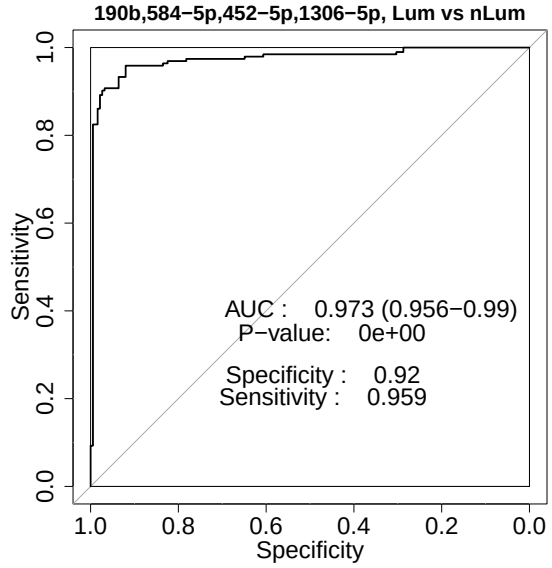
C)



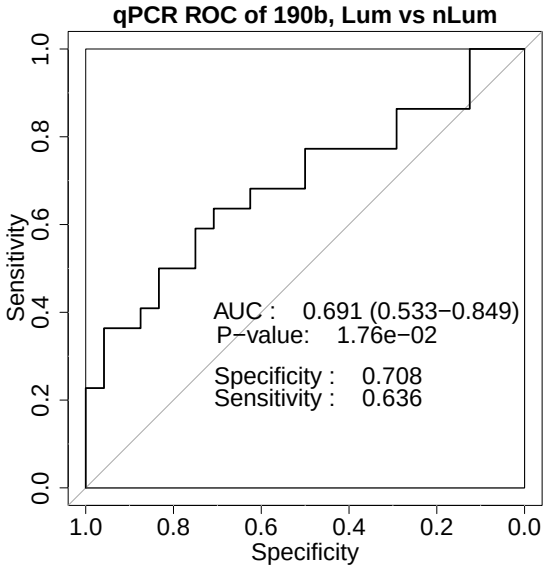
D)



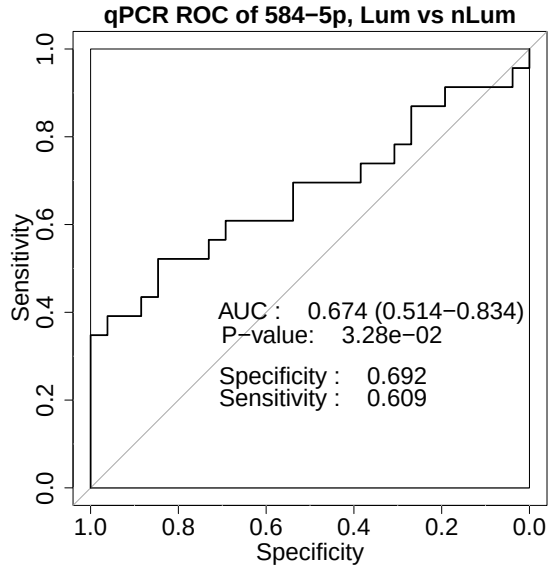
E)



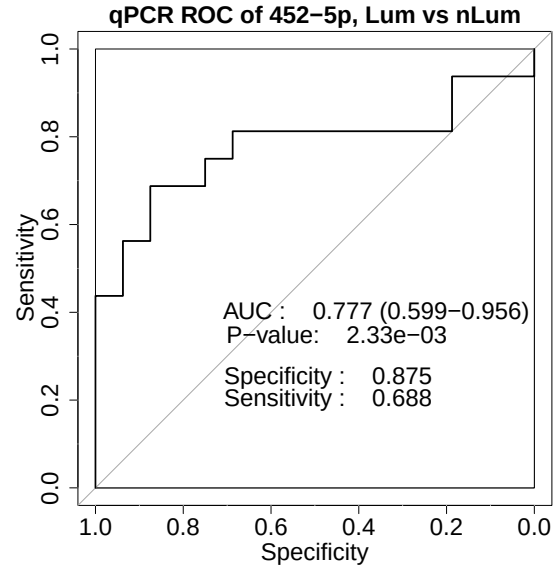
F)



