Identification of Differentially Expressed miRNA by Next Generation Sequencing in Locally Advanced Breast Cancer Patients of South Indian Origin

Shyam Kumar Tripathi, Jayanthi Mathaiyan*, Smita Kayal, Rajesh Nachiappa Ganesh

Abstract

**Purpose:** miRNAs are known to be aberrantly expressed in the serum, tissue, and Peripheral Blood Mononuclear Cells (PBMC) of cancer patients and could serve as potential noninvasive diagnostic markers for breast cancer. The aim of this study was to identify the differentially expressed miRNA using next-generation sequencing (NGS) from the paired PBMC samples from breast cancer patients and age-matched healthy individuals and explore their functional significance. **Methods:** In this study, PBMCs were employed for the detection of miRNAs by NGS in locally advanced breast cancer (LABC) women of South Indian origin who were divided into three age groups, (a) 40yrs-50yrs (b) 50yrs-60yrs and (c) 60yrs-70yrs, compared with age-matched control groups. **Results:** Four miRNAs (hsa-miR-192-5p, hsa-miR-24-2-2p, hsa-miR-3609, and hsa-miR-664b-3p) were found to be differentially expressed among LABC patients compared with age matched healthy women of the South Indian population. While miR-24-2-5p, miR3609, and miR-664b-3p were down-regulated, miR-192-5p was up-regulated. Gene Ontology (GO) annotations implicated miRNA with signaling pathways in peripheral nerve synapses, glutamatergic synapse, and cell morphogenesis, all of which play a pivotal role in the manifestation of cancer. **Conclusion:** Four miRNAs- 3 (While miR-24-2-5p, miR3609, and miR-664b-3p) downregulated and one upregulated (miR-192-5p) were identified as potential biomarkers for patients with locally advanced breast cancer. These markers could be validated in studies with a larger sample size.

**Keywords:** Next generation sequencing- miRNA- breast cancer

Introduction

Breast cancer (BC) is the most common cancer in women worldwide, affecting 12% of all women and leading to 450,000 deaths each year (Torre et al., 2015) which accounts for 15% of cancer related deaths among women.

Genomic profiling of tumors to predict response to therapies has become an approach of precision medicine (Stover et al., 2015). Identification of biomarkers, recognition of potential targets for therapy and plausible predictions using bioinformatics platforms have accelerated the research. Such multiple approach leads the discovery of microRNAs, a large family of non-coding RNAs that play crucial roles in cellular processes and disease manifestations.

Micro RNAs or miRNAs are non-coding single stranded RNAs of 20-22 nucleotides in length. They exert regulatory functions on the expression of multiple genes by initiating translational inhibition or degradation of their complementary mRNA targets. The varied characteristics of miRNAs allows for regulation of various cellular processes at different developmental stages (Loh et al., 2019). Various studies indicate that expression of miRNAs are tissue-specific and that are reflected in pathophysiological processes like tumorigenesis and metastasis. miRNAs can be detected not only in tissue but also in blood, serum, urine and other sources that are accessible with minimal invasiveness. Hence identification of miRNAs and the related target mRNAs has become the paramount focus of various clinical research studies, particularly cancer studies (Mallanna et al., 2010).

Expression levels of a few miRNAs were found to be associated with the initiation, development and manifestation of breast cancer, its prognosis and treatment at cellular and subcellular levels (Ohzawa et al., 2017). Dysregulation of miRNA expression and its role in the progression of mammary cancer was first described in 2005 by Iorio et al., (2005) and since then many miRNAs with aberrant expression in conjunction with breast cancer progression have been reported. An et al., (2017) have reported the plausible role of miRNA dysregulation in...
resistance to therapeutic procedures for breast cancer.

