

RESEARCH

Open Access



In silico analysis of differentially expressed genesets in metastatic breast cancer identifies potential prognostic biomarkers

Jongchan Kim

Abstract

Background: Identification of specific biological functions, pathways, and appropriate prognostic biomarkers is essential to accurately predict the clinical outcomes of and apply efficient treatment for breast cancer patients.

Methods: To search for metastatic breast cancer-specific biological functions, pathways, and novel biomarkers in breast cancer, gene expression datasets of metastatic breast cancer were obtained from Oncomine, an online data mining platform. Over- and under-expressed genesets were collected and the differentially expressed genes were screened from four datasets with large sample sizes ($N > 200$). They were analyzed for gene ontology (GO), KEGG pathway, protein-protein interaction, and hub gene analyses using online bioinformatic tools (Enrichr, STRING, and Cytoscape) to find enriched functions and pathways in metastatic breast cancer. To identify novel prognostic biomarkers in breast cancer, differentially expressed genes were screened from the entire twelve datasets with any sample sizes and tested for expression correlation and survival analyses using online tools such as KM plotter and bc-GenExMiner.

Results: Compared to non-metastatic breast cancer, 193 and 144 genes were differentially over- and under-expressed in metastatic breast cancer, respectively, and they were significantly enriched in regulating cell death, epidermal growth factor receptor signaling, and membrane and cytoskeletal structures according to the GO analyses. In addition, genes involved in progesterone- and estrogen-related signalings were enriched according to KEGG pathway analyses. Hub genes were identified via protein-protein interaction network analysis. Moreover, four differentially over-expressed (*CCNA2*, *CENPN*, *DEPDC1*, and *TTK*) and three differentially under-expressed genes (*ABAT*, *LRIG1*, and *PGR*) were further identified as novel biomarker candidate genes from the entire twelve datasets. Over- and under-expressed biomarker candidate genes were positively and negatively correlated with the aggressive and metastatic nature of breast cancer and were associated with poor and good prognosis of breast cancer patients, respectively.

Conclusions: Transcriptome datasets of metastatic breast cancer obtained from Oncomine allow the identification of metastatic breast cancer-specific biological functions, pathways, and novel biomarkers to predict clinical outcomes of breast cancer patients. Further functional studies are needed to warrant validation of their roles as functional tumor-promoting or tumor-suppressing genes.

Keywords: Breast cancer, Metastatic breast cancer, Prognosis, Oncomine, Gene ontology, Biomarkers

Correspondence: jkimatsgu@sogang.ac.kr

Department of Life Sciences, Sogang University, 35 Baekbeom-ro, Mapo-gu, Seoul 04107, Republic of Korea



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

World Health Organization reports that breast cancer is the most frequent female malignancy (www.who.int). Although conventional therapeutic strategies, including surgery, radiotherapy and chemotherapy, targeted therapies, and more recently immunotherapies [1, 2] dramatically prolonged the survival of breast cancer patients, the incidence and mortality rates of some subtypes continuously increase in recent years and the trend even varies depending on the race, age, or region [3, 4]. Identification of novel biomarkers in breast cancer is critical for accurate prognosis analysis and therapeutic efficacy prediction.

Stage IV breast cancers, in particular, are detrimental metastatic breast cancers (MBCs). MBCs are rarely curative, so their 5-year survival rate (26%) is much lower than localized cancer (99%) [5, 6]. Recently [7–14] and in the past, numbers of bioinformatic analyses have been conducted to identify key differentially expressed genes and enriched biological pathways or to evaluate the expression of a few specific genes in breast cancers, but such analysis using transcriptomes of MBCs has not been satisfactorily performed. The identification of biological functions and pathways enriched in MBCs is pivotal to search for appropriate treatment options that would minimize the adverse effects and increase the survival rates of this fatal disease.

ONCOMINE is a cancer microarray database and web-based data-mining platform containing 729 available datasets with 91,866 samples as of December 17th, 2020 (www.oncomine.org/) [15]. I searched for gene expression datasets generated with MBC patient samples and screened differentially over- and under-expressed genes. With the genesets, I attempted to analyze biological functions and pathways enriched in MBCs, to identify novel biomarker candidate genes positively and negatively correlated with the aggressive and metastatic nature of breast cancer and to validate their prognostic values in breast cancer. To do so, I conducted gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, protein-protein interaction (PPI) network analysis, hub gene identification, co-expression analysis, and Kaplan-Meier survival analyses with available online tools.

Ultimately, these analyses demonstrate that the identified genes may serve as potential prognostic biomarkers that accurately predict the clinical outcomes of breast cancer patients. The results also provide therapeutic implications that might be beneficial for treating metastatic breast cancer patients. Furthermore, the present study recapitulates the usefulness of Oncomine platform in identifying appropriate key pathways and biomarkers to suggest therapeutic opportunities and accurately predict the clinical outcomes of breast cancer patients.

