*Corresponding author:
E-mail: rajendra@cdac.in

Analysis of H3K4me3-ChIP-Seq and RNA-Seq data to understand the putative role of miRNAs and their target genes in breast cancer cell lines

Aneesh Kotipalli, Ruma Banerjee, Sunitha Manjari Kasibhatla, Rajendra Joshi*

HPC-Medical and Bioinformatics Applications Group, Centre for Development of Advanced Computing, Pune 411008, India

Breast cancer is one of the leading causes of cancer in women all over the world and accounts for ~25% of newly observed cancers in women. Epigenetic modifications influence differential expression of genes through non-coding RNA and play a crucial role in cancer regulation. In the present study, epigenetic regulation of gene expression by *in-silico* analysis of histone modifications using chromatin immunoprecipitation sequencing (ChIP-Seq) has been carried out. Histone modification data of H3K4me3 from one normal-like and four breast cancer cell lines were used to predict miRNA expression at the promoter level. Predicted miRNA promoters (based on ChIP-Seq) were used as a probe to identify gene targets. Five triple-negative breast cancer (TNBC)-specific miRNAs (miR153-1, miR4767, miR4487, miR6720, and miR-LET7I) were identified and corresponding 13 gene targets were predicted. Eight miRNA promoter peaks were predicted to be differentially expressed in at least three breast cancer cell lines (miR4512, miR6791, miR330, miR3180-3, miR6080, miR5787, miR6733, and miR3613). A total of 44 gene targets were identified based on the 3'-untranslated regions of downregulated mRNA genes that contain putative binding targets to these eight miRNAs. These include 17 and 15 genes in luminal-A type and TNBC respectively, that have been reported to be associated with breast cancer regulation. Of the remaining 12 genes, seven (*A4GALT*, *C2ORF74*, *HRCT1*, *ZC4H2*, *ZNF512*, *ZNF655*, and *ZNF608*) show similar relative expression profiles in large patient samples and other breast cancer cell lines thereby giving insight into predicted role of H3K4me3 mediated gene regulation via the miRNA-mRNA axis.

Keywords: breast neoplasms, ChIP-Seq, luminal-A/triple-negative, miRNA, RNA-Seq

Introduction

Breast cancer is one of the leading causes of death in women all over the world [1]. There are many subtypes in breast cancer identified based on the origin, hormone receptors expression, and response to treatment. There are four basic subtypes namely luminal-A, luminal-B, human epidermal growth factor receptor-2 (HER2) and triple-negative breast cancer (TNBC) [2]. Among all the subtypes, the TNBC is very aggressive and has poor prognosis compared to other subtypes and very few systemic treatment options are available other than chemotherapy [3]. Luminal-A (estrogen receptor [ER]+ and/or progesterone receptor [PR]+, HER2–) because of ER expression, has better prognosis compared to TNBC [2]. Epigenetic regulation via histone modification of microRNA (miRNA) pro-

moters is known to play a crucial role in breast cancer regulation [4]. Histone modifications wrapped around genes play an important role in gene regulation by providing access to transcription factors, RNA-polymerases, and other regulatory mechanisms [5]. There are several histone post-translational modifications identified earlier with a unique regulatory function for each of them. Using chromatin immunoprecipitation followed by sequencing (ChIP-Seq), one can identify the position of targeted protein binding (transcription factors, histone modifications) regions in the genome [6].

Epigenetic gene regulation occurs in three major ways namely DNA methylation of CpG islands, histone modifications, and non-coding RNA mediated [7]. Each regulatory level has a crucial role in normal cell development and diseases such as cancer and other non-communicable diseases. Gene level histone modifications can be used to predict the status of a gene whether it is active or inactive. H3K4me3 modification at the promoter level provides information about active genes, whereas H3K27me3 and H3K9me3 modifications provide the inactive status of the gene [8]. Previous studies provided evidence for correlation of histone modifications (H3K36me3, H3K9ac, H3K27ac, and H3K4me1) with gene expression. Combinations of two or more histone modifications at gene promoter and gene body level provides more resolution to predict gene activity [8].

Non-coding (nc) RNAs (LncRNAs, long intervening/intergenic noncoding RNA, miRNA, and small interfering RNA) play a major role in gene regulation of different biological processes such as cell cycle and proliferation along with developmental and metabolic processes [9]. The most widely studied ncRNAs are miRNAs which are small ncRNA ~22nt in length, evolutionarily conserved, and have a wide regulatory role in development and diseases. They play an important role in gene regulation by targeting complementary binding sites in untranslated region (UTR) regions of gene transcripts or by targeting promoters or other miRNA or LncRNA [10]. Interestingly, miRNAs are also involved in the upregulation of specific genes via binding to their promoters in the nucleus and thereby controlling gene expression [11]. An actively transcribed miRNA is able to regulate ~100–1,000 genes by complementary binding to their targeted genes. Based on conserved base-pairing homology, it has been predicted that ~60% of human genes are targeted by miRNAs [12]. In many cancers dysregulation of miRNA causes cancer progression and drug resistance. The role of miRNA deregulation in breast cancer was first reported in 2005 wherein the role of miR548 as an oncogenic regulator in breast cancer was elaborated [13]. There are other miRNAs such as let-7, miR145, miR200, and miR497 with a definitive role in breast cancer [14].

The present study is limited to normal-like, luminal-A and TN-

BC-claudin subtypes of breast cancer based on hormonal receptor expression. One each of normal subtype, ER positive subtype (Luminal-A) and ER negative (TNBC-claudin) were chosen for analysis. In the current study, ChIP-Seq data corresponding to histone modification H3K4me3 for one normal-like (MCF10A) and four breast cancer cell lines (luminal-A [MCF7, ZR751], TNBC [MB231, MB436]) were chosen to understand role of miRNA-gene promoter regulation of miRNA-mRNA axis. An attempt has been made to map the epigenetic expression patterns (histone H3K4me3) with RNA-sequencing (RNA-Seq) expression of miRNA targeted genes. Analysis of such combined data promises to provide insights into understanding epigenetic gene regulation (chromatin) as well as gene expression [15].

Methods

ChIP-Seq data were downloaded from Gene Expression Omnibus (GEO) for breast cancer pertaining to six cell lines, viz., normal-like (MCF10A and 76NF2V), luminal-A subtype (MCF7 and ZR751), and TNBC subtype (MB231 and MB436), each with one activation histone modification H3K4me3 and two replicates (sequenced as Illumina single-end reads) (Supplementary Tables 1, 2) [16]. RNA-Seq data for the above-mentioned cell lines with four replicates were also downloaded from GEO (Supplementary Table 3) [16].

