IET Systems Biology

Research Article

IET Journals
The Institution of Engineering and Technology

Identification of specific microRNA–messenger RNA regulation pairs in four subtypes of breast cancer

ISSN 1751-8849
Received on 22nd August 2019
Revised 4th October 2019
Accepted on 13th December 2019
E-First on 17th February 2020
doi: 10.1049/iet-syb.2019.0086
www.ietdl.org

Ling Guo[1,2], Aihua Zhang[1] ✉, Jie Xiong[3]

[1]*College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, 730050, People's Republic of China*
[2]*College of Electrical Engineering, Northwest University for Nationalities, Lanzhou, 730030, People's Republic of China*
[3]*Department of applied mathematics, Changsha University, Changsha, 410022, People's Republic of China*
✉ *E-mail: 435120071@qq.com*

Abstract: Four subtypes of breast cancer, luminal A, luminal B, basal-like, human epidermal growth factor receptor-enriched, have been identified based on gene expression profiles of human tumours. The goal of this study is to find whether the same groups' genes would exhibit different networks among the four subtypes. Differential expressed genes between each of the four subtypes and the normal samples were identified. The overlaps between the four groups of differentially expressed genes were used to construct regulations networks for each of the four subtypes. Univariate and multivariate Cox regressions were employed to test the genes in the four regulation networks. This study demonstrated that the common genes in four subtypes showed different regulation. Also, the hsa-miR-182 and decorin pair performs different functions among the four subtypes of breast cancer. The result indicated that heterogeneity of breast cancer is not only reflected in the different expression patterns among different genes, but also in the different regulatory networks of the same group of genes.

1 Introduction

Breast cancer (BRCA) is a serious threat to women's health. Owing to the heterogeneity and complexity of BRCA, how to effectively treat it becomes very difficult. Therefore, researchers have been devoting themselves to studying its complex molecular mechanism in order to improve the prognosis of patients. Histologically similar tumours may have different prognosis and may respond differently to therapy. It is commonly believed that these differences among histologically similar tumours are due to molecular differences. Five molecular subtypes of BRCA, luminal A (LumA), luminal B (LumB), human epidermal growth factor receptor 2 (HER2)-enriched, basal-like (Basal), and normal-like (Normal), have been identified [1]. LumA and LumB are characterised by the gene expression related to oestrogen receptor (ER) and/or progesterone receptor (PR) [2]. HER2 is characterised by frequent HER2/ERBB2 amplification (80%) [3]. The Basal subtype or as an alias, the triple negative BRCA, was reported to be the most invasive and have the poorest prognosis in clinical outcome [4]. Competing endogenous RNAs (ceRNAs) are messenger RNAs (mRNAs) which share similar microRNA (miRNA) binding site with each other. It has been widely acknowledged that ceRNAs can influence each other by competing for the miRNA for which they shared binding sites [5]. Different subtypes of BRCA require different treatments [6]. The four subtypes are not only different in specific markers and hormones, but also in other aspects. For example, differences between different races, changes in the number of copies of genes [7, 8]. To identify a gene signature for prognosis, many previous studies showed various methods for testing [9]. The research of identifying gene signature through regulating networks has more advantages at the system level [10–12]. Several research studies have employed a ceRNA network to analyse the subtypes of BRCA [13–15]. Additionally, miRNAs play a vital role in the regulation network. miRNA is a class of single-stranded, endogenous, small, evolutionarily conserved non-coding RNAs (ncRNAs). Generally, miRNAs negatively regulated gene expression by sequence-specific base pairing with their target mRNAs [16]. However, the study of ceRNAs is still in its infancy and the hypothesis for ceRNA is still controversial [17]. The relationship between miRNAs and diseases was a comprehensive study. Since the regulation between miRNA and mRNA was comprehensively studied [18–21]. With the in-depth study of miRNAs, it has been found that miRNAs in the exosomes play an important role in the pathogenesis of diseases [22, 23]. Therefore, it is necessary to study miRNAs in different BRCA subtypes. Therefore, this study is mainly focused on the regulatory network between miRNAs and mRNAs hoping can provide more reference for clinical practice.

Studying the differences among the four subtypes of BRCA is of great significance for the understanding of the differences of regulatory networks among different subtypes [24–26]. However, whether the same gene or the same regulatory network shares the same topological pattern among the four subtypes is rarely reported. Therefore, we proposed a hypothesis in this work for which the same gene has different networks among the four subtypes. In order to verify this, differentially expressed genes were analysed between each of the four subtypes and the normal tissue sample. The overlapping genes between the four groups of differentially expressed genes (DEGs) were then identified to construct the regulatory network for each of the four subtypes. For each subtype, different networks resulting from the same group of genes can reflect not only the heterogeneity of BRCA at the level of gene expression pattern differences but also the heterogeneity in regulatory relationships. Therefore, we focused mainly on revealing the heterogeneity in regulatory networks of miRNA–mRNA pairs.

2 Methods

2.1 Data collection

All BRCA data were collected from the UCSC cancer browser (https://xenabrowser.net/datapages/) database. The cohort of BRCA patients was obtained from the Phenotype dataset from the GDC TCGA BRCA within the database. The dataset Phenotype contains 1283 samples alongside with 187 identifiers.

2.2 Data preprocessing

Only the samples which include mRNAs, miRNAs, and clinical data were retained. Basal, HER2, LumA, and LumB were filtered according to the gene expression and clinical information, besides, normal samples were also collected from the dataset. Samples with

Table 1 Four subtypes of BRCA in the TCGA database

Subtypes	HR	P value	C-index
basal-like	3.767 (1.353–10.49)	0.011	0.645
HER2	3.671(1.011–13.35)	0.045	0.646
LumA	2.719(1.555–4.758)	<0.001	0.639
LumB	1.645(0.771–3.51)	0.198	0.611

survival time <30 days were excluded. Therefore, the sample size after the screening is different from the original GDC BRCA data. This resulted in 142 Basal, 62 HER2, 437 LumA, 194 LumB, and 113 Normal samples. Since we were only interested in gene expressions of mRNAs and miRNAs, the biomart tool of the Ensembl database (http://asia.ensembl.org/index.html) was used for further filtering of the data. 24,491 mRNAs were retained from a total number of 60,484 RNAs. Next, we only kept those mRNAs with expression values >1. As a result, 20,058 mRNAs and 538 miRNAs were retained for further study.