Methods

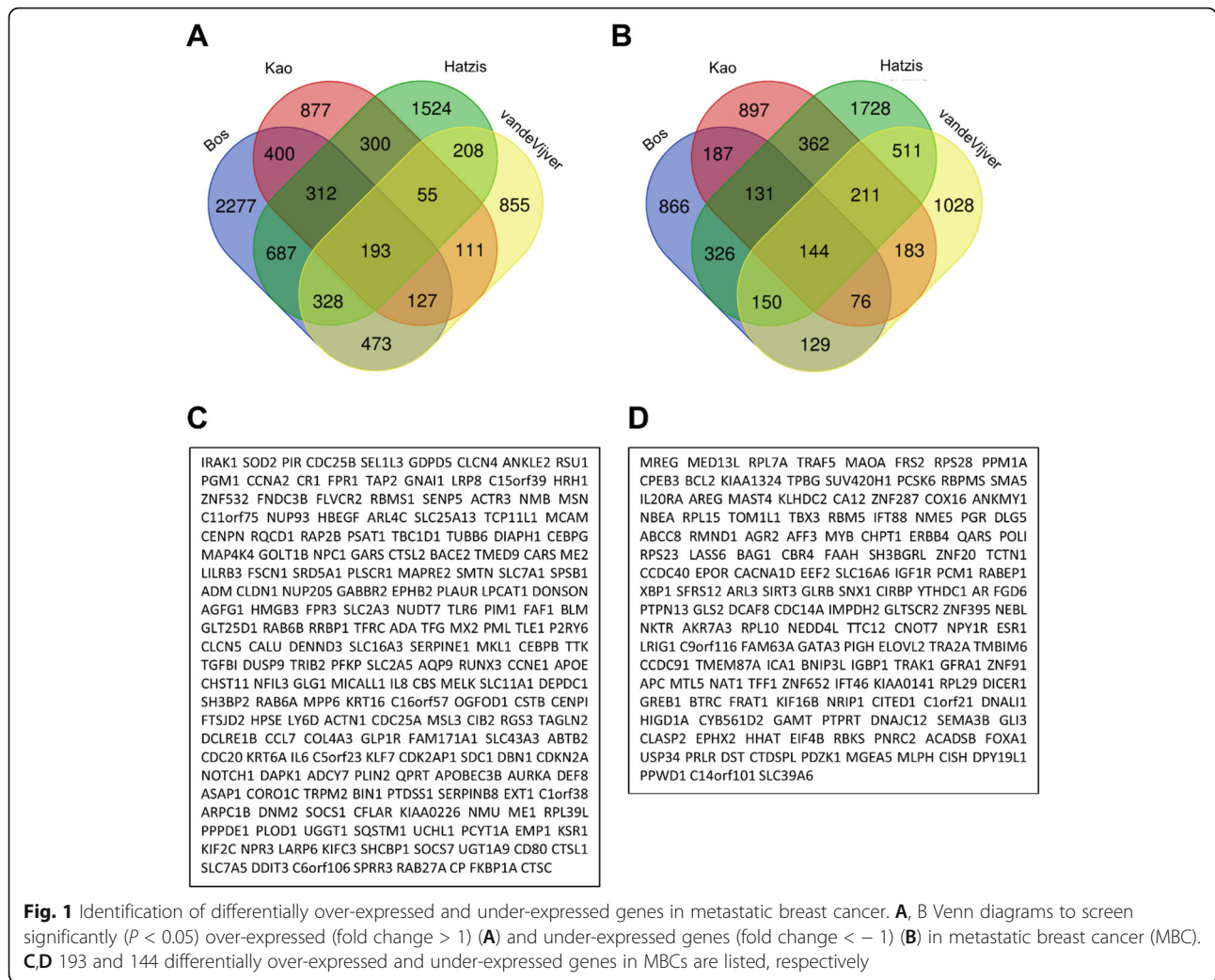
Dataset acquisition

To obtain microarray datasets, the publicly available Oncomine data-mining platform (<http://www.oncomine.org>) was analyzed. Datasets that profiled metastatic breast cancers (MBCs) were retrieved using filters including “breast cancer” (cancer type) and “metastatic event status at three years” (Clinical Outcome). A total of fourteen datasets were available under these filters and only transcriptome datasets were chosen (two genomic DNA studies were excluded): Bos Breast (N > 200), Desmedt Breast (N < 200), Hatzis Breast (N > 200), Kao Breast (N > 200), Loi Breast (N < 200), Loi Breast 3 (N < 200), Minn Breast 2 (N < 200), Schmidt Breast (N < 200), Symmans Breast 2 (N < 200), Symmans Breast (N < 200), vandeVijver Breast (N > 200), and Vantveer Breast (N < 200).

Determination of differentially expressed genesets

From four datasets with large sample sizes (N > 200), significantly over-expressed (fold change > 1) (DOE-L) or under-expressed (fold change < - 1) (DUE-L) genesets were selected based on their *P* values (*P* < 0.05) compared to the breast cancer patient samples with no metastatic events. 4797/2009 in Bos Breast, 3607/3564 in Hatzis Breast, 2375/2191 in Kao Breast, and 2350/2432 genes in vandeVijver Breast were significantly over-expressed/under-expressed, respectively. Using a Venn diagram drawing tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>), common genes were selected. In total, 193 and 144 genesets were differentially over-expressed (DOE-L) and under-expressed (DUE-L), respectively. These genesets were subjected to gene ontology, KEGG pathway, protein-protein interaction network analysis, and hub gene analyses to search for MBC-enriched genes, biological functions, and pathways.

To identify novel prognostic biomarkers, on the other hand, all twelve datasets with any sample sizes were analyzed. Differentially over-expressed (fold change > 1) or under-expressed (fold change < - 1) genesets with statistical significance (*P* < 0.05) were screened and examined. There was no single common gene found from all twelve datasets. However, four genes (*CCNA2*, *CENPN*, *TTK*, and *DEPDC1*) were differentially over-expressed (DOE-A) in eleven datasets (except Minn Breast 2) and one gene each was differentially under-expressed (DUE-A) in each of three groups of eleven datasets (the gene *ABAT* in the all twelve except Kao Breast, the gene *LRIG1* in the all twelve except Symmans Breast 2 and the gene *PGR* in the all twelve except Minn Breast 2).



Gene ontology (GO) and KEGG pathway analyses

Differentially expressed (DOE-L and DUE-L) genes obtained from four breast cancer datasets with large sample numbers ($N > 200$) were subjected to gene ontology (GO) and KEGG pathway analyses for functional and characteristic classification of enriched genes. To do so, 337 genes including 193 DOE-L and 144 DUE-L genesets were entered and analyzed at Enrichr (<https://amp.pharm.mssm.edu/Enrichr>), an online analysis tool. Genes were classified into three GO categories; Biological Process, Molecular Function, and Cellular Component. KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis for biological pathways was also conducted at Enrichr. The top ten GO terms and pathways were sorted according to their P values.