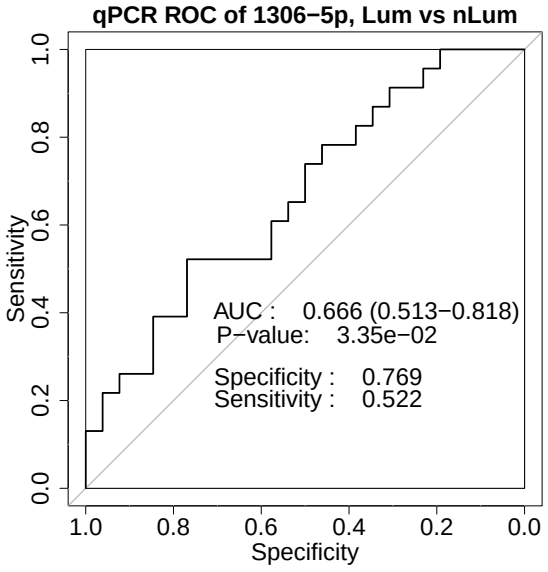
G)



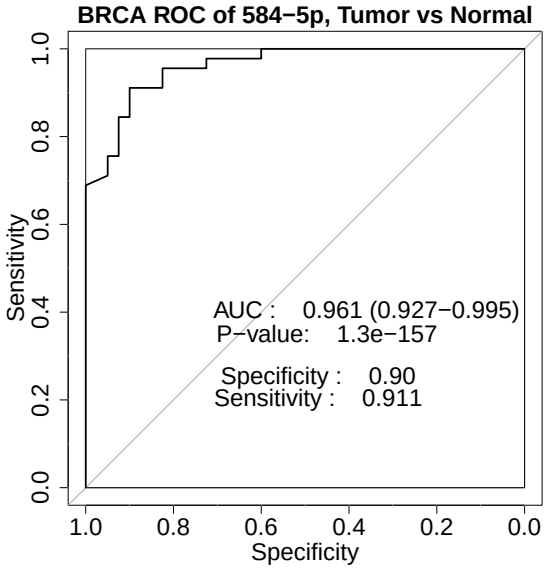
H)



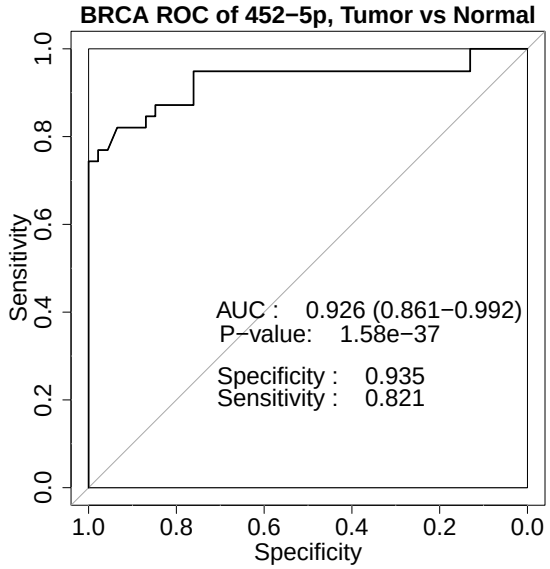
I)