2.3 Differential gene expression analysis

For identification of the DEGs, an R package called 'limma' and R software (version 3.5.0) was applied to analyse DEGs in different groups. Limma is used for the analysis of gene expression microarray data, especially the use of linear models for analysis designed experiments and the assessment of differential expression. RNAs (miRNAs and mRNAs) with $|\log_2 FC| \geq 1$ and adjust P-value (false discover rate) <0.05 were considered to be significantly differentially expressed. We analysed DEGs between each subtype and the Normal samples. In addition, we also analysed DEGs among the four subtypes.

2.4 Identification of miRNA–mRNA networks in four subtypes

Different subtypes would exhibit different regulation networks. It is important to find driver genes and core genes in four subtypes. The overlap of DEGs in miRNAs and mRNAs was identified by the Venn method which is implemented as R package named 'VennDiagram'. The miRWalk2.0 (http://zmf.umm.uniheidelberg.de/apps/zmf/mirwalk2/) database, which is so far the only freely accessible database, which is supplying the biggest available collection of predicted and experimentally verified miRNA–target interactions, was used in the current study for identification of the targets of the differentially expressed miRNAs.

The overlapped miRNAs and mRNAs between the four subtypes were used to construct a regulation network. From the previously mentioned miRNA–target database, all miRNA–target interactions between the miRNAs and the mRNAs were identified. The correlations between miRNAs and mRNAs in different subtypes may differ. Therefore, Pearson correlation coefficients were calculated, and miRNA–mRNA pair with a correlation coefficient >0.5 and <−0.5 was considered to be significant. The resulted sub-networks would be different among the four subtypes.

2.5 Independent dataset validation

The Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/pubmed/geo) was employed to be the validation dataset. An online web system called KM-plotter was used to assess the effect of miRNAs or mRNAs on survival. The system provides a total number of 5143 BRCA samples and the ability to select all kinds of subtypes of BRCA and therapy types. In the current study, Basal and LumB subtypes were chosen to verify our result.

An online web tool called KM-plotter, which can be used to assess the relevance of the expression levels of various genes on the clinical outcome both in untreated and treated BRCA patients, was used to assess the effect of the genes on survival [27].

2.6 Prognostic driver gene analysis

The 'survival' and 'survminer' packages in R software were used for the analysis of the prognosis in the four subtypes. All driver genes were analyzed by the univariate Cox regression. Prognostic genes with P-value < 0.05 were considered to be significantly associated with survival. Multiple Cox regression analysis (MCA) was used to test the independent factor which was significantly factor in univariate Cox regression. Prognostic index (PI) of driver genes was calculated by a linear combination of the gene expressions and the coefficients of Cox regression. PI is a prognostic index vector and the jth element of PI is the prognostic index of the jth patient, i.e.

$$\mathrm{PI}_j = \sum_i \beta_i \times G_{ij},$$

where β_i is the regression coefficient of the ith variable (in this context, the ith gene/miRNA) estimated by the MCA; G_{ij} is the observed value of the ith variable in the jth sample (in this context, the expression value of the ith gene/miRNA in the jth sample).

The hazard ratio (HR) was calculated from $\exp(\beta)$, and β was the coefficient from Cox regression. In this study, P-value ≤ 0.05 was considered to be significant for the log-rank test. The value of the concordance index (C-index) >0.6 was considered as the good performance of a model. HR and 95% confidence interval (CI) were calculated to identify low-risk (HR <1) or high-risk gene signature model (HR > 1). Kaplan–Meier curve was employed to estimate the differences between the high- and low-risk patients.

3 Result

3.1 Subtypes of BRCA

Four subtypes of BRCA were correlated with the overall survival (OS) of BRCA patients by Kaplan–Meier survival analysis (Supplementary Fig. S1). Each of the four subtypes of Cox regression analysis is listed in Table 1.

3.2 Gene expression analysis

The DEGs between each of the four subtypes and the normal one was analysed. The overlapping genes between these four groups of DEGs were then identified. There resulted in 102 mRNAs and 31 miRNAs, the detailed information of these RNAs is listed in Supplementary Figs. S2 and S3.

3.3 Gene ontology (GO) enrichment analysis

GO enrichment was employed in R package 'ClusterProfiler' [28] for displaying biological process, molecular function and cellular component of DEGs. The results are shown in Supplementary Fig. S4.

The four groups of DEGs resulted from a comparison between each of the four subtype samples and the normal sample (as mentioned above) was analysed by a web tool named Metascape (http://metascape.org). The result is shown in Fig. 1 with biological processes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways correspond with miRNAs and mRNAs, respectively.

The result of GO enrichment shows that (i) the four subtypes shared three common pathways and biological processes (gene silencing by miRNA in cancer and negative regulation of angiogenesis); (ii) the differentially expressed miRNAs for the Basal sample included more biological processes; and (iii) four subtypes share similar gene functions.

3.4 Overlapping RNAs between the four subtypes

For analysis of the intrinsic properties of the four subtypes, the overlapped genes among the four groups of DEGs were identified, and the process is visualised by the Venn method as shown in Fig. 2.