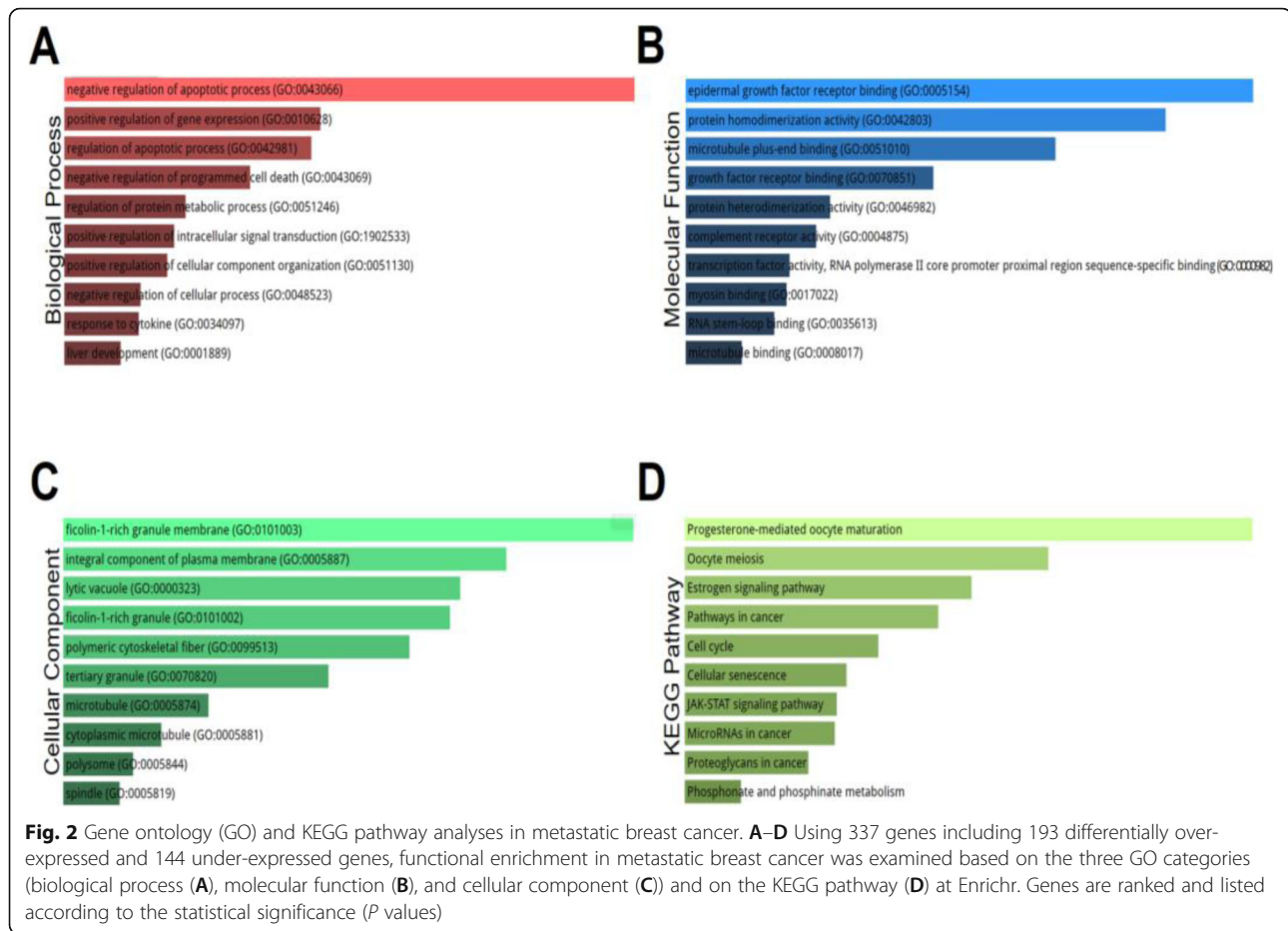
Protein-protein interactions (PPIs) and hub protein identification

To examine the protein-protein interaction network within the differentially expressed genesets, I utilized the

Search Tool for the Retrieval of Interacting Genes/Proteins (STRING). In total, 193 DOE-L and 144 DUE-L genesets were separately entered and their protein-protein interaction networks were analyzed. The networks were created, exported, and entered into Cytoscape, the network analysis/visualization tool to identify hub proteins from the complex networks. Among eleven “node ranking methods” [16], I analyzed the networks by Degree and both top ten hub proteins (hub_oe and hub_ue) were screened and ranked based on their number of interactors.

Comparison of biomarker candidate gene expression between basal-like/triple-negative and other subtypes of breast cancers

Two online RNA-seq databases (Cancer Cell Line Encyclopedia (CCLE) for human breast cancer cell lines and bc-GenExMiner (version 4.3) for breast cancer patient samples) were used to compare the expression levels of four DOE-A and three DUE-A genes between



basal-like/triple-negative and other subtypes of breast cancer. For CCLE, basal-like/triple-negative breast cancer (BL/TNBC) and luminal breast cancer cell lines were determined based on the literature [17–19]. For bc-GenExMiner, basal-like and TNBCs were determined by Prediction Analysis of Microarray 50 (PAM50) test and immunohistochemistry (IHC), respectively.

Kaplan-Meier survival analyses

Survival tests including relapse-free survival (RFS), overall survival (OS), distant metastasis-free survival (DMFS), and post-progression survival (PPS) were performed using KM plotter at <http://kmpplot.com> with Jetset best probe sets. MRFS was tested at <http://bcgenex.centregauducheau.fr> with all microarray datasets. All survivals were compared between the patients with high or low expression of each gene and the patient cohorts were split into two groups according to the median gene expression.

Protocol registration

The research protocol used in this study has been registered in PROSPERO database (registration #CRD42021247804).

Statistical analysis

Statistical analyses were performed according to the pre-set analytic methods of each online tool. Two-tailed, unpaired *t*-tests were performed for comparing gene expression with CCLE dataset analysis following grouping the breast cancer cell lines into either luminal or BL/TNBC. *P* < 0.05 was considered statistically significant.

Results

Identification of differentially expressed genesets in metastatic breast cancers.

We identified differentially over-expressed (DOE-L) and under-expressed (DUE-L) genes in metastatic breast cancer (MBC) by utilizing the OncoPrint database (Tables S1 and S2). A total of 193 DOE-L and 144 DUE-L genes were selected (Fig. 1) as described in Methods.

Functional and characteristic classification of enriched genes in metastatic breast cancer.