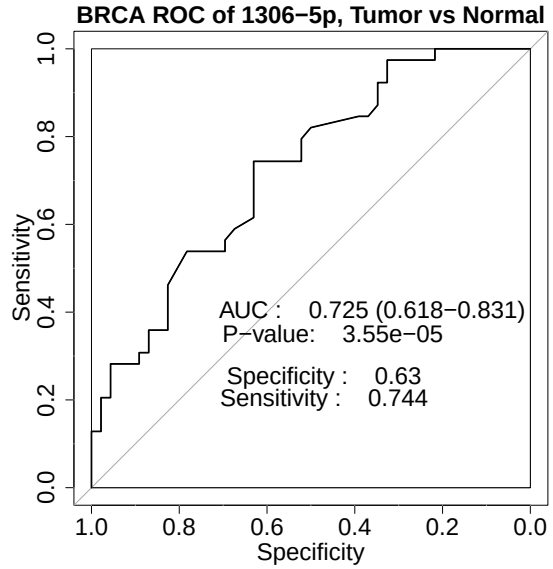
A)


BRCA ROC of 584−5p, Tumor vs Normal
AUC : 0.961 (0.927−0.995)
P−value: 1.3e−157
Specificity : 0.90
Sensitivity : 0.911
Sensitivity
Specificity

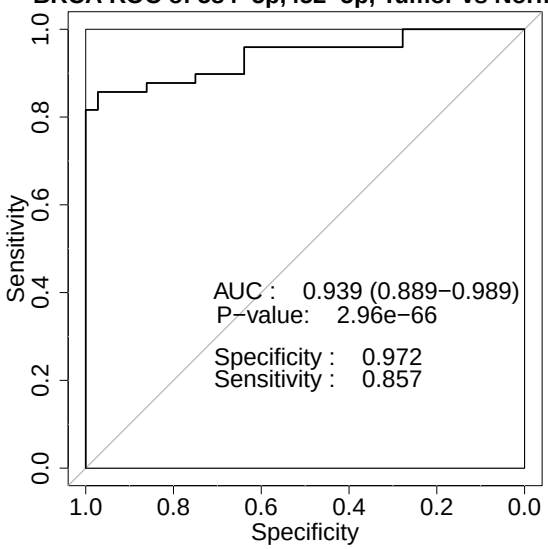
B)


BRCA ROC of 452−5p, Tumor vs Normal
AUC : 0.926 (0.861−0.992)
P−value: 1.58e−37
Specificity : 0.935
Sensitivity : 0.821
Sensitivity
Specificity

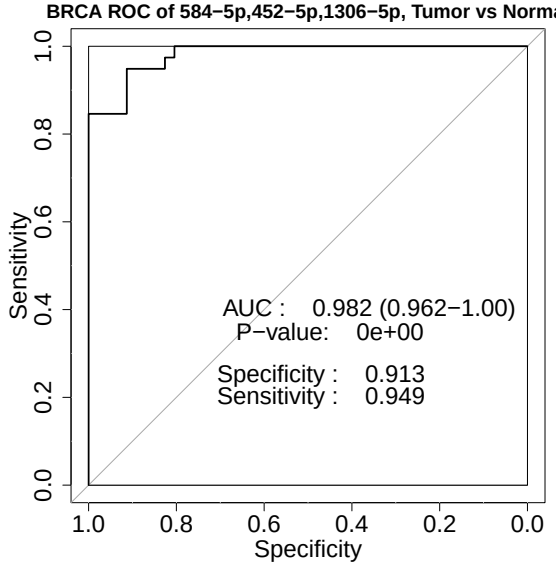
C)


BRCA ROC of 1306−5p, Tumor vs Normal
AUC : 0.725 (0.618−0.831)
P−value: 3.55e−05
Specificity : 0.63
Sensitivity : 0.744
Sensitivity
Specificity