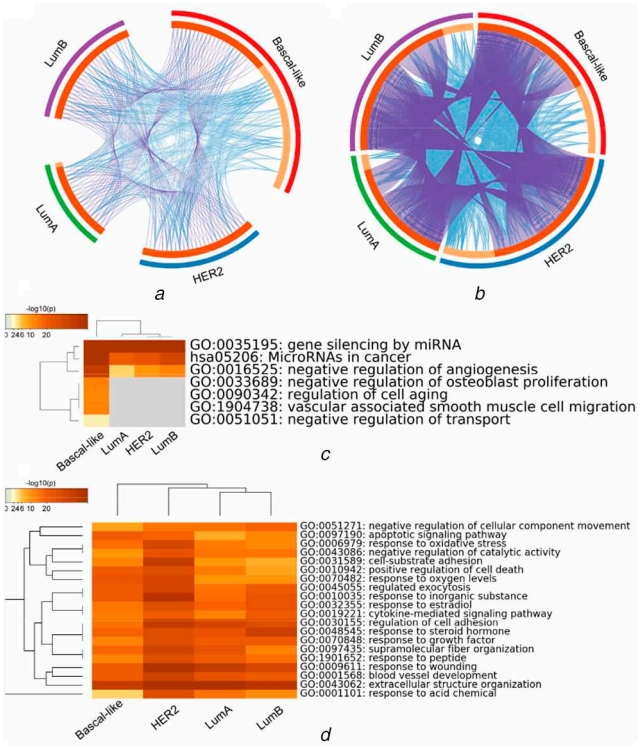

Fig. 1 *GO enrichment of miRNAs and mRNAs*

(a) Overlap of functions in the differentially expressed miRNAs for the four subtypes, ***(b)*** Overlap of functions in the differentially expressed mRNAs for the four subtypes, ***(c)*** GO and KEGG enrichment of the differentially expressed miRNAs for each of the four subtypes, ***(d)*** GO and KEGG enrichment of the differentially expressed mRNAs for each of the four subtypes

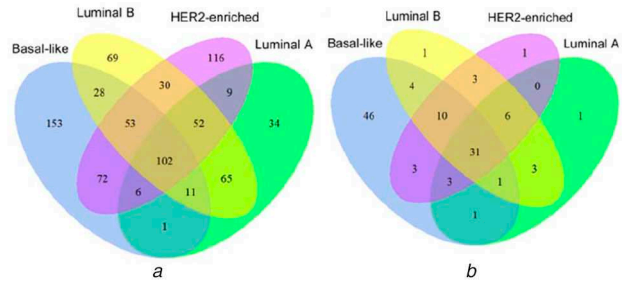

Fig. 2 *Venn diagrams of the four groups of DEGs*

(a) Overlapping mRNAs among the four groups of DEGs, ***(b)*** Overlapping miRNAs among the four groups of DEGs

For further analysis of the difference between the four subtypes, the DEGs were analysed between the four subtypes (Supplementary Fig. S5). The result shows that the greatest difference was found to be between the Basal and LumA samples. Generally, miRNAs negatively regulate the expression of RNAs, so the expression of both of them shows the opposite trend. Moreover, the results show that the proportion of high expression of miRNAs was higher and that of low expression of mRNA was higher.

3.5 Network construction and analysis

Four networks were constructed from the overlapped DEGs (102 mRNAs and 31 miRNAs) for each of the four subtypes. The miRNA–mRNA relationships were collected from the same database, and the correlation between miRNA and mRNA is quite different between the four subtypes. The networks of the four subtypes are shown in Fig. 3.

3.6 Special sub-network analysis

We extracted special networks based on retaining those gene pairs with correlation coefficients >0.5 and <−0.5 (as mentioned in the

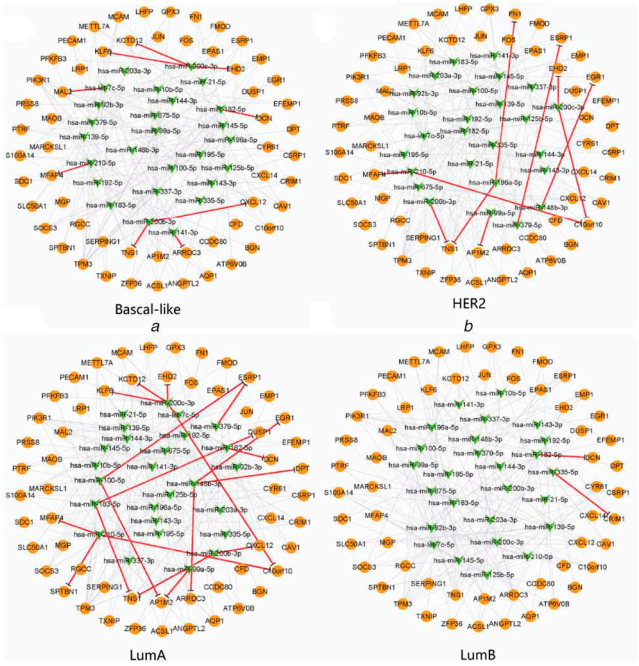

Fig. 3 *Regulation networks for each of the four subtypes*

(a) Regulation network for Basal subtype, ***(b)*** Regulation network for HER2 subtype, ***(c)*** Regulation network for LumA subtype, ***(d)*** Regulation network for LumB subtype

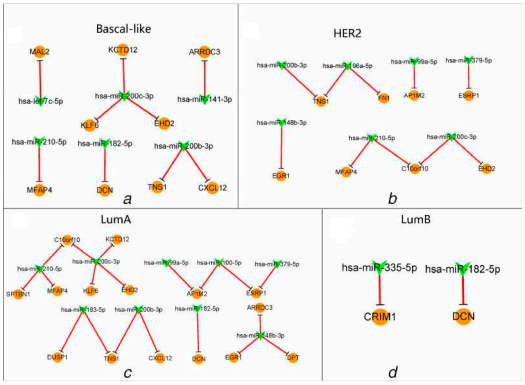

Fig. 4 *Sub-networks constructed by filtering based on correlation coefficients*

(a) Sub-network of Basal subtype, ***(b)*** Sub-network of HER2 subtype, ***(c)*** Sub-network of LumA subtype, ***(d)*** Sub-network of LumB subtype

method section) from each of the four networks. The resulting sub-networks are shown in Fig. 4.