Having perceived the role of miRNAs and their role in breast cancer development, the race is on for the discovery of novel miRNAs and their target mRNAs so as to identify a novel target or therapeutic means to fight cancer. However, the discovery and profiling of miRNAs using microarrays and RT-qPCR has limitations including low sensitivity to miRNA expression, difficulty deducing structural changes and the inability to identify novel miRNAs. To overcome this limitation massive parallel sequencing technology like NGS would be a better alternative to discover novel miRNA which occurs at low frequency. The objective of this pilot study was to identify differentially expressed miRNA in the PBMC of patients with locally advanced breast cancer compared with age and gender-matched healthy control by using NGS. In particular, we have focused on the mechanism of profiling miRNA expression associated with breast cancer through examining the expression of their targets, followed by pathway analyses. Hence aim of the present study was to identify the differentially expressed miRNA using next-generation sequencing (NGS) from the paired PBMC samples of locally advanced breast cancer patients and age-matched healthy individuals that have the potential to affect breast cancer tumorigenesis and progression.

Materials and Methods

Study Population and Characteristics

Fifteen patients who were diagnosed with LABC during the year 2018 were selected from a regional cancer center in South India for the current study. Patients who had not undergone chemotherapy or radiotherapy but had histological evidence of disease with South Indian descent were selected and grouped into three age groups of five patients each - (a) 40yrs-50yrs (b) 50yrs-60yrs and (c) 60yrs-70yrs. Patients diagnosed with triple-negative breast cancer were excluded from the study. Simultaneously, fifteen age and gender matched healthy individuals with no history of chronic disease or acute illnesses were identified as controls for the study. The institutional ethics committee approved the study (JIP/IEC/2017/0267). All patients had not undergone chemotherapy or radiotherapy but had histological evidence of disease with South Indian descent and were selected and grouped into three age groups of five patients each - (a) 40yrs-50yrs (b) 50yrs-60yrs and (c) 60yrs-70yrs. Patients diagnosed with triple-negative breast cancer were excluded from the study. Simultaneously, fifteen age and gender matched healthy individuals with no history of chronic disease or acute illnesses were identified as controls for the study. The institutional ethics committee approved the study (JIP/IEC/2017/0267). All the details were explained and written informed consent was taken before including the subject into the study. Blood samples were collected from the study and control groups and stored for further analysis.

RNA isolation and sequencing

For the identification of differentially expressed miRNAs, 5 mL of blood was collected in EDTA tubes from breast cancer patients and their corresponding controls. The samples were then subjected to density gradient centrifugation and peripheral blood mononuclear cells (PBMC) were isolated. RNA was isolated from the PBMCs using miRNA easy kit (Qiagen, Gmbtt Germany) in accordance with the manufacturer’s instructions. The yield of RNA was checked by Nanodrop (Bio photometer plus) and Fluorimeter (Qubit4) and the samples were stored at -80°C. The samples were divided into three groups of five each based on the age of patients and controls for expression studies as shown in Table 1.

RNA samples were pooled for each age group and a single sample per group was obtained. This was done as an attempt to eliminate biological variations as a result of differential gene expressions between samples from the same age group.

The six pooled RNA samples were quantified by using the Qubit BR Assay Kit (Invitrogen cat# Q10211). RNA purity check was done by using QIAxpert and RIN (RNA integrity number) was assessed on Tape station using RNA screen tape (Agilent, Cat# 5067-5576) to detect purity, concentration, and integrity of RNA sample. The RIN number is based on the numbering system from 1 to 10 with 1 being most degraded and 10 the most intact. Next-generation sequencing (NGS) with Illumina (HiSeq2500) technology was used for the miRNA study as explained below.

Statistical Analysis

Differential expression analysis was performed with read count of the known miRNA using DESeq2 to compare miRNA expression between LABC patient and aged matched healthy individual. Log2FC≥ (1) for upregulated and log2FC≤ (-1) for downregulated miRNAs with p-adj ≤0.05 was considered.

Library preparation and Sequencing

The extracted RNA was used to prepare small RNA libraries using the NEB Next Multiplex Small RNA Library Prep kit protocol (NEB, #E7560S) and in accordance with the manufacturer’s protocol. Qubit High Sensitivity Assay (Invitrogen, Cat# Q32852) was employed for the quantification of the same. The libraries were then pooled and sequencing was carried out on the Illumina HiSeq 2500.