The raw reads were checked for quality using FastQC (version 0.11.7) [17]. BWA-MEM (Burrows-Wheeler alignment–Maximal Exact Matches) version 0.7.17 was used for alignment with reference genome build hg38 [18]. Samtools (version 1.8) was used to manage replicates and for sam to bam conversion [19]. ChIP-Seq analysis was done using model-based analysis of ChIP-Seq (MACS2, version 2.1.1.20160309) [20]. Peak calling was done using narrowpeak as H3K4me3 generates narrow histone marks [21]. p-value thresholds for peak calling were set to 0.001 for all samples and all replicates (Fig. 1A) [22]. Raw ChIP-Seq data of both H3K4me3 and corresponding input sequence was used for peak calling. To identify the reproducibility within the biological replicates IDR2.0.3 tool was used with a threshold of 0.05 to obtain statistically significant peaks [23,24]. Pseudo replicate analysis was carried out to identify low reproducible replicates which satisfy the criteria of N1/N2 ≥ 2 and Np/Nt ≥ 2 (where N1 represents the number of replicate1 self-consistent peaks, and N2 represents the number of replicate two self-consistent peaks; Np represents the number of peaks consistent between pooled pseudoreplicates, and, Nt represents the number of peaks consistent between true replicates) [24]. Peaks were annotated using HOMER (v3.12) tool [25]. Peaks corresponding to miRNA promoter genes were extract-
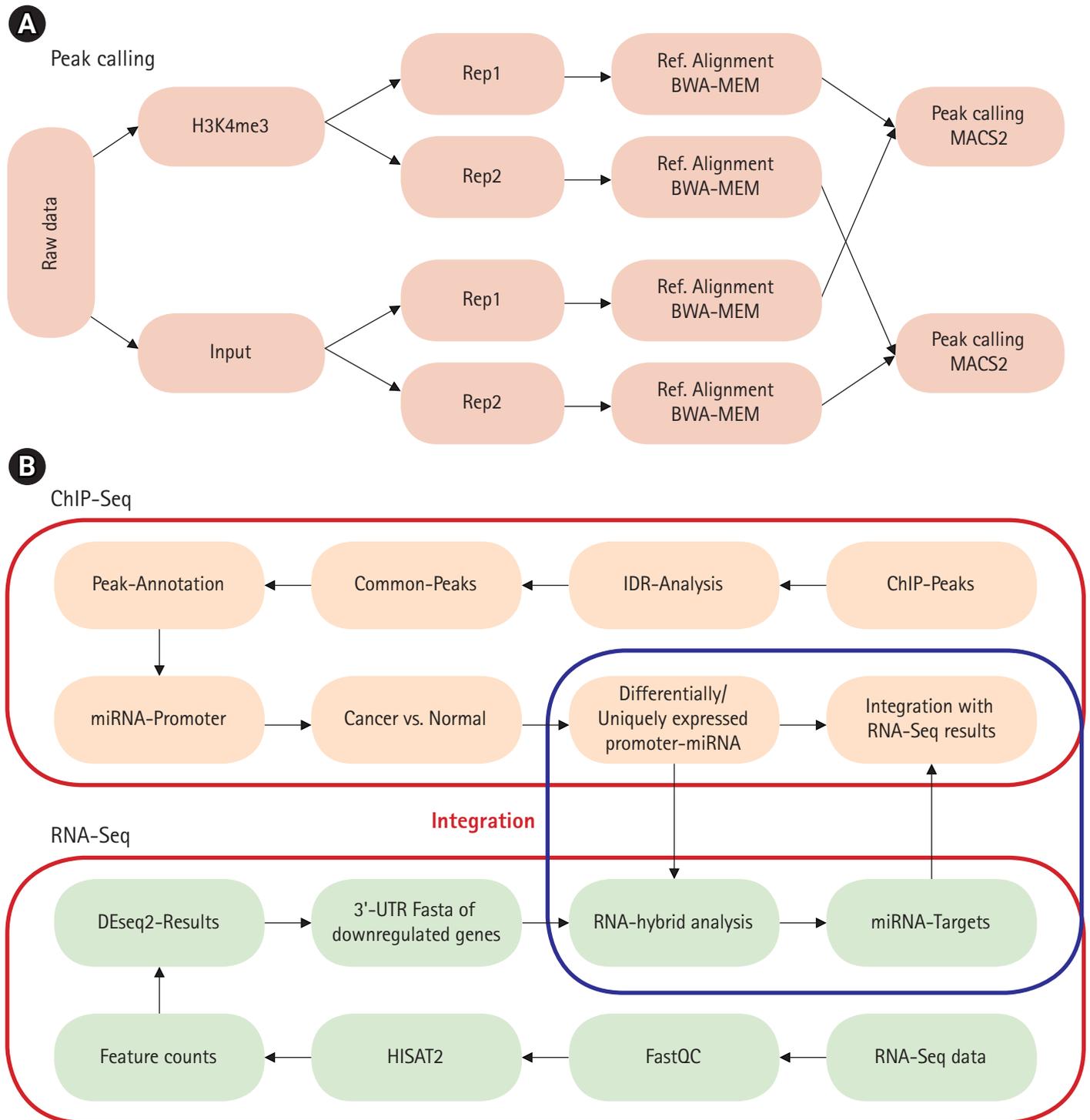
Fig. 1. (A) Chromatin immunoprecipitation sequencing (ChIP-Seq) peak calling workflow for miRNA promoter prediction. (B) ChIP-Seq and RNA sequencing (RNA-Seq) data integration workflow for prediction of miRNA-mRNA interaction via 3′-untranslated region (3′-UTR) binding target prediction.

ed using *in-house* scripts. Cis-regulatory Element Annotation System (CEAS -0.9.9.7) was used to get statistics on ChIP enrichment for genomic features such as chromosomes, promoters, gene bodies, or exons, to infer genes that are most likely to be regulated by a binding factor [26]. For RNA-Seq analysis HISAT2 (v2.1.0) was used for alignment of raw reads [27] and feature counts (v1.5.0p1) tool was used to count reads mapped on to each gene [28]. DEseq2 (v1.24) was used for differential gene expression of subtypes [29]. Downregulated and upregulated genes were selected based on log-2fold change > 2 as per guidelines for analysis of multi-omics data [30]. Differential expression of genes for normal-like (MCF10A) vs. luminal-A (MCF7 and ZR751) and normal-like (MCF10A) vs. TNBC (MB231 and MB436) was carried out (Supplementary Fig. 1). miRNA sequences were extracted and checked for complementarity with 3′-UTRs of downregulated genes. RNAhybrid server (https:// bibiserv.cebitec.uni-bielefeld.de/rnahybrid) was used to predict miRNA-mRNA interactions with seedmatch (2–8 bp) using the helix constraint "from" and "to" parameter along with cutoff of ≤ –25 kcal/mol [31]. In the current analysis, we used stringent minimum free energy (≤ –25 kcal/mol, except for miR3613 (≤ – 18 kcal/mol) due to low GC content) to predict strong putative targets with high miRNA-mRNA duplex binding stability [31]. These energy cutoffs were used based on previous studies of miR1306-ADAM10 duplex that have been experimentally validated [32]. The workflow for ChIP-Seq and RNA-Seq data integration is depicted in Fig. 1B.