To analyze the functional enrichment of the differentially expressed genes in MBCs, I examined gene ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis using 337 differentially

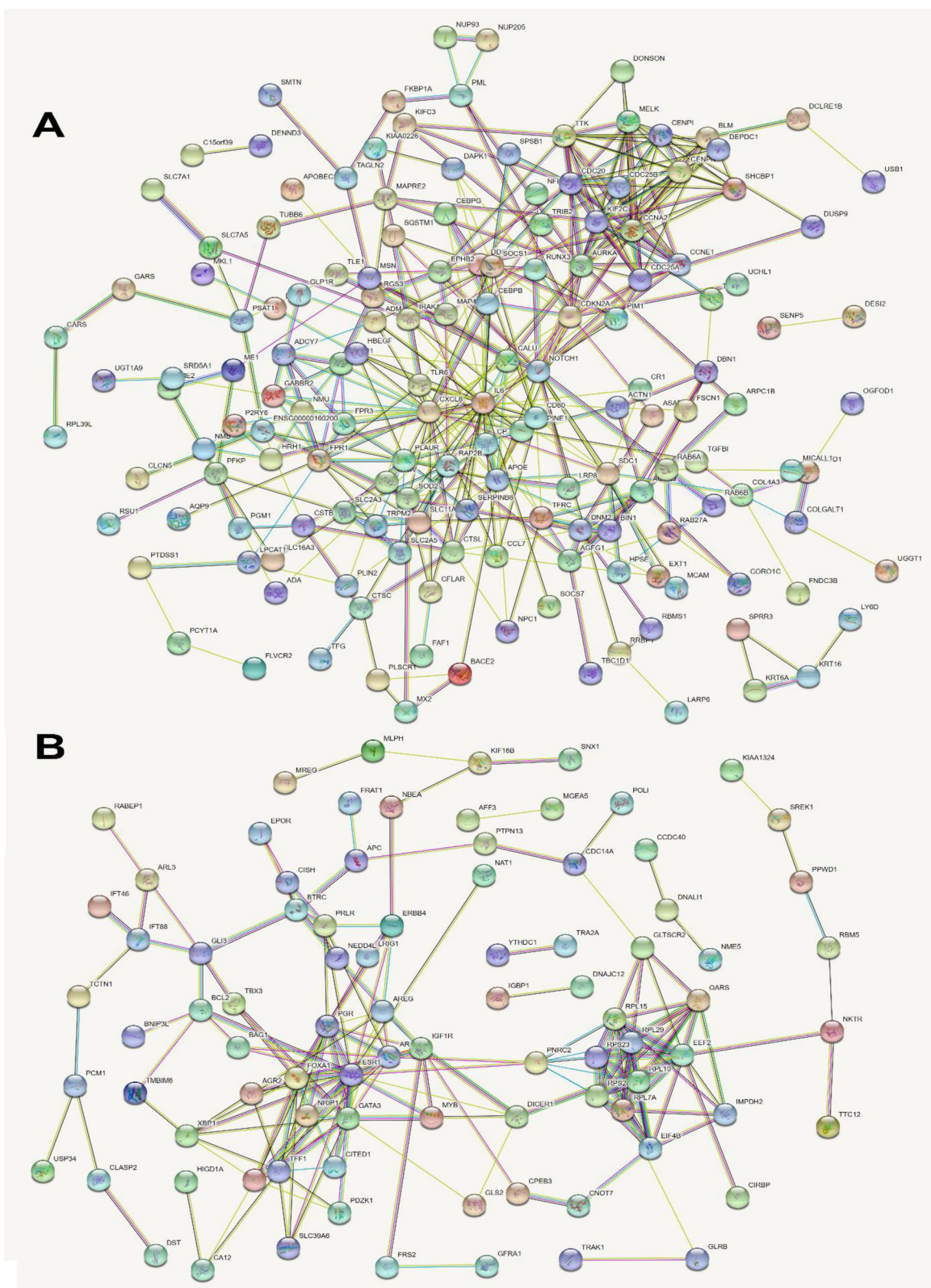


Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Protein-protein interaction networks of differentially expressed genes in metastatic breast cancer. **A** Protein-protein interaction (PPI) among 193 differentially over-expressed genes was examined to analyze the functional protein association network in metastatic breast cancer using STRING, an online tool. In total, 192 nodes (disconnected nodes are hidden) and 407 edges are presented. PPI enrichment P value is 1.11×10^{-16} , which implies this network has significantly more interactions than a network with a randomly chosen set of proteins. **B** Protein-protein interaction (PPI) among 144 differentially under-expressed genes was examined to analyze the functional protein association network in metastatic breast cancer using STRING. In total, 143 nodes (disconnected nodes are hidden) and 190 edges are presented. PPI enrichment P value is 6.99×10^{-14} , which implies this network has significantly more interactions than a network with a randomly chosen set of proteins. Circles and lines in **A** and **B** indicate genes and interactions, respectively. The line colors indicate the types of interaction evidence.

expressed genes (193 DOE-L and 144 DUE-L genes). They were classified into three GO categories including biological process (BP), molecular function (MF), and cellular component (CC). For BP, genes are significantly enriched in the GO terms including negative regulation of the apoptotic process, positive regulation of gene expression, regulation of the apoptotic process, negative regulation of programmed cell death, and regulation of protein metabolic process (Fig. 2A). For MF, genes are significantly enriched in the GO terms including epidermal growth factor receptor binding, protein homodimerization activity, microtubule plus-end binding, growth factor receptor binding, and protein heterodimerization activity (Fig. 2B). For CC, genes are significantly enriched in the GO terms including ficolin-1-rich granule membrane, an integral component of the plasma membrane, lytic vacuole, ficolin-1-rich granule, and polymeric cytoskeletal fiber (Fig. 2C). Therefore, these results suggest that genes regulating cell death, gene expression, protein metabolism, signal transduction, and protein-protein binding are significantly enriched in MBCs. Also, KEGG pathway analysis demonstrates that genes involved in progesterone-mediated oocyte maturation, oocyte meiosis, estrogen signaling pathway, pathways in cancer, and cell cycle are also significantly enriched in MBCs (Fig. 2D).