Bioinformatics analysis: Figure 1

The sequencing of samples resulted in generation of 0.9 to 1.7 GB raw data from each sample. The raw data was subjected to quality check and further analysis.

Pre-processing of data

Raw fastq sequences (PRJNA615345) thus generated were de-multiplexed with CASAVA v1.8 pipeline. The raw reads were quality checked with FastQC. FastQC tool analyzes the base quality score distribution, sequence quality score distribution, average base content per read, GC distribution in the reads and read-length distribution. Unwanted sequences such as adapter sequences, primers, poly-A tails and other similar sequences are trimmed using Cutadapt tool (http://www.bioinformatics.babraham.ac.uk/projects/cutadapt). Structural RNA contamination including ribosomal RNAs and transfer RNA sequences were removed using Bowtie2 (version2.2.4) with default parameters (Table 1 in Supplementary data).

Prediction of miRNAs

To identify both known and novel miRNAs, and to determine differential gene expression profiles, the quality controlled reads were aligned to the reference human genome (Human genome - hg19) and to a miRNA database, such as miRBase (http://www.mirbase.org/). The
reads which map to multiple positions within a genome and/or map to known small RNA coordinates (e.g., snRNA, rRNA, tRNA), along with any reads that do not map to the reference genome, were discarded.

miRDeep2 is a software package which allows both expression profiling and identification of novel miRNAs. The quality checked reads were mapped to the reference human genome (hg19) from Illumina iGenomes (http://support.illumina.com/sequencing/sequencing_software/igenome.html). miRDeep2 is used to identify known as well as novel micro-RNA.

Differential expression analysis
The sRNA sequences were aligned against mature miRNA sequences from miRbase22. Reads were aggregated into counts and the expression values were quantified. Differential expression analysis was done based on normalized deep sequencing counts in RPM (Reads Per million) against known miRNAs using DESeq2. Log2 fold changes with positive (Log2fc≥1) values were taken as up-regulated genes and the same with negative values (Log2fc≤-1) as down-regulated genes.

Functional analysis of target genes
The potential functions of the differentially expressed miRNAs were identified using the Amigo tools in which the targets of the up and down-regulated miRNAs were used as the input for the gene set enrichment analysis. This include GO (Biological process, cellular component, and molecular function) and Reactome Pathway. In each category, 10 prime values were considered for further analysis. A p-value of less than 0.05 was considered statistically significant.

Results
The present study was aimed to identify potential miRNA signatures related with breast carcinoma in PBMCs on an NGS platform. Expression profiles for the paired samples were generated on the Illumina HiSeq2500. More than 15 million reads were obtained for each sample. The reads were subjected to quality control checks using FastQC and trimmed where necessary to retain high quality sequences for further analysis using Cutadapt. In addition, using Bowtie, low-quality sequence reads were discarded (Q < 20). Also, other RNA contamination and sequences <17 bases were quality filtered using the same, 1-2% data loss was observed at every step of pre-processing.

miRNA prediction, identification and quantification
Using miRDeep2, the reads were aligned to the reference human genome [Human genome - hg19 from Illumina iGenomes (http://support.illumina.com/sequencing/sequencing_software/igenome.html)] and to a miRNA database, miRBase (http://www.mirbase.org/). This was done to identify both known and novel miRNAs. The sample wise distribution of known as well as novel miRNAs is summarized in (Table 2 in supplementary data) while the structures for the same with the highest read count in each of the samples are shown in (Figure 1a and 1b in supplementary data). In the known miRNA top 10 predominant miRNA were found across the 6 sample are shown in (Figure 2 in supplementary data), of which miR-21 identified as the key regulator of oncogenic processes.

DESeq2 was used to compute the differential expression of miRNAs for the control and cancer groups. A total of 640 miRNAs were found to be differentially expressed (Table 3 in supplementary data) in PBMCs of breast cancer patients as compared to that in gender and age-matched healthy women. Out of which only four

| Sl. No | Age group | Number of patient samples | Number of control samples |
|-------|-----------|---------------------------|--------------------------|
| 1     | 40-50     | 5                         | 5                        |
| 2     | 50-60     | 5                         | 5                        |
| 3     | 60-70     | 5                         | 5                        |

Table 1. Pooled Groups for NGS Analysis
miRNAs were found to be significantly differentially expressed with p-adjusted values (FDR) value ≤ 0.05 (Table 2).