Relative expressions of the gene targets identified were verified using the UALCAN (http://ualcan.path.uab.edu) and CCLE (Broad Institute Cancer Cell Line Encyclopedia, https://portals. broadinstitute.org/ccle) databases. *ACTB* gene was used as a control for relative expression analysis using the CCLE database. The UALCAN hosts the relative expression of genes across normal versus different cancer types from the TCGA (The Cancer Genome Atlas) cancer resource associated with clinicopathological data [33]. The breast cancer cell line data (60) available in CCLE database was also incorporated into the study for validation. Kaplan-Meier (KM) plots from Human Protein Atlas were used for survival analysis (https:// www.proteinatlas.org).

Results

ChIP-Seq analysis

All the ChIP-Seq datasets passed the quality check (Supplementary Fig. 2) and > 86% of reads were mapped to the reference genome for all replicates of H3K4me3 (for each cell line) used in the study. The number of peaks in the biological replicates varied from 27875 to 64652 for different cell lines (Fig. 2). Reproducibility analysis of peaks (obtained for the replicates) enabled identification of statistically significant peaks (threshold 0.05) (Table 1, Supplementary Figs. 3, 4) which are common between biological replicates [21]. In the normal cell line (MCF10A) and luminal-A subtype (cell lines MCF7 and ZR751), 16,601 peaks (59.7%), 13,008 peaks (53.4%) and 10,158 peaks (48.1%) passed the reproducibility threshold respectively. In the triple-negative subtype (cell lines MB231 and MB436), 14,339 peaks (65.7%) and 16,549 peaks (61.2%) passed the threshold. The highest percentage of overlapping peaks between the replicates was observed in cell line MB231 (65.7%) whereas the least percentage of peaks that passed the threshold was seen in cell line ZR751 (48.1%). All cell lines except 76NF2V cell line generated reproducibility ≥ 2, which indicated that the peaks

Fig. 2. Total number of peaks predicted using MACS2 for biological replicates with H3K4me3 histone modification (p = 0.001). Replicate 1 and 2 are coloured as green and blue respectively.

Table 1. Details of overlapping peaks obtained after IDR analysis of all cell lines (biological replicates)

Type	Cell line	Common peaks between replicates/IDR–pass peaks	Percentage of common peaks
Normal–like	MCF10A	16,601/27,802	59.7
Luminal-A type	MCF7	13,008/24,382	53.4
	ZR751	10,158/21,103	48.1
Triple–negative type	MB231	14,339/21,821	65.7
	MB436	16,577/27,067	61.2

Table 2. List of H3K4me3 regulated miRNA promoter-specific peaks across cell lines

Subtype	Cell line	No. of peaks
Normal	MCF10A	53
Luminal-A	MCF7	44
	ZR751	44
TNBC	MB231	42
	MB436	54

were reproducible and statistically significant. Hence, cell line 76NF2V was not used for further analysis because of the low reproducibility of replicates, N1/N2 was 3.370 (Supplementary Table 4). Chromosomal-level distribution of ChIP-peaks is available in Supplementary Fig. 5.

miRNA promoter prediction analysis

miRNA promoter regions were identified for each cell-line (Table 2, Supplementary Table 5). Peaks corresponding to miRNA-gene promoters that are common and unique between normal versus cancerous cell lines were identified (normal vs. TNBC, normal vs. luminal-A, and TNBC vs. luminal-A) (Tables 3, 4, Fig. 3). The majority of the miRNAs predicted have been reported to have a role in breast cancer (Supplementary Table 6).

Cell line-specific miRNAs obtained in this study have been listed in Table 3. Few of these miRNAs have been validated previously [34]. It is interesting to note that there are no common miRNAs between both the luminal-A cell lines used in this study; however, five TNBC-specific miRNAs viz., miR153-1, miR4767, miR4487, miR6720, and miR-LET7I were exclusively found in both the TNBC cell lines. Identification of target genes belonging to TNBC-specific miRNAs was carried out (Supplementary Tables 7, 8). It is to be mentioned that with the cutoff criteria for target-gene identification used in this study (refer to Methods section), no targets were found for miR153-1. Of the five miRNA promoters found to be upregulated in the TNBC cell lines, 3 miRNAs, viz., miR-153-1, miR-6720, and miR-LET7I were found to have similar relative expression in TCGA data samples. Of these, miR-LET7I was found to have higher expression in TNBC
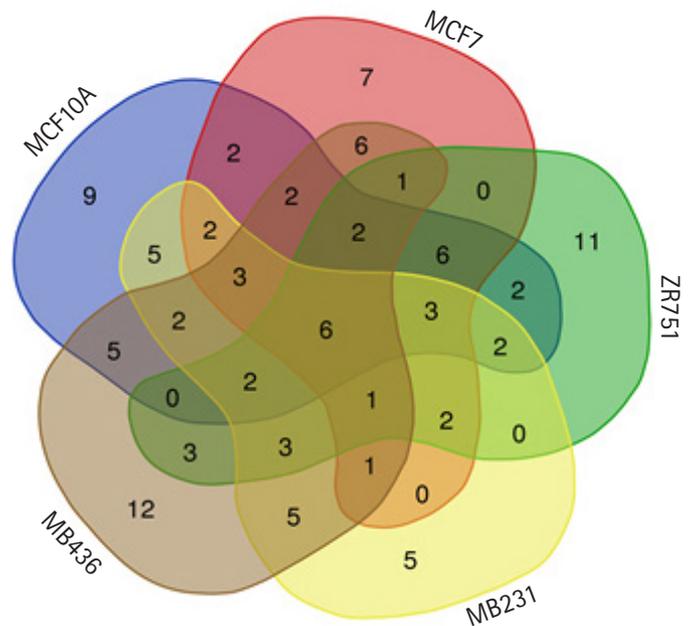
Fig. 3. Common and unique miRNAs predicted across different cell lines.

(Supplementary Fig. 6).

Eight miRNAs obtained are found to be common across two cancer subtypes (Table 4). miR4512 was observed in all cancerous cell lines, both luminal-A (MCF7 and ZR751) and TNBC (MB231 and MB436) subtypes. miR3180-3 was observed in luminal-A (MCF7 and ZR751) and TNBC subtypes (MB231 and MB436). miR6791 and miR330 were observed to be common in three cancer cell lines, two luminal-A (MCF7 and ZR751) and one TNBC (MB231) subtypes. miR5787, miR6733, and miR3613 were observed in three cancer cell lines, two TNBC (MB231 and MB436) and luminal-A (ZR751) subtypes. miR6080 was observed to be present in three cancer cell lines, two TNBC (MB231 and MB436) and luminal-A (MCF7) subtypes. All the eight miRNAs listed above have been used for further downstream analysis to identify their putative gene targets based on mRNA expression data. Of the eight miRNA promoters found to be upregulated across breast cancer cell lines, the relative expression of three miRNAs, viz., miR-

Table 3. Predicted cell line specific miRNAs

Cell line	No. of unique miRNAs	miRNA
MCF10A – Normal-like	9	miR4790
		miR4687
		miR4530
		miR6892
		miR4520-1
		miR548AJ1
		miR4279
		miR1470
		miR3675
MCF7 – Luminal-A	7	miR4734
		miR4520-2
		miR4521
		miR4519
		miR4497
		miR4477B
		miR1244-3
ZR751 – Luminal-A	11	miR6850
		miR4761
		miR200C
		miR4738
		miR1282
		miR4781
		miR7706
		miR4756
		miR6090
		miR375
		miR6515
MB231 – TNBC	5	miR1260B
		miR1258
		miR7704
		miR574
		miR4651
MB436 – TNBC	12	miR34B
		miR1184-3
		miR6875
		miR6790
		miR11401
		miR4482
		miR6743
		miR148A
		miR544B
		miR4799
		miR4466
		miR9-3

TNBC, triple-negative breast cancer.