3.7 Prognostic miRNA–mRNA network analysis

For testing of the prognostic genes in the four sub-networks, step multivariate Cox regression was applied to analyse HR for each RNA. Cox regression with Akaike information criterion for model selection was applied on genes in each of the four subtypes. The result is shown in Table 2.

The survival analysis result of special genes in Basal and HER2 showed that PI consisting of these special genes that can effectively predict high-risk and low-risk patients (P-value < 0.001) (Figs. 5*a* and *d*). The distribution of PI value shown in Figs. 5*b* and *e*. Receiver operating characteristic curve (ROC) curve shows that special miRNAs and mRNAs can effectively predict 10 years of survival area under curve ((AUC) > 0.8) (Figs. 5*c* and *f*).

The results of the survival analysis in LumA and LumB are shown in Fig. 6. The special regulation pairs of miRNAs and mRNA in LumA showed a significant difference between high- and low-risk patients (Fig. 6*a*). The distribution of PI in LumA and LumB is shown in Figs. 6*b* and *e*. The ROC curve of special genes in LumA showed good performance in predicting 10 years. In

Table 2 List of prognostic genes in each subtype

RNAs	Gene symbol	Description	HR	95% CI	*P*-value
Basal-like	hsa-miR-182–5p		1.79	1.10–2.92	0.02
	DCN	decorin	1.73	1.07–2.82	0.03
	ARRDC3	arrestin domain-containing protein 3	0.47	0.22–1.04	0.06
HER2	*EGR1*	early growth response protein 1	0.55	0.31–0.97	0.04
	ESRP1	epithelial splicing regulatory protein 1	2.17	0.87–5.45	0.08
	FN1	fibronectin	0.58	0.36–0.94	0.03
LumA	hsa-miR-210–5p		0.70	0.50–0.98	0.04
	DPT	dermatopontin	1.37	1.03–1.82	0.03
	ESRP1	epithelial splicing regulatory protein 1	1.58	0.97–2.57	0.06
	EHD2	EH domain-containing protein 2	0.50	0.29–0.86	0.01
LumB	hsa-miR-355–5p		0.87	0.65–1.16	0.09
	hsa-miR-182–5p		1.08	0.74–1.57	0.70
	CRIM1	cysteine-rich motor neuron 1 protein	1.02	0.66–1.57	0.91
	DCN	decorin	1.25	090–1.75	0.20

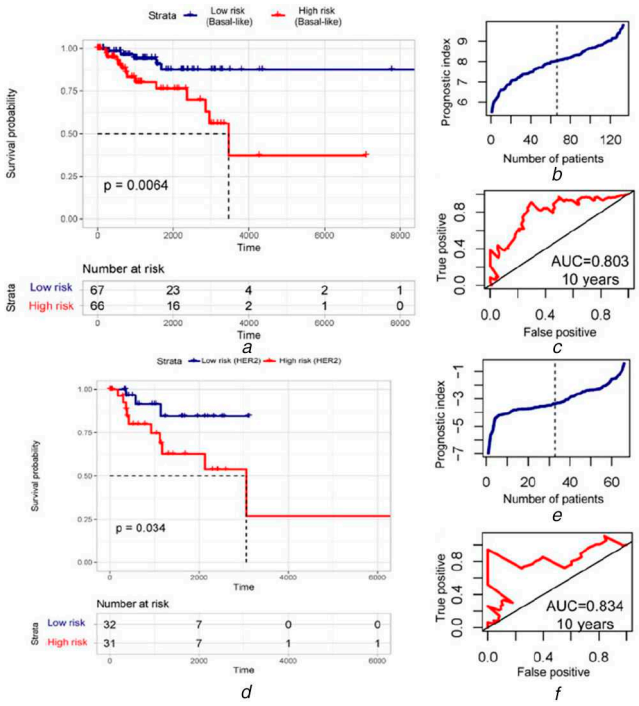

Fig. 5 *Kaplan–Meier curve for the validation of genes in each of the two sub-networks of Basal and HER2 subtypes*

(a) Kaplan–Meier survival curve of high-risk and low-risk groups classified by genes in special sub-network of Basal-like, ***(b)*** PI distribution in patients with Basal-like, ***(c)*** AUC of ROC for predicting PI of OS in Basal-like subtype in 10 years, ***(d)*** Kaplan–Meier survival curve of high-risk and low-risk groups classified by genes in special sub-network of HER2 in 10 years, ***(e)*** PI distribution in patients with HER2, ***(f)*** AUC of ROC for predicting PI of OS in HER2 subtype

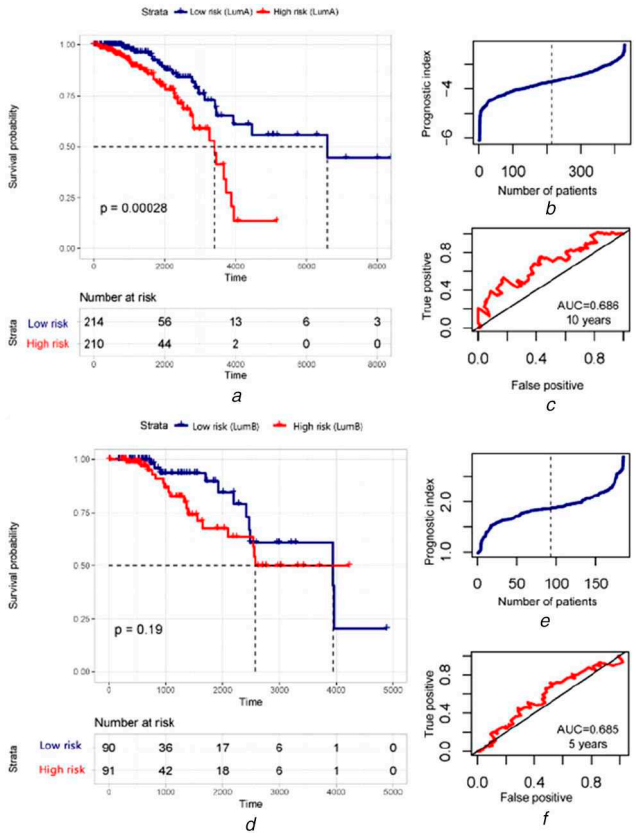

Fig. 6 *Kaplan–Meier curve for validation genes in sub-network of LumA and LumB subtypes*