Interactome networks of the differentially expressed genes and identification of hub genes in metastatic breast cancer

Protein-protein interaction (PPI) provides insights into molecular function and diseases including cancer [20]. To explore PPI networks of the differentially expressed genes in MBCs, I utilized STRING, an online protein-protein interaction prediction tool, which visualizes potential interaction networks based on experimentally proven interaction data and computational prediction [21]. DOE-L (Fig. 3A) and DUE-L genes (Fig. 3B) were separately subjected to PPI analysis. In total, 192 nodes and 407 edges from DOE-L genes and 143 nodes and 190 edges from DUE-L genes were predicted after excluding disconnected nodes. Of note, their PPIs were predicted significantly more than those of a randomly chosen set of proteins.

To identify hub genes based on the PPI networks, I exported each network and examined them according to the degree of connectivity (DC) using Cytoscape software. In the DOE-L geneset, *IL6* (DC = 32), *CXCL8* (DC = 27), *AURKA/NOTCH1* (DC = 21), *CDC20/CCNA2/APOE* (DC = 17), *CDKN2A* (DC = 16), and *KIF2C/TTK* (DC = 15) were ranked as top ten hub genes (hub_oe) (Fig. 4A). In addition, *ESR1* (DC = 22), *FOXA1/GATA3* (DC = 14), *EEF2* (DC = 13), *RPL7A/TFF1* (DC = 12), *RPL15/AR/PGR*

A		B	
Rank	Node	Rank	Node
1	IL6	1	ESR1
2	CXCL8	2	FOXA1
3	AURKA	2	GATA3
3	NOTCH1	4	EEF2
5	CDC20	5	RPL7A
5	CCNA2	5	TFF1
5	APOE	7	RPL15
8	CDKN2A	7	AR
9	KIF2C	7	PGR
9	TTK	10	IGF1R

Fig. 4 Identification of hub genes in PPI networks. **A,B** The PPI networks among the differentially over-expressed (**A**) and under-expressed genes (**B**) were exported separately and entered into Cytoscape, the network analysis tool. The top ten hub genes (**A**, hub_oe; **B**, hub_ue) are ranked according to their number of interactors

Table 1 Potential biomarker candidate genes in metastatic breast cancer gene expression datasets

Potential biomarker candidate genes (over-expressed; DOE-A)	Relevant 11 studies
<i>CCNA2</i>	Bos Breast, Desmedt Breast, Hatzis Breast, Kao Breast, Loi Breast, Loi Breast 3, Schmidt Breast, Symmans Breast 2, Symmans Breast, vandeVijver Breast, Vantveer Breast
<i>CENPN</i>	
<i>DEPDC1</i>	
<i>TTK</i>	
Potential biomarker candidate genes (under-expressed; DUE-A)	Relevant 11 studies
<i>ABAT</i>	Bos Breast, Desmedt Breast, Hatzis Breast, Loi Breast, Loi Breast 3, Minn Breast 2, Schmidt Breast, Symmans Breast 2, Symmans Breast, vandeVijver Breast, and Vantveer Breast
<i>LRIG1</i>	Bos Breast, Desmedt Breast, Hatzis Breast, Kao Breast, Loi Breast, Loi Breast 3, Minn Breast 2, Schmidt Breast, Symmans Breast, vandeVijver Breast, and Vantveer Breast
<i>PGR</i>	Bos Breast, Desmedt Breast, Hatzis Breast, Kao Breast, Loi Breast, Loi Breast 3, Schmidt Breast, Symmans Breast 2, Symmans Breast, vandeVijver Breast, and Vantveer Breast

(DC = 11), and *IGF1R* (DC = 10) in DUE-L genes were ranked as top ten hub genes (hub_ue) (Fig. 4B).

Identification of novel biomarker candidate genes for breast cancer

As shown in Table 1, four (*CCNA2*, *CENPN*, *DEPDC1*, and *TTK*; DOE-A) and three genes (*ABAT*, *LRIG1*, and *PGR*; DUE-A) were identified as differentially over- and under-expressed genes, respectively, as described in “Methods” and they were selected as novel biomarker candidate genes for breast cancer and were subjected to the subsequent analyses.

Identification of PPI hub genes co-expressed with potential biomarker candidates

I attempted to find PPI hub genes (hub_oe and hub_ue) the most significantly and positively co-expressed with four DOE-A and three DUE-A novel biomarker candidate genes, respectively. Among the top ten hub_oe genes (Fig. 4A), *KIF2C* was the only gene that is the most significantly ($P < 0.0001$) and positively co-expressed with all four potential biomarker candidate genes (*AURKA* ($r = 0.75$) was co-expressed as positively as *KIF2C* ($r = 0.75$) with *CENPN*) (Fig. S1A). Among the top ten hub_ue genes (Fig. 4B), *ESR1* was the only gene that is the most significantly ($P < 0.0001$) and positively co-expressed with all three potential biomarker candidate genes (*FOXAI* ($r = 0.63$) was co-expressed as positively as *ESR1* ($r = 0.63$) with *LRIG1*) (Fig. S1B)