330, miR-3613, and miR-6733 were found to be complementary in studies reported in TCGA data samples using UALCAN webserver

Table 4. Predicted TNBC and luminal-A specific miRNAs common across (≥3) cancer cell lines

Cell line	No. of common miRNAs	miRNA
MB231 (TNBC)	1	miR4512
MB436 (TNBC)		
MCF7 (Luminal-A)		
ZR751 (Luminal-A)		
MB231 (TNBC)	2	miR6791
MCF7 (Luminal-A)		miR330
ZR751 (Luminal-A)		
MB436 (TNBC)	1	miR3180-3
MCF7 (Luminal-A)		
ZR751 (Luminal-A)		
MB231 (TNBC)	1	miR6080
MB436 (TNBC)		
MCF7 (Luminal-A)		
MB231 (TNBC)	3	miR5787
MB436 (TNBC)		miR6733
ZR751 (Luminal-A)		miR3613

TNBC, triple-negative breast cancer.

(Supplementary Fig. 7). The relative expression of the other five miRNAs in this resource was found to be insufficient to draw any conclusion.

RNA-Seq analysis

All the RNA-Seq datasets passed the quality check (>28) and hence were retained for further analyses (Supplementary Fig. 8). About 96% reads mapped for cell lines MCF10A, MCF7, and MB231 whereas, for cell lines ZR751 and MB436 >93% mapping was observed (Supplementary Table 9). In normal-like vs. luminal-A type, a total of 1,189 genes were upregulated (Supplementary Table 10) and 687 genes were downregulated (Supplementary Table 11). In normal-like vs. TNBC type, a total of 954 genes were upregulated (Supplementary Table 12) and 167 genes were downregulated (Supplementary Tables 13, 14, Supplementary Fig. 9). Five miRNAs specific to the TNBC cell lines were further studied to identify their binding to downregulated 3′-UTR gene targets (Supplementary Tables 8, 15).

TNBC-specific miRNA target analysis of the downregulated genes helped in the identification of 13 genes (Supplementary Table 7). Of the 13 genes, *ADAMTSL1*, *STC2*, *CPA4*, and *NUPR1* have been previously reported (Supplementary Table 6). It is interesting to note that *FOXL2* is a target for multiple miRNAs, viz., miR4767, miR4487, and miR6720 in TNBC cell lines. Comparison of these target genes to other breast cancer cell lines from the CCLE database revealed that all of them have low expression as compared to the *ACTB* control gene (Supplementary Fig. 10).

Genes *STC2, CPA4,* and *NUPR1* were found to have a relatively higher expression amongst the 13 target genes of TNBC cell lines. In larger breast cancer samples obtained from TCGA, with the exception of *SPOCK2, CPA4, C1orf228,* and *NFE2,* the other target genes are found to have relatively low expression in TNBC as compared to normal samples (Supplementary Fig. 11). Relative expression of these genes in other cancer subtypes hints at the down-regulatory effect of TNBC-specific miRNAs. Survival plots of most of the downregulated genes (with the exception of *NUPR1, CPA4, EPHA3, ADAMTSL1,* and *ATP13A4*) were found to be associated with poor patient survival (Supplementary Fig. 12).

Eight miRNA promoters that are common across more than three cancer cell lines were also used as probes to identify the gene targets (Table 4, Supplementary Tables 16-18). A total of 44 down-regulated gene targets were identified across luminal-A and TNBC subtypes. In normal-like (MCF10A) vs. luminal-A (MCF7 and ZR751) downregulated genes, 17 genes have been predicted and their role in breast cancer has been reported earlier (*RERG, IG-FBP6, SPATA18, AXL, BMF, FXYD5, PTRF, RUNX2, UGT8, CFB, CSF3, HEG1, PLAU, PTER, S100A3, SNURF,* and *WIPF1*) [35-51] (Fig. 4, Supplementary Tables 6, 19, Supplementary Fig. 13). In normal-like (MCF10A) vs. TNBC (MB231 and MB436) downregulated genes, 15 genes have been predicted in this study and their role in breast cancer have also been previously reported (*TNFSF10, TMEM47, IQGAP2, FAT4, NUPR1, HOXC13, PRRX1, STC2,* AC108941.2, *ADAMTSL1, ARHGEF5, BNC1, CPA4, PPL,* and *TNFRSF10D*) [52-66] (Fig. 5, Supplementary Tables 6, 20, Supplementary Fig. 14).

Of the remaining 12 target genes identified, nine genes in luminal-A were identified to be regulated by their corresponding miRNAs (gene *A4GALT* targeted by miR3180-3, miR4512, and miR6791; gene *C10orf55* targeted by miR330, miR3180-3, miR5787, and miR6791; gene *C2ORF74* targeted by miR330 and miR5787; gene *ZC4H2* targeted by miR330 and miR5787; gene *ZNF512* targeted by miR330, miR3180-3, miR5787, and miR6791; gene *ZNF655* targeted by miR5787; gene *ZNF71* targeted by miR5787 and miR6791; gene *HCG2042738* targeted by miR6791; gene *HRCT1* targeted by miR4512 and miR5787) (Table 5). Similarly, three genes in TNBC were also identified to be regulated by their corresponding miRNAs (gene *HIST3H2A* targeted by miR6791; *ZNF608* targeted by miR5787; *ELOVL4* targeted by miR5787) (Supplementary Table 18). Comparison of these 12 target genes to other breast cancer cell lines from the CCLE database revealed that all of them have low expression as compared to the *ACTB* control gene (Supplementary Fig. 15). Genes *HIST3H2A* and *C2ORF74* were found to have a relatively higher expression amongst the 12 target genes. In larger datasets of breast cancer, with the exception of *ZNF71* and *HIST3H2A,* all other gene targets were found to be downregulated (Supplementary Figs. 13, 14, 16−18). This observation supports the probable role of miRNA-mRNA axis in gene regulation. The down-regulation of *A4GALT, C2ORF74, HRCT1, ZC4H2, ZNF512, ZNF655, ZNF608,* and *HIST3H2A* genes were found to be independently associated with poor survival in breast cancer patients (Table 5, Supplementary Fig. 19). It

Table 5. Predicted gene targets of differentially regulated miRNAs in breast cancer cell lines (TNBC and luminal-A) proposed using ChIP-Seq–RNA-Seq integrated analysis