(a) Kaplan–Meier survival curve of high-risk and low-risk groups classified by genes in special sub-network of LumA, ***(b)*** PI distribution in patients with LumA, ***(c)*** AUC of ROC for predicting PI of OS in LumA subtype in 10 years, ***(d)*** Kaplan–Meier survival curve of high-risk and low-risk groups classified by genes in special sub-network of LumB, ***(e)*** PI distribution in patients with LumB, ***(f)*** AUC of ROC for predicting PI of OS in LumB subtype in 5 years

LumB, the special genes could not classify patients significantly. However, the ROC curve showed that the PI of biomarkers in LumB can predict 5 years of survival (Fig. 6*f*).

3.8 Validating biomarkers in the external dataset and cell lines

For validation of the result of the TCGA datasets, a pair of representative regulation relationships (hsa-miR-182–5p and decorin (*DCN*)), which appeared in Basal and LumB was selected for analysis. As shown in Fig. 7, the regulation of hsa-miR-182–5p and *DCN* is significantly associated with overall survival in Basal subtype, but not in the LumB subtype. The result also shows that hsa-miR-182–5p and *DCN* were closely associated with the overall survival of Basal (log-rank P-value < 0.1). However, hsa-miR-182–5p and *DCN* were not associated with the overall survival of LumB.

The verification of independent datasets (GEO database in KM-plotter) basically coincides with the results of training data sets (TCGA database). The validation results showed that the same RNAs might play different roles in different subtypes.

In order to further verify that heterogeneity of the gene signatures in the four subtypes, we compared their expression among all subtypes (Fig. 8). Although all of the expression values of the miRNAs and mRNAs are up-regulated compared with the normal samples, these RNAs exhibit different expression in each subtype.

From Fig. 8, the results showed that these special prognostic genes of each subtype exhibit different expression in different

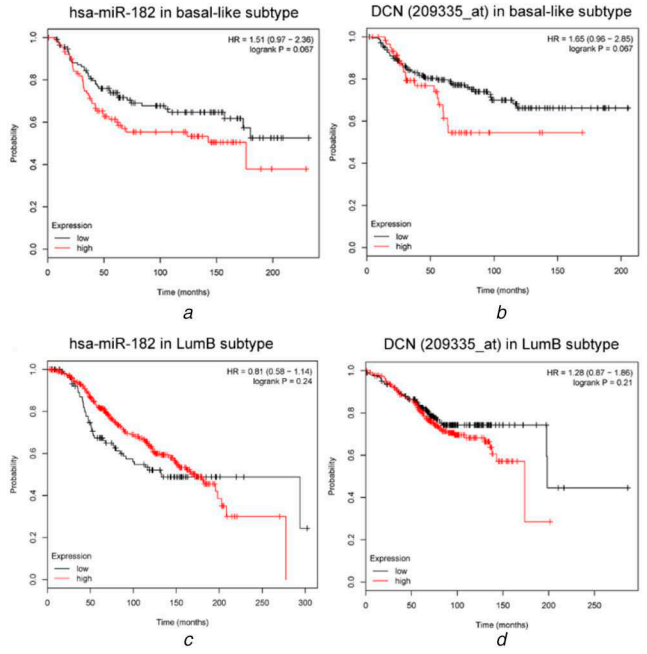

Fig. 7 *Validation of hsa-miR182 and DCN in KM-plotter dataset*
(a) Kaplan–Meier curve of hsa-miR-182 expression in Basal subtype, ***(b)*** Kaplan–Meier curve of DCN expression in Basal subtype, ***(c)*** Kaplan–Meier curve of hsa-miR-182 expression in LumB subtype, ***(d)*** Kaplan–Meier curve of DCN expression in LumB subtype

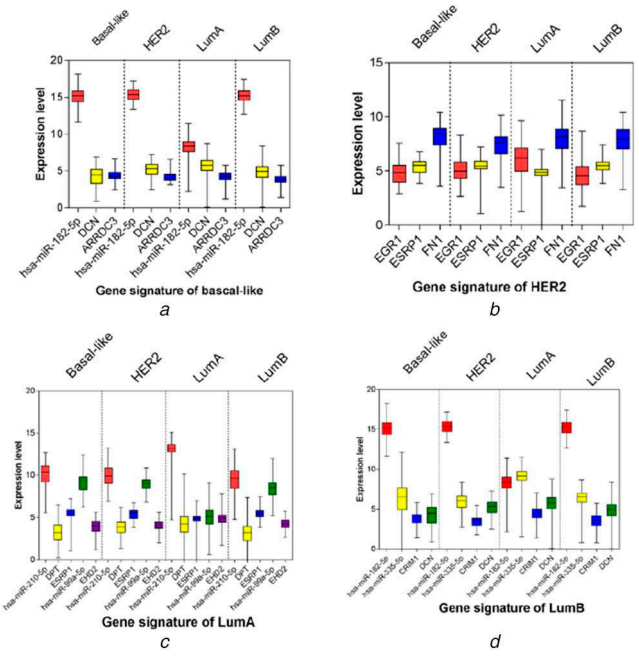

Fig. 8 *Cross-validation of special prognostic genes in each subtype*
(a) Expression of Basal prognostic genes in four subtypes, ***(b)*** Expression of HER2 prognostic genes in four subtypes, ***(c)*** Expression of LumA prognostic genes in four subtypes, ***(d)*** Expression of LumB prognostic genes in four subtypes

subtypes. Some of them show different expression among different subtypes, and some of them show no significant difference among different subtypes. For example, hsa-miR-182–5p exhibits high expression in Basal subtype and low expression in LumA, the prognostic genes (*EGE1*, *ESRP1*, and *FN1*) of HER2 exhibit similar expression in all of the four subtypes.