The numbers of targets for the four differentially expressed miRNAs were predicted using two online web servers-miRDB (https://mirdb.org) and miRWalk. The target numbers ranged from 477 to 2984 (Table 2). These were analyzed using a free tool provided by Ghent University’s Bioinformatics & Evolutionary Genomics group (http://bioinformatics.psb.ugent.be/webtools/Venn/) for Venn diagram analysis (Figure 4) and five genes (MOB1B, KIAA0930, SH3TC2, RUNX2, CADM1) were...
found to be common targets for the four miRNAs.

Functional Annotations of mRNAs (Targets of miRNA)

Functional annotation is a method of recognizing functional elements along the sequence of a given genome. Functional insights into microRNAs are gained from the pathways or biological processes that are overrepresented in their target genes. Moreover, numerous genes could be regulated by a single miRNA and the vice versa is also possible. Functional annotation was performed using the miRNet web server for selected genes that were predicted as miRNA targets. GO annotations in terms of molecular functioning and Reactome pathway were not observed significantly for miRNA-192. Significant up-regulation was observed with biological processes and specifically in the regulation of the localized translations. A similar pattern was observed with cellular components wherein the transferase complexes were highly up-regulated. However, the functional analysis for the down-regulated miRNAs- miR-24-2-5p, miR3609 and miR-664b-3p, demonstrated that the target genes were involved in a variety of positive regulation processes like cell morphogenesis, nervous system development, neuron differentiation. Cellular component analysis identified a significant down regulation of the cell leading edge and cytoplasmic vesicles. Down regulation of glutamatergic synapse was also observed. GO molecular functioning demonstrated association of the down-regulated miRNA with purine ribonucleoside binding, cytoskeletal protein binding, and protein kinase activity while the Reactome pathway analysis demonstrated association with signaling pathways involving Fc-gamma Receptor activation, antimicrobial peptides and gene expression. Further analysis found that their target genes were enriched in pathways involved in many cancers including prostate, colorectal cancer, and breast cancer (Figure 3a, 3b and 3c in supplementary data).

Discussion

Our study identified four differentially expressed miRNAs (miR-3609, miR-24-2-5p, miR-664b-3p and miR-192) associated with breast cancer patient when compared with healthy individual. Of the four differentially expressed miRNAs, miR-192 was downregulated in breast cancer patients when compared to healthy individuals. This tumor suppressor property is supported by other studies in different types of cancer where miR192 was found to be downregulated. The study done by Zhang et al reported mir-192 acts as a tumor suppressor by inhibiting cell proliferation and inducing apoptosis in breast cancer (Zhang et al., 2019). In contrast, there is some evidence to suggest that mir-192 act as oncomir by promoting tumor growth in breast cancer patients. (Bhardwaj et al., 2019; Zhou et al., 2018).

The present study documented the downregulation of miR-24-2-5p in healthy individuals compare to LABC patients which indicate its oncomir nature. According to Du et al. miR-24 enhances invasion and progression of tumour by targeting PTPN9 (phosphatases tyrosine-protein phosphatase non-receptor type 9) and PTPRF (receptor-type tyrosine-protein phosphatase F) thereby promoting EGF (Epidermal growth factor) (Du et al., 2013). Its expression, however, was down regulated after chemotherapy (Camps et al., 2014; Sochor et al., 2014).

miR-3609 was found to be downregulated in healthy individual as compared with LABC patients which indicate its oncomir nature. According to our finding, miR-664b-3p was downregulated in healthy individual as compared with LABC patients which confirm its oncogenic nature.
with neoadjuvant chemotherapy in LABC patients which support its oncogenic nature (Khraman et al., 2018). In contrast, study by Wu et al confirmed tumour suppressor nature of miR-664 by suppressing the invasion and proliferation in breast cancer by targeting Insulin Receptor Substrate 1 (Wu et al., 2019). Hence further studies are needed to validate the role of miRNA-664b-3p on oncomir or tumour suppressor in breast cancer.