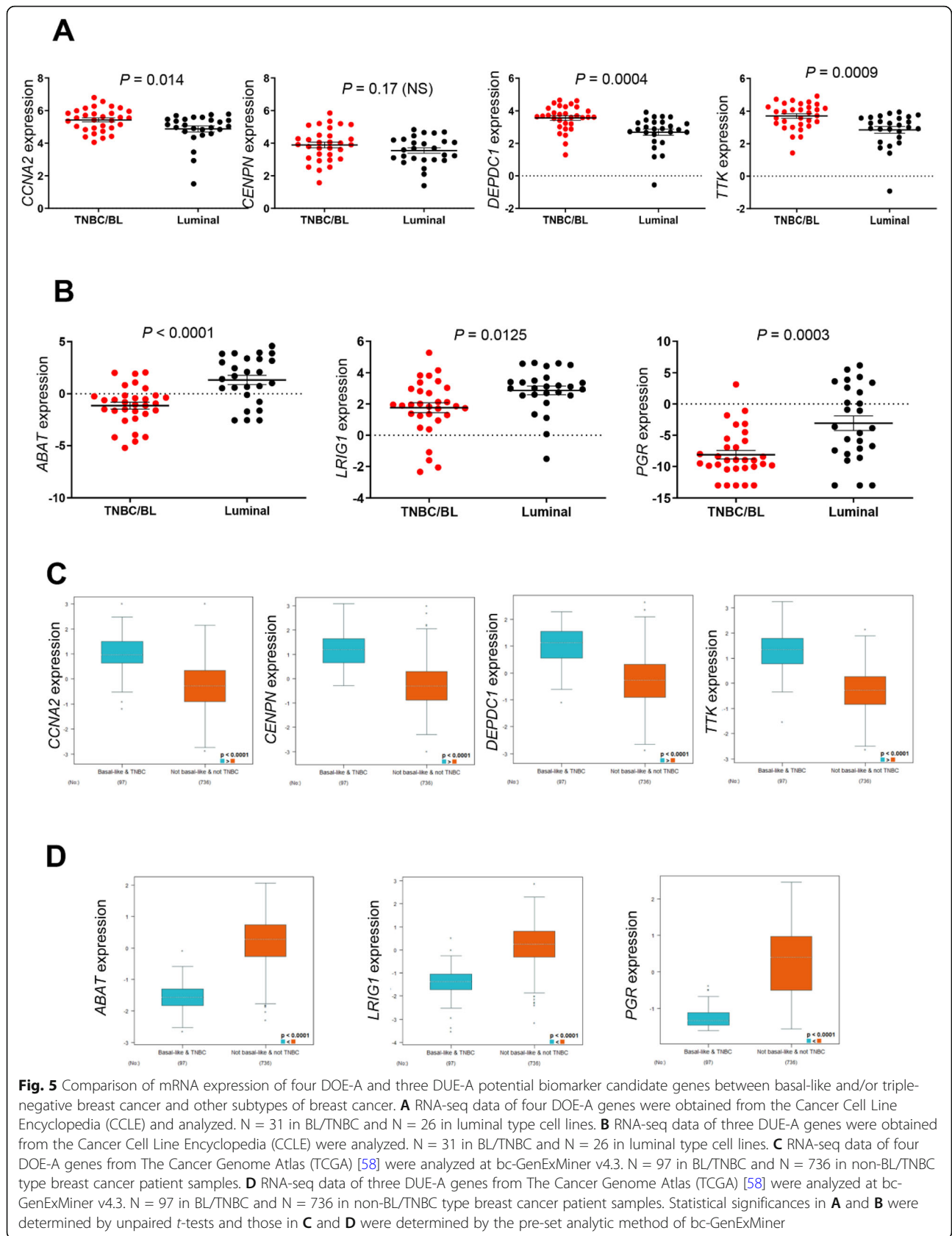
Examination of the expression correlation of potential biomarker candidate genes with the aggressive and metastatic nature of breast cancer

Basal-like (BL) and/or triple-negative breast cancers (TNBCs) are considered an aggressive and highly metastatic subtype of breast cancer often associated with poor clinical outcomes [22–27]. To examine the

expression correlation of the potential biomarker candidate genes (DOE-A and DUE-A) with the aggressive and metastatic nature of breast cancer, I compared their expression levels in BL/TNBCs with those in other breast cancer subtypes. First, I extracted RNA-seq expression data of human breast cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE). A total of 57 human breast cancer cell lines have expression information in the database and their subtypes were determined based on previous reports [17–19]. Among them, 26 are luminal and 31 are BL/TNBCs. The expression levels of *CCNA2*, *DEPDC1*, and *TTK* (DOE-A) were significantly higher in BL/TNBCs than in luminal breast cancer cell lines (Fig. 5A). The expression of all three DUE-A genes, on the other hand, was significantly lower in BL/TNBC cell lines, compared to luminal breast cancer cell lines (Fig. 5B).

Next, I chose to further investigate whether this correlation in cell lines could also be applied to human breast cancer patient samples. Using an online tool called bc-GenExMiner (version 4.3), I compared the gene expression between BL/TNBCs and other subtypes of human breast cancers. Consistent with cell line analysis, all four DOE-A genes were expressed significantly more (Fig. 5C) and all three DUE-A genes were expressed significantly less in BL/TNBCs than in non-BL and non-TNBCs (Fig. 5D). Together, the results in Fig. 5 strongly demonstrate that seven potential biomarker candidate genes (DOE-A and DUE-A) are positively and negatively correlated with the aggressive and metastatic nature of breast cancer, respectively.

Additionally, I examined two of the most significantly co-expressed hub genes (*KIF2C* and *ESR1*) shown in Fig. S1 and found that *KIF2C* and *ESR1* were significantly up- and downregulated in BL/TNBCs, respectively, compared to luminal breast cancer cell lines (Figs. S2A and S2B). Moreover, in human breast cancer patient samples,