Additionally, hsa-miR-182–5p and *DCN* can significantly classify Basal-like and LumB subtypes. Thus, logistics regression [29] applied to test the performance of these two biomarkers. The results showed that the AUC of the ROC curve was 0.62, which represents good performance (Fig. 9).

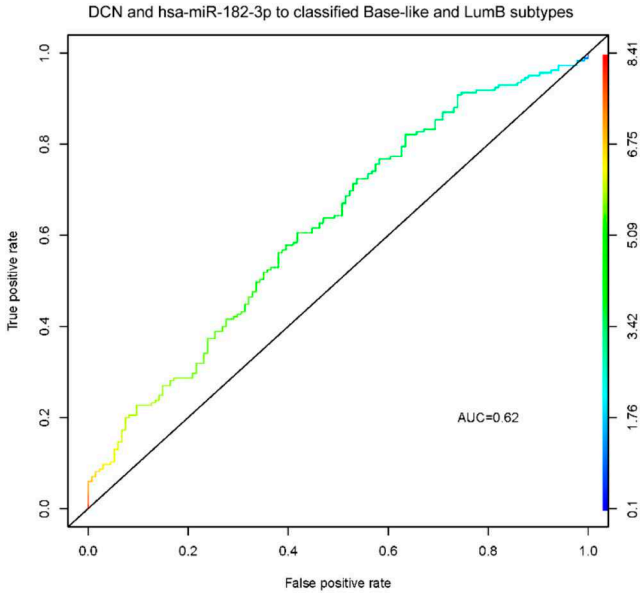

Fig. 9 *ROC curve validated the performance of biomarkers in Basal-like and LumB subtype*

ATCC cell lines database (https://www.atcc.org/en/Products/Cells_and_Microorganisms/Cell_Lines.aspx) was employed to classify BRCA cell lines into four subtypes. Secondly, we used GSE40057 dataset to analyse biomarker that was analysed from our results. External BRCA cell lines are very complicated. Each subtype of BRCA included many cell lines. The cell lines corresponding to each subtype have different characteristics. For example, LumB ER+, PR+/–, HER2 + Ki67 high, usually endocrine responsive, variable to chemotherapy. HER2+ in LumB are BT474, ZR-75 trastusumab responsive. In this study, we used a biomarker of DCN expression to validate in LumB and Basal-like subtypes. The results are shown in Fig. 10. Also, the result shows that DCN expression is significantly different between LumB (BT474 and ZR-75–1) and one of the types of Basal-like (Hs578 T). However, the expression of miRNAs was not sufficient to perform analysis. Thus, this study used an expression of mRNA as external data to validate.

4 Discussion

The heterogeneity of BRCA in molecular and cellular levels and the large number of genes potentially involved in cell growth, death and differentiation suggest the importance of studying multiple genetic alterations in concert [30]. Several research studies have suggested that gene expression-based signature can be used to predict prognosis and it outperformed the risk assessment based solely on clinicopathologic factors [31, 32]. Also, many previous publications have focused on the expression pattern of coding or non-coding genes in order to obtain some biomarkers for each subtype [33, 34]. In fact, the complexity of BRCA is not only reflected in the gene expression of the four subtypes, but also in the complication of regulation within coding genes, non-coding genes, and proteins. Generally, previous works have mainly focused on the differences in gene expressions and regulatory networks between different subtypes [23, 24, 35, 36].

In the present study, we focused on common genes that are differentially expressed between the four subtypes and normal tissues. The expression of the selected genes in every tumour subtype is significantly higher (P-value < 0.01) than that of the normal breast tissues. We analysed the regulation network of these genes, the result shows that the same genes have completely different regulatory networks in different subtypes (Fig. 3). The complexity of BRCA is not only reflected in the differential expression of individual genes, but also in the different regulation modes among the same genes. Based on the miRNA–target interaction database and the negative correlation expression level between miRNAs and mRNAs, the special network was constructed for each subtype (Table 3).

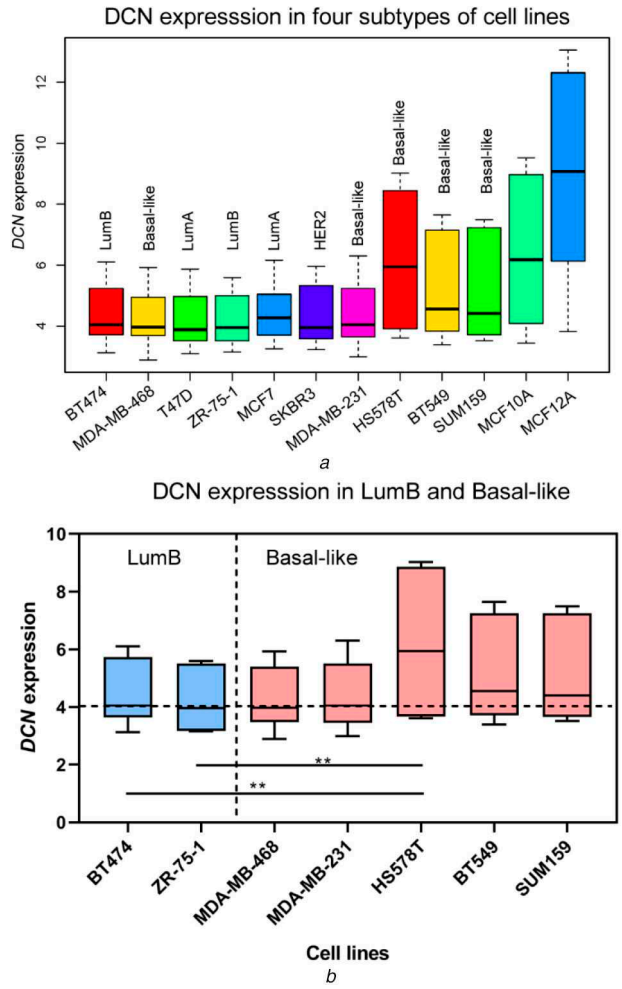

Fig. 10 *External cell line dataset of DCN expression in LumB and Basal-like subtypes*