No.	Gene name		miRNA	Status in TCGA/CCLE/survival plot
Luminal-A				
1	A4GALT	Alpha 1,4-galactosyltransferase	miR4512, miR6791, miR3180-3	Down/down/poor–survival
2	C10orf55	Chromosome 10 open reading frame 55	miR6791, miR330, miR3180-3, miR5787	Down/down/high–survival
3	C2ORF74	Chromosome 2 open reading frame 74	miR330, miR5787	Down/down/poor–survival
4	HCG2042738	Isoform CRA_b and AC124312.1	miR6791	Down/down/poor–survival
5	HRCT1	Histidine rich carboxyl terminus 1	miR4512, miR5787	Down/down/poor–survival
6	ZC4H2	Zinc-finger family of protein	miR330, miR5787	Down/down/poor–survival
7	ZNF512	Zinc-finger protein 512	miR3180-3, miR6080, miR5787, miR6733	Down/down/poor–survival
8	ZNF655	Zinc-finger protein 655	miR5787	Down/down/poor–survival
9	ZNF71	Zinc-finger protein 71	miR6791, miR5787	Up/up/high–survival
TNBC				
1	HIST3H2A	Histone cluster 3 H2A	miR6791	Up/up/poor–survival
2	ZNF608	Zinc-finger protein 608	miR5787	Down/down/poor–survival
3	ELOVL4	ELOngation of very long chain fatty acids-4	miR5787	Down/down/high–survival

TNBC, triple-negative breast cancer; ChIP-Seq, chromatin immunoprecipitation sequencing; RNA-Seq, RNA-sequencing; TCGA, The Cancer Genome Atlas; CCLE, Broad Institute Cancer Cell Line Encyclopedia.
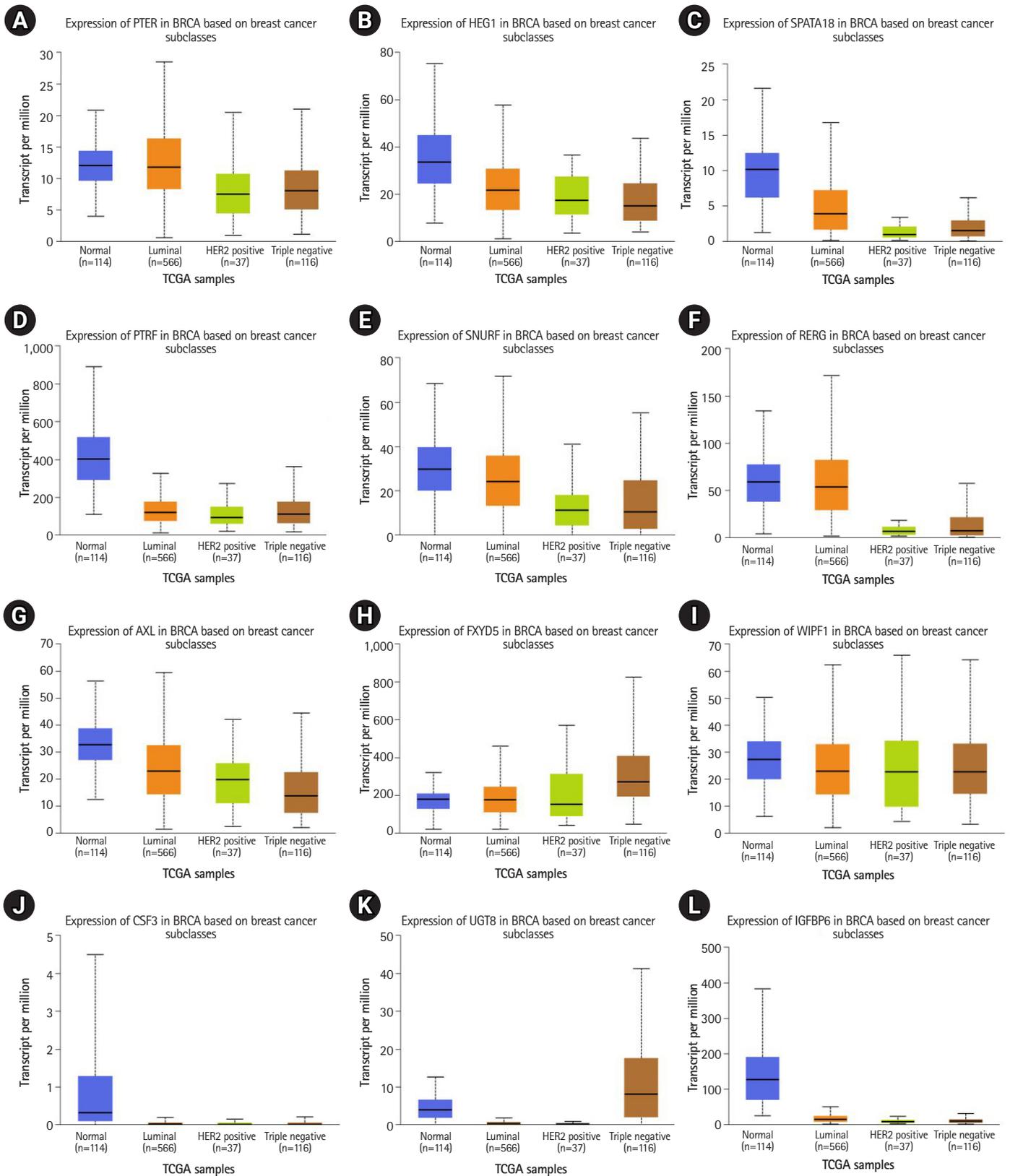
Fig. 4. Relative gene expression (The Cancer Genome Atlas [TCGA] breast cancer samples) of luminal-A downregulated gene targets (12 of the total 17 genes) previously reported in breast cancer that correlate with predicted miRNA binding analysis: (A) PTER, (B) HEG1, (C) SPATA18, (D) PTRF, (E) SNURF, (F) RERG, (G) AXL, (H) FXYD5, (I) WIPF1, (J) CSF3, (K) UGT8, and (L) IGFBP6.
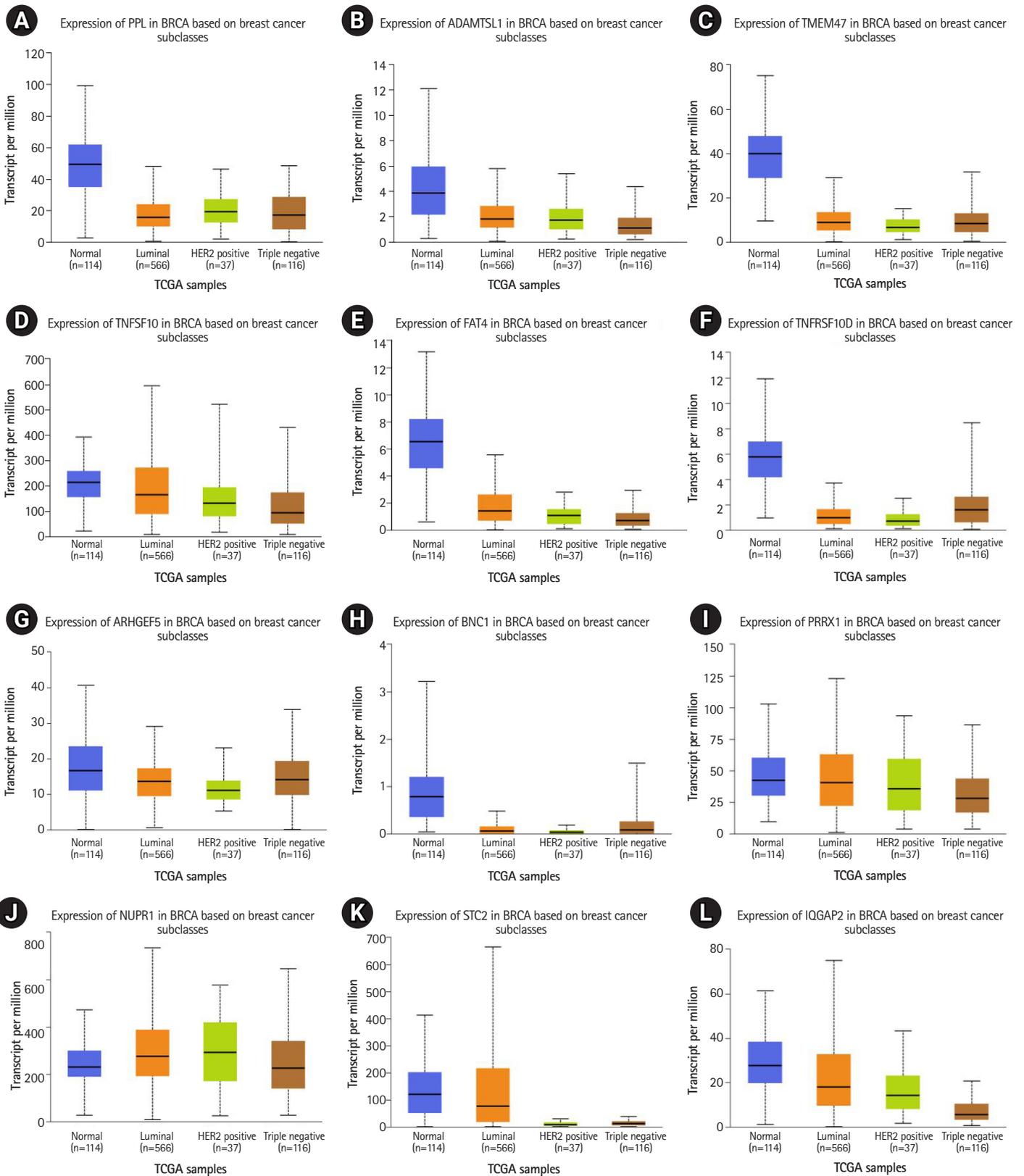
Fig. 5. Relative gene expression (TCGA breast cancer samples) of triple-negative breast cancer downregulated gene targets (12 of the total 15 genes) previously reported in breast cancer that correlate with predicted miRNA binding analysis: (A) PPL, (B) ADAMTSL1, (C) TMEM47, (D) TNFSF10, (E) FAT4, (F) TNFRSF10D, (G) ARHGEF5, (H) BNC1, (I) PRRX1, (J) NUPR1, (K) STC2, and (L) IQGAP2.