These conflicting evidences about the roles of miRNAs (miR-664b-3p, miR-3609, miR-192) in breast cancer pathogenesis illustrates these miRNAs exert both oncogenic and tumour suppressive effects (Khraman et al., 2018; Wu et al., 2019). The overall net oncogenic or tumour suppressive effect of a miRNA depends partly upon the balance between miRNA-mediated upregulation or downregulation of oncogenic and tumour suppressive pathways, as well as tumour immune system interactions and tumour-modifying extrinsic factors. There is a need for further in vitro or in vivo experiments, to confirm the relevance of these miRNAs.

Functional annotation from this study implicate miRNA with the manifestation of breast cancer through signaling pathways associated with glutamatergic synapse, cell morphogenesis and neuroepithelial synapses. Glutamate regulation has been previously reported in association with cancer progression and its potential as a biomarker for therapeutic purposes has been identified (Stepulak et al., 2014; Prickett and Samules, 2012). Through the peripheral sympathetic nervous system a physiological stress can be exerted which in turn enhances the metastasis of breast cancer (Pimentel et al., 2013). Also, dysregulation of transcription and translation have been time and again reported in association with the metastasis of various cancers (Imai et al., 2005). Based on the functional annotations and Reactome pathway analysis it can be assumed that the differentially expressed miRNA identified herein manifest tumor progression by acting on signaling pathways associated with peripheral synapses and cell morphogenesis.

In conclusion, four differentially expressed mi-RNAs: hsa-miR-192-5p, hsa-miR-24-2-2p, hsa-miR-3609, and hsa-miR-664b-3p were identified in the blood sample of patients with locally advanced breast cancer compared with that of age matched healthy controls. These mi–RNAs were elucidated through massive parallel sequencing and their functions have been mapped. The bioinformatics analysis showed their role in breast cancer development and may act as prognosis or diagnostic biomarker. Further analysis is required to validate these findings in more number of subjects for their plausible role in breast cancer both as diagnostic and prognostic biomarkers.

**Author Contribution Statement**

Jayanthi Mathaiyan: Conceptualization, Methodology, manuscript reviewing and editing. Shyam Kumar Tripathi: Data curation, manuscript writing and editing. Rajesh NG, Smita Kayal: Supervision, guidance in doing and writing manuscript. Availability of data and materials: Authors consent to the availability of data and materials.

**Acknowledgements**

We thank Medgenome for NGS analysis.

**Funding**

There was no external funding involved in this work. An intramural grant from Jawaharlal Institute of Postgraduate Medical Education and Research. (grant no.JIP/PHASEII/32) was obtained for the study.

**Approval**

The study was approved by JIPMER Scientific Advisory Committee (JASC 48/2017/69)

**Ethical Declaration**

The study has been carried out in accordance with the guidelines of Declaration of Helsinki, and ethics approval from JIPMER Hospital (JIP/IEC/2017/0267) written informed consent was obtained from all participants before entering the study.

**Conflict of interests**

The authors declare that they have no competing interests.

**References**

An X, Sarmiento C, Tan T, Zhu H (2017). Regulation of multidrug resistance by microRNAs in anti-cancer therapy. *Acta Pharmaceutica Sinica B*, 7, 38–51.

Bhardwaj V, Mandal A (2019). Next-Generation Sequencing Reveals the Role of Epigallocatechin-3-Gallate in Regulating Putative Novel and Known microRNAs Which Target the MAPK Pathway in Non-Small-Cell Lung Cancer A549 Cells. *Molecules (Basel, Switzerland)*, 24, 368.

Camp C, Saini HK, Mole DR, et al (2014). Integrated analysis of microRNA and mRNA expression and association with HIF binding reveals the complexity of microRNA expression regulation under hypoxia. *Mol Cancer*, 13, 28.

Du WW, Fang L, Li M, et al (2013). MicroRNA miR-24 enhances tumor invasion and metastasis by targeting PTEN and PTPRF to promote IGF signaling. *J Cell Sci*, 126, 1440–53.