(a) DCN expression in four subtypes cell lines, ***(b)*** DCN expression validation between LumB and Basal-like subtypes

Table 3 PI of a gene signature in four subtypes

Factors	HR (95%CI)	C-index	*P*-value	Log-rank test
$PI_{Bascal-like}$	3.77(1.35–10.49)	0.65	0.01	0.01
PI_{HER2}	3.67(1.01–13.35)	0.65	0.04	0.03
PI_{LumA}	2.72(1.56–4.76)	0.64	5×10^{-4}	3×10^{-4}
PI_{LumB}	1.65(0.77–3.51)	0.61	0.20	0.19

We extracted genes that contain regulatory networks in four subtypes and calculated their HRs in each subtype. The genes in the Basal, HER2 and LumA subtypes are significantly associated with survival. The gene signature of a special regulation in LumB was not associated with survival. Although these genes in LumB do not predict overall survival, they can be found to effectively predict the survival of patients with this subtype in 5 years (Figs. 6*d* and *f*).

Obtaining special miRNA–mRNA networks from common genes is useful for the understanding of the BRCA subtypes. In the present study, we mainly study the relationship in each subtype. Although hsa-miR-182–5p and *DCN* have been reported many times as noted biomarkers in BRCA [37–39], the regulation pair between the two of them were few reported in the Basal subtype of BRCA. In the present study, regulation of hsa-miR-182–5P and *DCN* shows that they can predict overall survival in the Basal subtype, but not in LumB. The same trend is reflected in the result of validation data (Fig. 7). The results also showed that the hsa-miR-200 family (including hsa-miR-200a/200b/200c/141/429) plays a vital role in BRCA (Fig. 4). Also, the results showed that the miR-200 family plays a different role in different subtypes. Hsa-miR-141–3p is one of the members of the hsa-miR-200 family. Hsa-miR-141–3p and *ARRDC3* regulatory pair only in the Basal subtype. Also, *ARRDC3* has been reported to contribute to the aggressiveness of the basal-like subtype [40], and this gene was also identified in this study. We expect to find specific regulatory networks and genes in each subtype from the same gene set. The results make it clearer that the same genes play different roles in different subtypes.

5 Acknowledgments

This work was supported by the National Nature Science Foundation of China (61903306) and the China Postdoctoral Science Foundation (2018M633606).

6 References

[1] Roman, R., Perou, C.M., Fraser, S.W., *et al.*: 'Breast cancer molecular subtypes respond differently to preoperative chemotherapy', *Clin. Cancer Res.*, 2005, **11**, pp. 5678–5685

[2] Haughian, J.M., Pinto, M.P., Chuck, H.J., *et al.*: 'Maintenance of hormone responsiveness in luminal breast cancers by suppression of notch', *Proc. Natl. Acad. Sci. U.S.A.*, 2012, **109**, pp. 2742–2747

[3] Sircoulomb, F., Bekhouche, I., Finetti, P., *et al.*: 'Genome profiling of ERBB2-amplified breast cancers', *BMC Cancer*, 2010, **10**, p. 539

[4] Brewster, A.M., Mariana, C.M., Powel, B.: 'Epidemiology, biology, and treatment of triple-negative breast cancer in women of African ancestry', *Lancet Oncol.*, 2014, **15**, pp. e625–e634

[5] Salmena, L., Poliseno, L., Tay, Y., *et al.*: 'A ceRNA hypothesis: the Rosetta stone of a hidden RNA language?', *Cell*, 2011, **146**, pp. 353–358

[6] Lehmann, B.D., Bauer, J.A., Chen, X., *et al.*: 'Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies', *J. Clin. Invest.*, 2011, **121**, (7), pp. 2750–2767

[7] Chuthapisith, S., Permsapaya, W., Warnnissorn, M., *et al.*: 'Breast cancer subtypes identified by the ER, PR and HER-2 status in Thai women', *Asian Pac. J. Cancer Prev.*, 2012, **13**, (2), pp. 459–462

[8] Pan, X.Y., Hu, X.H., Zhang, Y.H., *et al.*: 'Identification of the copy number variant biomarkers for breast cancer subtypes', *Mol. Genet. Genomics*, 2019, **294**, (1), pp. 95–110

[9] Mallik, S., Zhao, Z.: 'Graph-and rule-based learning algorithms: a comprehensive review of their applications for cancer type classification and prognosis using genomic data', *Brief. Bioinform.*, 2019, doi: 10.1093/bib/bby120

[10] Mallik, S., Maulik, U.: 'MiRNA-TF-gene network analysis through ranking of biomolecules for multi-informative uterine leiomyoma dataset', *J. Biomed. Inf.*, 2015, **57**, pp. 308–319

[11] Bhadra, T., Mallik, S., Bandyopadhyay, S.: 'Identification of multiview gene modules using mutual information-based hypograph mining', *IEEE Trans. Syst. Man Cybern., Syst.*, 2017, **49**, (6), pp. 1119–1130

[12] Cai, M., Li, L.: 'Subtype identification from heterogeneous TCGA datasets on a genomic scale by multi-view clustering with enhanced consensus', *BMC Med. Genet.*, 2017, **10**, (Suppl 4), p. 75

[13] Zhou, S., Wang, L., Yang, Q., *et al.*: 'Systematical analysis of lncRNA–mRNA competing endogenous RNA network in breast cancer subtypes', *Breast Cancer Res. Treat.*, 2018, https://doi.org/10.1007/s10549-018-4678-1