needs to be mentioned that relative expression data and survival plots for gene *HCG2042738* could not be obtained due to insufficient annotation.

Discussion

The interplay between epigenetic gene regulation through histone modifications and other regulatory mechanisms like ncRNA is of great interest in cancer biology. In the present analysis, the role of H3K4me3 in miRNA expression based on promoter level peaks has been studied using ChIP-Seq and RNA-seq data integration. To achieve the same, a novel approach of mapping data derived from ChIP-Seq (miRNA promoter peaks) and RNA-Seq (targets of 3′-UTRs of genes binding to miRNA) was used to understand epigenetic regulation that may aid in the identification of subtype and cell line specific miRNAs [15,16].

In normal-like cell line MCF10A, of the nine unique miRNAs identified, miR4530 was found to have a role in the suppression of cell proliferation, promote angiogenesis and induce apoptosis by targeting gene VASH1 (Vasohibin 1) in breast carcinoma [67]. Hence, promoter-level epigenetic regulation of miR4530 by H3K4me3 may have a protective role in normal-like subtypes. miR34B was observed to be present only in cell line MB436 (TNBC subtype). miR34B has high expression in TNBC tumors compared to normal types. Expression of miR34B highly correlates with clinical outcome of patients. Notch2 (notch receptor 2) gene that has a role controlling cell differentiation, is a direct target for miR34B [68]. miR6875 was observed in TNBC cell line MB436. According to previous reports, a high expression of miR6875 was observed in early breast cancer patients [69]. miR574-5p attenuates proliferation, migration, and epithelial mesenchymal transition (EMT) in TNBC cells by targeting genes BCL11A (BAF chromatin remodeling complex subunit) and SOX2 (SRY-Box transcription factor 2) to inhibit the SKIL (SKI like proto-oncogene)/TAZ (Tafazzin)/CTGF (connective tissue growth factor) axis [70].

Of the five TNBC subtype-specific miRNAs, mir153, miR6720, and miR-LET7I were found to be upregulated in larger breast cancer datasets belonging to TCGA. miR153 has been reported to have a tumor suppressor role and has been suggested as a prognostic marker for TNBC [34].

The majority of the predicted gene targets (total 44) overlap with previous experimental studies and include 32 gene targets (Figs. 4, 5) of eight miRNAs (miR4512, miR6791, miR330, miR3180-3, miR6080, miR5787, miR6733, and miR3613) which are identified in more than three breast cancer cell lines and absent in normal-like cell lines. Overexpression of miR330-3p in breast cancer cell lines has been reported earlier, which results in greater invasiveness in-vi-

tro, and miR330-3p–overexpressing cells also metastasize more aggressively *ex-ovo* [71]. Gene *CCBE1* (collagen and calcium binding EGF domains 1) is a direct target of miR330-3p, and knockout of *CCBE1* results in a greater invasive capacity [71]. Exosomal expression of miR3613-3p promotes breast cancer cell proliferation and metastasis. It has been previously reported that miR3613-3p levels were negatively correlated to *SOCS2* (suppressor of cytokine signaling 2) expression in breast cancer tissues [72]. Few genes were observed to be targeted by multiple miRNAs (like *A4GALT* and *FOXL2* targeted by three miRNAs each) as it is known that miR-NAs can regulate multiple targets based on seed match and sequence similarity between miRNA-mRNA [10].

Of the remaining 12 gene targets, relative gene expression of genes *A4GALT, C2ORF74, HRCT1, ZC4H2, ZNF512, ZNF655,* and *ZNF608* agree with the proposed hypothesis of H3K4me3 regulated miRNA-mRNA axis in large patient data (TCGA samples) along with their relative expression in other breast cancer cell lines (CCLE database). These genes were associated with poor survival based on KM plots (Human Protein Atlas). The proposed methodology of miRNA-mRNA regulation when analyzed in the context of other histone modifications like H3K27me3, H3K4me1, H3K9me3 will enable better insights into the underlying mechanism of breast cancer regulation.