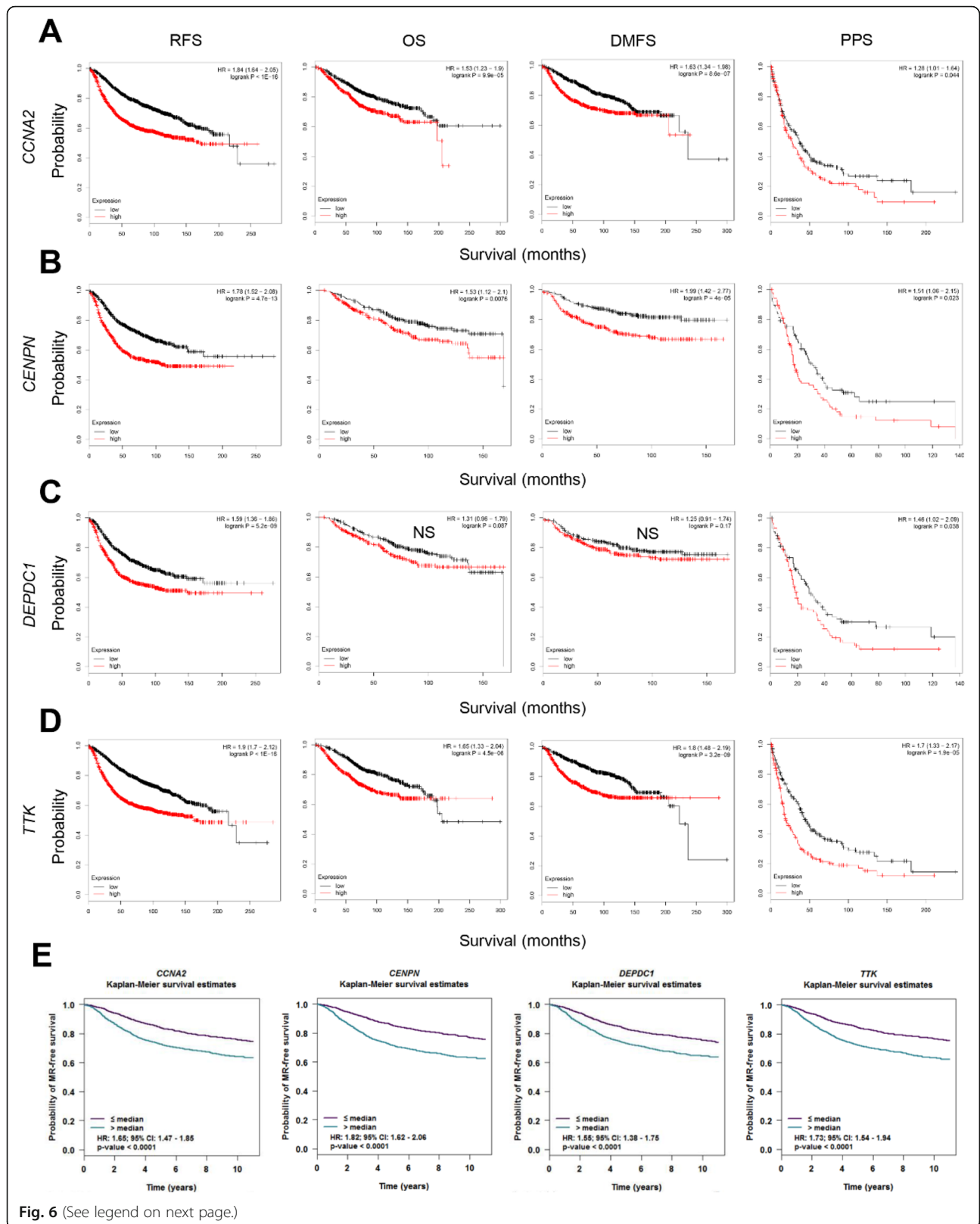


Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Correlation between the expression levels of potential prognostic biomarkers (DOE-A genes) and patient survival. **A–D** Relapse-free, overall, distant metastasis-free, and post-progression survival of four DOE-A genes (*CCNA2* in **A**; *CENPN* in **B**; *DEPDC1* in **C**; *TTK* in **D**) were stratified by the expression levels of each gene (low or high). Expression data were analyzed using KM plotter (<http://kmplot.com/>). JetSet best probes were selected and patients (for *CCNA2*, N = 3951 in RFS, = 1402 in OS, = 1746 in DMFS and = 414 in PPS; for *CENPN*, N = 1764 in RFS, = 626 in OS, = 664 in DMFS and = 173 in PPS; for *DEPDC1*, N = 1764 in RFS, = 626 in OS, = 664 in DMFS and = 173 in PPS; for *TTK*, N = 3951 in RFS, = 1402 in OS, = 1746 in DMFS and = 414 in PPS) were split by median expression. NS, not significant. **E** Metastatic relapse-free survival of four DOE-A genes were stratified by the expression levels of each gene (low or high). Microarray expression data were analyzed using bc-GenExMiner v4.3 (<http://bcgenex.centregauducheau.fr/>). Patients (for *CCNA2* and *TTK*, N = 4533; for *CENPN* and *DEPDC1*, N = 4359) were split by median expression. Statistical analyses were performed by pre-set analytic methods. HRs (hazardous ratios) and 95% CIs (confidence intervals) are indicated

the result was consistent (Figs. S2C and S2D). The data suggest that *KIF2C* and *ESR1*, two co-expressed hub genes, are also positively and negatively correlated with the aggressive and metastatic nature of breast cancer, respectively.

Examination of prognostic values of the biomarker candidate genes in breast cancer patients

To examine the prognostic values of four DOE-A and three DUE-A potential biomarkers in predicting breast cancer patient survival, I explored the correlation between their expression levels and the patients' clinical outcomes. For DOE-A genes, high levels of *CCNA2*, *CENPN*, and *TTK* expression were significantly associated with poor prognosis in all four available patient survivals when analyzed with KM plotter (RFS, relapse-free survival; OS, overall survival; DMFS, distant metastasis-free survival; PPS, post-progression survival). High expression of *DEPDC1*, another DOE-A gene, was significantly associated with poor prognosis only in RFS and PPS (Fig. 6A–D). Besides, high levels of all four DOE-A gene expressions were significantly correlated with metastatic relapse-free survival (MRFS) when analyzed with bc-GenExMiner (version 4.3) (Fig. 6E). For DUE-A genes, on the other hand, high levels of all three DUE-A gene expression were significantly associated with good patient RFS, OS, DMFS (except PPS) (Fig. 7A–C) and MRFS (Fig. 7D). I also examined two of the most significantly co-expressed hub genes, *KIF2C* and *ESR1*, and found that they were also significantly associated with poor and good clinical outcomes, respectively, in all five survival analyses (Fig. S3).

Discussion

Because of the limitations in the classical TNM staging system, The American Joint Committee on Cancer (AJCC) 8th edition added biological factors including estrogen and progesterone receptor expression and human epidermal growth factor 2 status for clinical prognostic staging in combination with the TNM staging [28]. Furthermore, when available, the use of multigene expression assays is recommended as stage modifiers [28]. By comparing the multigene assay panels recommended in AJCC 8th edition, I found that three biomarker genes

(*ESR1*, *PGR*, and *KIF2C*) were already included in at least one of the panels and the rest six biomarker genes (*CCNA2*, *CENPN*, *DEPDC1*, *TTK*, *ABAT*, and *LRIG1*) were not included in any of them. This suggests that the present study applied reliable analytic methods that could reproduce the prognostic value of some biomarkers as well as present meaningful novel prognostic biomarkers. Each multigene panel is, however, limited to use only in patients with specific stages and pathology, which implicates that the biomarker genes identified in the present study need additional validations to confirm their proper utility in the particular patient groups based on the stages and pathology.