Li D, Wang X, Yang M (2019). miR-3609 sensitizes breast cancer cells to adriamycin by blocking the programmed death-ligand 1 immune checkpoint. *Exp Cell Res*, 380, 20–8.

Fitzpatrick C, Bendek MF, Briones M, et al (2019). Mitochondrial ncRNA targeting induces cell cycle arrest and tumour growth inhibition of MDA-MB-231 breast cancer cells through reduction of key cell cycle progression factors. *Cell Death Dis*, 10, 423–41.

Imai Y, Ishikawa E, Asada S, Sugimoto Y (2005). Estrogen-mediated post transcriptional down-regulation of breast cancer resistance protein/ABCG2. *Cancer Res*, 65, 596–04.

Iorio MV, Ferracin M, Liu CG, et al (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer Res*, 65, 7065–70.

Khraman M, Röske A, Lauffer T, et al (2018). MicroRNA in diagnosis and therapy monitoring of early-stage triple-negative breast cancer. *Sci Rep*, 8, 11584–98.

Loh, HY, Norman BP, Lai KS, et al (2019). The Regulatory Role of MicroRNAs in Breast Cancer. *Int J Mol Sci*, 20, 4940–52.

Mallanna SK, Rizzino A (2010). Emerging roles of microRNAs in the control of embryonic stem cells and the generation of induced pluripotent stem cells. *Developmental Biol*, 344,
Identification of Deferentially Expressed miRNA in Breast Cancer by Next Generation Sequencing

Ohzawa H, Miki A, Teratani T, et al (2017). Usefulness of miRNA profiles for predicting pathological responses to neoadjuvant chemotherapy in patients with human epidermal growth factor receptor 2-positive breast cancer. Oncol Lett, 13, 1731–40.

Prickett TD, Samuels Y (2012). Molecular pathways: dysregulated glutamatergic signaling pathways in cancer. Clin Cancer Res, 18, 4240–6.

Pimentel MA, Chai MG, Le CP, et al (2013). Metastatic Cancer: Clinical and biological perspective. Sympathetic Nervous System Regulation of Metastasis. Landes Bioscience, Austin Texas USA, pp 169-79.

Slattery ML, Trivelas A, Pellatt AJ, et al (2017). Genetic variants in the TGFβ-signaling pathway influence expression of miRNAs in colon and rectal normal mucosa and tumor tissue. Oncotarget, 8, 16765–83.

Sochor M, Basova P, Pesta M, et al (2014). Oncogenic microRNAs: miR-155, miR-19a, miR-181b, and miR-24 enable monitoring of early breast cancer in serum. BMC Cancer, 14, 448-56.

Stepulak A, Rola R, Polberg K, Ikonomidou C (2014). Glutamate and its receptors in cancer. J Neural Transmission (Vienna, Austria : 1996), 121, 933–44.

Stover DG, Wagle N (2015). Precision medicine in breast cancer: genes, genomes, and the future of genomically driven treatments. Curr Oncol Rep, 17, 15.

Torre LA, Bray F, Siegel RL, et al (2015). Global cancer statistics, 2012. CA Cancer J Clin, 65, 87–08.

Website [Internet]. [cited 2019 Apr 16] Available from: http://www.who.int/cancer/detection/breastcancer/en/index1.html

Wu L, Li Y, Li J, Ma D (2019). MicroRNA-664 Targets Insulin Receptor Substrate 1 to Suppress Cell Proliferation and Invasion in Breast Cancer. Oncol Res, 27, 459–67.

Zhang Y, He Y, Lu LL, et al (2019). miRNA-192-5p impacts the sensitivity of breast cancer cells to doxorubicin via targeting peptidylprolyl isomerase A. Kaohsiung J Med Sci, 35, 17-23.

Zhou Y, Zheng X, Lu J, et al (2018). Ginsenoside 20(S)-Rg3 Inhibits the Warburg Effect Via Modulating DNMT3A/MiR-532-3p/HK2 Pathway in Ovarian Cancer Cells. Cell Physiol Biochem, 45, 2548-59.

This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.