ORCID

Aneesh Kotipalli: https://orcid.org/0000-0002-7548-3954
Ruma Banerjee: https://orcid.org/0000-0001-7890-8666
Sunitha Manjari Kasibhatla: https://orcid.org/0000-0003-0320-2209
Rajendra Joshi: https://orcid.org/0000-0003-1299-0091

Authors' Contribution

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

work is funded by the National Supercomputing Mission (NSM) of the Government of India. The authors thank anonymous reviewers for their valuable suggestions and constructive criticism in improving the manuscript.

Supplementary Materials

Supplementary data can be found with this article online at http://www.genominfo.org.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin 2015;65:5-29.

2. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006;295:2492-2502.

3. Collignon J, Lousberg L, Schroeder H, Jerusalem G. Triple-negative breast cancer: treatment challenges and solutions. Breast Cancer (Dove Med Press) 2016;8:93-107.

4. Luan QX, Zhang BG, Li XJ, Guo MY. MiR-129-5p is downregulated in breast cancer cells partly due to promoter H3K27m3 modification and regulates epithelial-mesenchymal transition and multi-drug resistance. Eur Rev Med Pharmacol Sci 2016; 20:4257-4265.

5. Yan H, Tian S, Slager SL, Sun Z. ChIP-seq in studying epigenetic mechanisms of disease and promoting precision medicine: progresses and future directions. Epigenomics 2016;8:1239-1258.

6. Furey TS. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet 2012;13:840-852.

7. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell 2012;150:12-27.

8. Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet 2011;12:7-18.

9. Prabhu KS, Raza A, Karedath T, Raza SS, Fathima H, Ahmed EI, et al. Non-coding RNAs as regulators and markers for targeting of breast cancer and cancer stem cells. Cancers (Basel) 2020;12: 351.

10. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009;136:215-233.

11. Place RF, Li LC, Pookot D, Noonan EJ, Dahiya R. MicroRNA-373 induces expression of genes with complementary promoter sequences. Proc Natl Acad Sci U S A 2008;105:1608-1613.

12. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. PLoS Biol 2005;3:e85.

13. Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res 2005;65:7065-7070.

14. Khalife H, Skafi N, Fayyad-Kazan M, Badran B. MicroRNAs in breast cancer: new maestros defining the melody. Cancer Genet 2020;246-247:18-40.

15. Ramos J, Felty Q, Roy D. Integrated Chip-Seq and RNA-Seq data analysis coupled with bioinformatics approaches to investigate regulatory landscape of transcription modulators in breast cancer cells. Methods Mol Biol 2020;2102:35-59.

16. Franco HL, Nagari A, Malladi VS, Li W, Xi Y, Richardson D, et al. Enhancer transcription reveals subtype-specific gene expression programs controlling breast cancer pathogenesis. Genome Res 2018;28:159-170.

17. Andrews S. FASTQC: a quality control tool for high throughput sequence data. Cambridge: Babraham Institute, 2010. Accessed 2021 May 30. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

18. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26:589-595.

19. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-2079.

20. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9:R137.

21. Bailey T, Krajewski P, Ladunga I, Lefebvre C, Li Q, Liu T, et al. Practical guidelines for the comprehensive analysis of ChIP-seq data. PLoS Comput Biol 2013;9:e1003326.

22. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc 2012;7:1728-1740.

23. Li Q, Brown JB, Huang H, Bickel PJ. Measuring reproducibility of high-throughput experiments. Ann Appl Stat 2011;5:1752-1779.

24. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res 2012;22: 1813-1831.

25. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010;38:576-589.

26. Shin H, Liu T, Manrai AK, Liu XS. CEAS: cis-regulatory element annotation system. Bioinformatics 2009;25:2605-2606.

27. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 2019;37:907-915.

28. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923-930.

29. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.

30. Hollbacher B, Balazs K, Heinig M, Uhlenhaut NH. Seq-ing answers: current data integration approaches to uncover mechanisms of transcriptional regulation. Comput Struct Biotechnol J 2020;18:1330-1341.

31. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. RNA 2004; 10:1507-1517.

32. Augustin R, Endres K, Reinhardt S, Kuhn PH, Lichtenthaler SF, Hansen J, et al. Computational identification and experimental validation of microRNAs binding to the Alzheimer-related gene ADAM10. BMC Med Genet 2012;13:35.

33. Chandrashekar DS, Bashel B, Balasubramanya SA, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia 2017;19:649-658.

34. Shi D, Li Y, Fan L, Zhao Q, Tan B, Cui G. Upregulation of miR-153 inhibits triple-negative breast cancer progression by targeting ZEB2-mediated EMT and contributes to better prognosis. Onco Targets Ther 2019;12:9611-9625.

35. Hsu PC, Ho JY, Yu CP. RERG involvement in the RAS pathway and ER-dependent transcription in breast cancer. J Clin Oncol 2019;37(15 Suppl):e14638.

36. Nikulin SV, Raigorodskaya MP, Poloznikov AA, Zakharova GS, Schumacher U, Wicklein D, et al. In vitro model for studying of the role of IGFBP6 gene in breast cancer metastasizing. Bull Exp Biol Med 2018;164:688-692.

37. Bornstein C, Brosh R, Molchadsky A, Madar S, Kogan-Sakin I, Goldstein I, et al. SPATA18, a spermatogenesis-associated gene, is a novel transcriptional target of p53 and p63. Mol Cell Biol 2011;31:1679-1689.

38. Colavito SA. AXL as a target in breast cancer therapy. J Oncol 2020;2020:5291952.

39. Hornsveld M, Tenhagen M, van de Ven RA, Smits AM, van Triest MH, van Amersfoort M, et al. Restraining FOXO3-dependent transcriptional BMF activation underpins tumour growth and metastasis of E-cadherin-negative breast cancer. Cell Death Differ 2016;23:1483-1492.

40. Lee YK, Lee SY, Park JR, Kim RJ, Kim SR, Roh KJ, et al. Dysadherin expression promotes the motility and survival of human breast cancer cells by AKT activation. Cancer Sci 2012;103: 1280-1289.

41. Bai L, Deng X, Li Q, Wang M, An W, Deli A, et al. Down-regulation of the cavin family proteins in breast cancer. J Cell Biochem 2012;113:322-328.

42. McDonald L, Ferrari N, Terry A, Bell M, Mohammed ZM, Orange C, et al. RUNX2 correlates with subtype-specific breast cancer in a human tissue microarray, and ectopic expression of Runx2 perturbs differentiation in the mouse mammary gland. Dis Model Mech 2014;7:525-534.

43. Cao Q, Chen X, Wu X, Liao R, Huang P, Tan Y, et al. Inhibition of UGT8 suppresses basal-like breast cancer progression by attenuating sulfatide-alphaVbeta5 axis. J Exp Med 2018;215:1679-1692.

44. Suman S, Basak T, Gupta P, Mishra S, Kumar V, Sengupta S, et al. Quantitative proteomics revealed novel proteins associated with molecular subtypes of breast cancer. J Proteomics 2016;148:183-193.