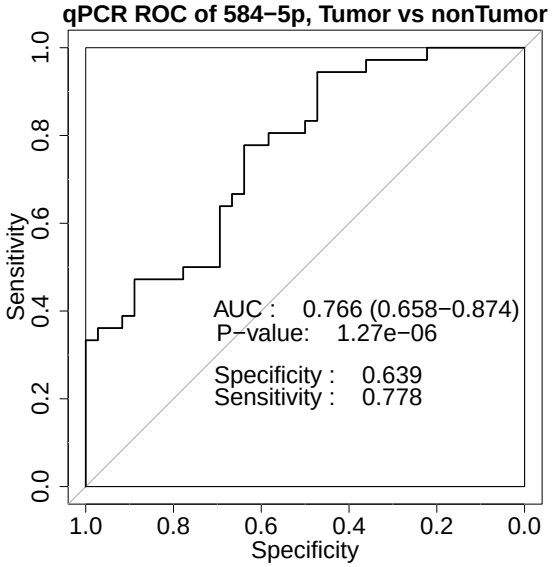
D)


BRCA ROC of 584−5p,452−5p, Tumor vs Normal
AUC : 0.939 (0.889−0.989)
P−value: 2.96e−66
Specificity : 0.972
Sensitivity : 0.857
Sensitivity
Specificity

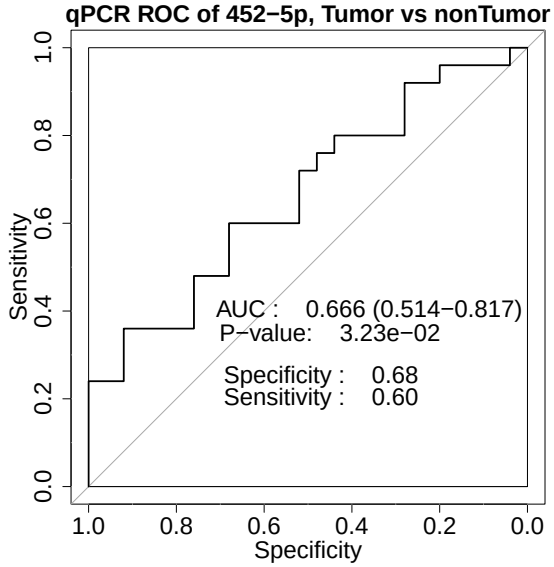
E)


BRCA ROC of 584−5p,452−5p,1306−5p, Tumor vs Normal
AUC : 0.982 (0.962−1.00)
P−value: 0e+00
Specificity : 0.913
Sensitivity : 0.949
Sensitivity
Specificity

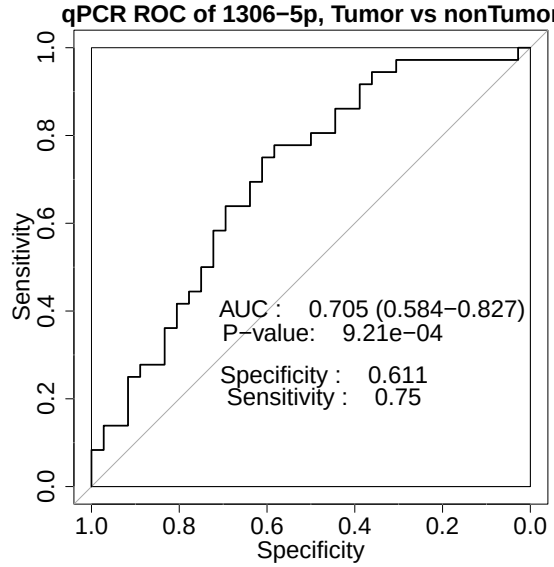
F)


qPCR ROC of 584−5p, Tumor vs nonTumor
AUC : 0.766 (0.658−0.874)
P−value: 1.27e−06
Specificity : 0.639
Sensitivity : 0.778
Sensitivity
Specificity

G)


qPCR ROC of 452−5p, Tumor vs nonTumor
AUC : 0.666 (0.514−0.817)
P−value: 3.23e−02
Specificity : 0.68
Sensitivity : 0.60
Sensitivity
Specificity

H)


qPCR ROC of 1306−5p, Tumor vs nonTumor
AUC : 0.705 (0.584−0.827)
P−value: 9.21e−04
Specificity : 0.611
Sensitivity : 0.75
Sensitivity
Specificity