[14] Conte, F., Fiscon, G., Chiara, M., *et al.*: 'Role of the long non-coding RNA PVT1 in the dysregulation of the ceRNA–ceRNA network in human breast cancer', *PLOS ONE*, 2017, **12**, p. e0171661

[15] Zhou, X., Liu, J., Wang, W.: 'Construction and investigation of breast-cancer-specific ceRNA network based on the mRNA and miRNA expression data', *IET Syst. Biol.*, 2014, **8**, pp. 96–103

[16] Chen, X., Xie, D., Zhao, Q., *et al.*: 'MicroRNAs and complex diseases: from experimental results to computational models', *Brief. Bioinform.*, 2019, **20**, pp. 515–539

[17] Thomson, D.W., Dinger, M.E.: 'Endogenous microRNA sponges: evidence and controversy', *Nat. Rev. Genetics*, 2016, **17**, (5), pp. 272–283

[18] Chen, X., Wang, L., Qu, J., *et al.*: 'Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics'*, 2018, **34**, pp. 4256–4265

[19] Chen, X., Huang, L.: 'LRSSLMDA: Laplacian regularized sparse subspace learning for MiRNA-disease association prediction', *PLOS Comput. Biol.*, 2017, **13**, p. e1005912

[20] Chen, X., Yin, J., Qu, J., *et al.*: 'MDHGI: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction', *PLOS Comput. Biol.*, 2018, **14**, p. e1006418

[21] Aqil, M., Naqvi, A.R., Mallik, S., *et al.*: 'The HIV Nef protein modulates cellular and exosomal miRNA profiles in human monocytic cells', *J. Extracell. Vesicles*, 2014, **3**, (1), p. 23129

[22] Aqil, M., Mallik, S., Bandyopadhyay, S., *et al.*: 'Transcriptomic analysis of mRNAs in human monocytic cells expressing the HIV-1 Nef protein and their exosomes', *BioMed Res. Int.*, 2015, **2015**, p. 492395

[23] De, A.-J.U.G., Velázquez-Caldelas, T.E., Espinal-Enríquez, J., *et al.*: 'Transcriptional network architecture of breast cancer molecular subtypes', *Front. Physiol.*, 2016, **7**, p. 568

[24] Alcalá-Corona, S.A., Anda-Jáuregui, G.D., Espinal-Enríquez, J., *et al.*: 'Network modularity in breast cancer molecular subtypes', *Front. Physiol.*, 2017, **8**, p. 915

[25] Nasser, S.: 'Context-specific gene regulatory networks subdivide intrinsic subtypes of breast cancer', *BMC Bioinf.*, 2011, **12**, pp. 3–12
[26] Györffy, B., Lanczky, A., Eklund, A.C., *et al.*: 'An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients', *Breast Cancer Res. Treat.*, 2010, **123**, pp. 725–731
[27] Mallik, S., Zhao, Z.: 'Identification of gene signatures from RNA-seq data using pareto-optimal cluster algorithm', *BMC Syst. Biol.*, 2018, **12**, (Suppl 8), p. 126
[28] Yu, G., Wang, L.G., Han, Y., *et al.*: 'Clusterprofiler: an R package for comparing biological themes among gene clusters', *OMICS*, 2012, **16**, (5), pp. 284–287
[29] Sørlie, T., Perou, C.M., Tibshirani, R.T., *et al.*: 'Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications', *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, pp. 10869–10874
[30] Harris, A., Glas, A.M., Floore, A., *et al.*: 'Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer', *J. Natl. Cancer Inst.*, 2006, **98**, pp. 1183–1192
[31] Nielsen, T.O., Parker, J.S., Leung, S., *et al.*: 'A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor–positive breast cancer', *Clin. Cancer Res.*, 2010, **16**, pp. 5222–5232
[32] Sørlie, T., Wang, Y., Xiao, C., *et al.*: 'Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms', *BMC Genomics*, 2006, **7**, pp. 127–127
[33] Zhang, S., Wang, J., Ghoshal, T., *et al.*: 'lncRNA gene signatures for prediction of breast cancer intrinsic subtypes and prognosis', *Genes*, 2018, **9**, (2), p. 65
[34] Qin, S., Ma, F., Chen, L.: 'Gene regulatory networks by transcription factors and microRNAs in breast cancer', *Bioinformatics*, 2015, **31**, pp. 76–83
[35] Xiao, B., Zhang, W., Chen, L., *et al.*: 'Analysis of the miRNA–mRNA–lncRNA network in human estrogen receptor-positive and estrogen receptor-negative breast cancer based on TCGA data', *Gene*, 2018, **658**, pp. 28–35
[36] Keerthana, K., Steptoe, A.L., Martin, H.C., *et al.*: 'MicroRNA-182-5p targets a network of genes involved in DNA repair', *RNA*, 2013, **19**, pp. 230–242
[37] Chang, Y.Y., Kuo, W.H., Hung, J.H., *et al.*: 'Deregulated microRNAs in triple-negative breast cancer revealed by deep sequencing', *Mol. Cancer*, 2015, **14**, p. 36
[38] Li, S.J., Chen, D.L., Zhang, W.B., *et al.*: 'Prognostic value of stromal decorin expression in patients with breast cancer: a meta-analysis', *J. Thorac. Dis.*, 2015, **7**, p. 1939
[39] Kelemen, L.E., Couch, F.J., Shahana, A., *et al.*: 'Genetic variation in stromal proteins decorin and lumican with breast cancer: investigations in two case-control studies', *Breast Cancer Res.*, 2008, **10**, (6), pp. 1–11
[40] Soung, Y.H., Pruitt, K., Chung, J.: 'Epigenetic silencing of ARRDC3 expression in basal-like breast cancer cells', *Sci. Rep.*, 2014, **4**, p. 3846