45. Hollmen M, Karaman S, Schwager S, Lisibach A, Christiansen AJ, Maksimow M, et al. G-CSF regulates macrophage phenotype and associates with poor overall survival in human triple-negative breast cancer. Oncoimmunology 2016;5:e1115177.

46. Zhao YR, Wang JL, Xu C, Li YM, Sun B, Yang LY. HEG1 indicates poor prognosis and promotes hepatocellular carcinoma invasion, metastasis, and EMT by activating Wnt/beta-catenin signaling. Clin Sci (Lond) 2019;133:1645-1662.

47. Lin M, Zhang Z, Gao M, Yu H, Sheng H, Huang J. MicroRNA-193a-3p suppresses the colorectal cancer cell proliferation and progression through downregulating the PLAU expression. Cancer Manag Res 2019;11:5353-5363.

48. Hung CM, Liu LC, Ho CT, Lin YC, Way TD. Pterostilbene enhances TRAIL-induced apoptosis through the induction of death receptors and downregulation of cell survival proteins in TRAIL-resistance triple negative breast cancer cells. J Agric Food Chem 2017;65:11179-11191.

49. Gianni M, Terao M, Kurosaki M, Paroni G, Brunelli L, Pastorelli R, et al. S100A3 a partner protein regulating the stability/activity of RARalpha and PML-RARalpha in cellular models of breast/lung cancer and acute myeloid leukemia. Oncogene 2019;38: 2482-2500.

50. Saville B, Poukka H, Wormke M, Janne OA, Palvimo JJ, Stoner M, et al. Cooperative coactivation of estrogen receptor alpha in ZR-75 human breast cancer cells by SNURF and TATA-binding protein. J Biol Chem 2002;277:2485-2497.

51. Garcia E, Machesky LM, Jones GE, Anton IM. WIP is necessary for matrix invasion by breast cancer cells. Eur J Cell Biol 2014;93:413-423.

52. He W, Wang Q, Xu J, Xu X, Padilla MT, Ren G, et al. Attenuation of TNFSF10/TRAIL-induced apoptosis by an autophagic sur-

vival pathway involving TRAF2- and RIPK1/RIP1-mediated MAPK8/JNK activation. Autophagy 2012;8:1811-1821.

53. Guo L, Zhang K, Bing Z. Application of a coexpression network for the analysis of aggressive and nonaggressive breast cancer cell lines to predict the clinical outcome of patients. Mol Med Rep 2017;16:7967-7978.

54. Kumar D, Hassan MK, Pattanaik N, Mohapatra N, Dixit M. IQGAP2 acts as a tumor suppressor in breast cancer and its reduced expression promotes cancer growth and metastasis by MEK/ERK signalling pathways. Preprint at https://doi.org/10.1101/651034 (2019).

55. Hou L, Chen M, Zhao X, Li J, Deng S, Hu J, et al. FAT4 functions as a tumor suppressor in triple-negative breast cancer. Tumour Biol 2016;37:16337-16343.

56. Chowdhury UR, Samant RS, Fodstad O, Shevde LA. Emerging role of nuclear protein 1 (NUPR1) in cancer biology. Cancer Metastasis Rev 2009;28:225-232.

57. Li C, Cui J, Zou L, Zhu L, Wei W. Bioinformatics analysis of the expression of HOXC13 and its role in the prognosis of breast cancer. Oncol Lett 2020;19:899-907.

58. Dong J, Lv Z, Chen Q, Wang X, Li F. PRRX1 drives tamoxifen therapy resistance through induction of epithelial-mesenchymal transition in MCF-7 breast cancer cells. Int J Clin Exp Pathol 2018;11:2629-2635.

59. Hou J, Wang Z, Xu H, Yang L, Yu X, Yang Z, et al. Stanniocalcin 2 suppresses breast cancer cell migration and invasion via the PKC/claudin-1-mediated signaling. PLoS One 2015;10:e0122179.

60. Ji XW, Zhou TY, Lu Y, Wei MJ, Lu W, Cho WC. Breast cancer treatment and sulfotransferase. Expert Opin Ther Targets 2015;19:821-834.

61. Li Z, Guo X, Wu Y, Li S, Yan J, Peng L, et al. Methylation profiling of 48 candidate genes in tumor and matched normal tissues from breast cancer patients. Breast Cancer Res Treat 2015;149:767-779.

62. Debily MA, Camarca A, Ciullo M, Mayer C, El Marhomy S, Ba I, et al. Expression and molecular characterization of alternative transcripts of the ARHGEF5/TIM oncogene specific for human breast cancer. Hum Mol Genet 2004;13:323-334.

63. Pangeni RP, Channathodiyil P, Huen DS, Eagles LW, Johal BK, Pasha D, et al. The GALNT9, BNC1 and CCDC8 genes are frequently epigenetically dysregulated in breast tumours that metastasise to the brain. Clin Epigenetics 2015;7:57.

64. Bademler S, Ucuncu MZ, Tilgen Vatansever C, Serilmez M, Ertin H, Karanlik H. Diagnostic and prognostic significance of carboxypeptidase A4 (CPA4) in breast cancer. Biomolecules 2019;9:103.

65. Sun C, Gu Y, Chen G, Du Y. Bioinformatics analysis of stromal molecular signatures associated with breast and prostate cancer. J Comput Biol 2019;26:1130-1139.

66. Sanlioglu AD, Korcum AF, Pestereli E, Erdogan G, Karaveli S, Savas B, et al. TRAIL death receptor-4 expression positively correlates with the tumor grade in breast cancer patients with invasive ductal carcinoma. Int J Radiat Oncol Biol Phys 2007;69:716-723.

67. Zhang T, Jing L, Li H, Ding L, Ai D, Lyu J, et al. MicroRNA-4530 promotes angiogenesis by targeting VASH1 in breast carcinoma cells. Oncol Lett 2017;14:111-118.

68. Svoboda M, Sana J, Redova M, Navratil J, Palacova M, Fabian P, et al. MiR-34b is associated with clinical outcome in triple-negative breast cancer patients. Diagn Pathol 2012;7:31.

69. Xie Y, Du J, Liu Z, Zhang D, Yao X, Yang Y. MiR-6875-3p promotes the proliferation, invasion and metastasis of hepatocellular carcinoma via BTG2/FAK/Akt pathway. J Exp Clin Cancer Res 2019;38:7.

70. Zhang KJ, Hu Y, Luo N, Li X, Chen FY, Yuan JQ, et al. miR5745p attenuates proliferation, migration and EMT in triplenegative breast cancer cells by targeting BCL11A and SOX2 to inhibit the SKIL/TAZ/CTGF axis. Int J Oncol 2020;56:1240-1251.

71. Mesci A, Huang X, Taeb S, Jahangiri S, Kim Y, Fokas E, et al. Targeting of CCBE1 by miR-330-3p in human breast cancer promotes metastasis. Br J Cancer 2017;116:1350-1357.

72. Liu Y, Yang Y, Du J, Lin D, Li F. MiR-3613-3p from carcinoma-associated fibroblasts exosomes promoted breast cancer cell proliferation and metastasis by regulating SOCS2 expression. IUBMB Life 2020;72:1705-1714.