CCNA2 encodes Cyclin A2 which functions as a cell cycle regulator and its expression is elevated in many human cancers. Moreover, *CCNA2* gene dysregulation is shown to be associated with poor prognosis [29–31]. *CENPN* encodes Centromere Protein N, which is important for the assembly of a multi-protein complex called kinetochore [32]. *DEPDC1* encodes DEP domain containing 1 protein, which has been shown to act as a transcription regulator by forming a complex with ZNF224, a member of the Krueppel C2H2-type zinc-finger protein family [33]. *TTK* encodes a dual-specificity protein kinase that can phosphorylate tyrosine and serine/threonine (threonine tyrosine kinase) and has crucial roles in regulating the spindle assembly checkpoint [34]. It is often overexpressed in breast tumors [35] and confers radioresistance [36]. *ABAT* encodes 4-aminobutyrate aminotransferase, which metabolizes GABA (γ -aminobutyric acid), a neurotransmitter. This gene expression is downregulated in inflammatory breast cancer and low expression of *ABAT* is correlated with a poor tamoxifen treatment outcome [37]. Moreover, it suppresses breast cancer metastasis [38]. *LRIG1* encodes a protein that negatively regulates epidermal growth factor receptor signaling, and its tumor-suppressive effects in cancer have been demonstrated [39–43]. *PGR* encodes the progesterone receptor, a member of the steroid receptor superfamily. Its expression is higher in luminal type A breast cancer than other aggressive breast cancer subtypes [44] and studies have demonstrated that progesterone receptor-positive (PR+) breast cancers are associated with better prognosis [45–47]. Furthermore,

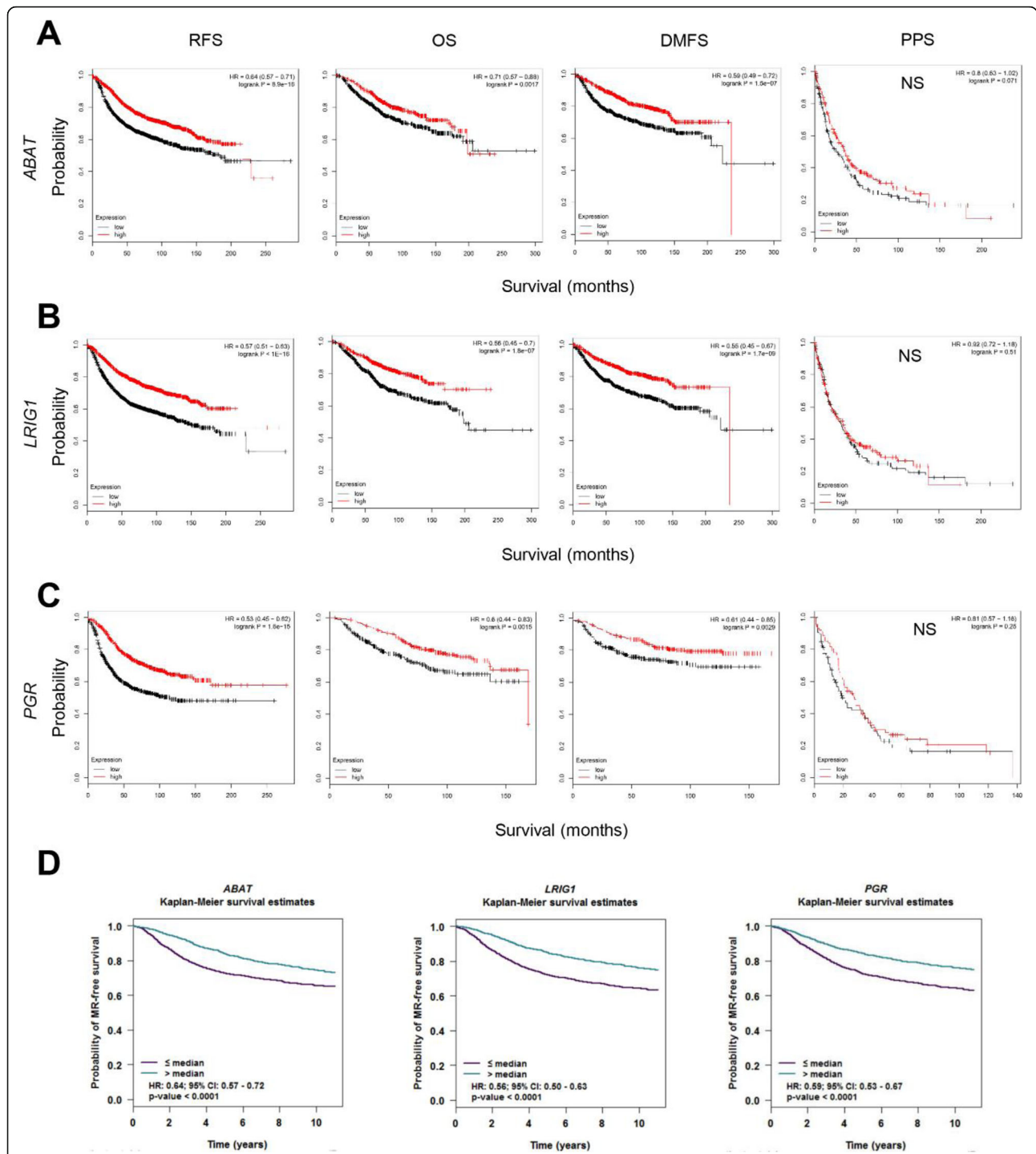


Fig. 7 Correlation between the expression levels of potential prognostic biomarkers (DUE-A genes) and patient survivals. **A–C** Relapse-free, overall, distant metastasis-free, and post-progression survival of three DUE-A genes (*ABAT* in **A**; *LRIG1* in **B**; *PGR* in **C**) were stratified by the expression levels of each gene (low or high). Expression data were analyzed using KM plotter (<http://kmplot.com/>). JetSet best probes were selected and patients (for *ABAT* and *LRIG1*, N = 3951 in RFS, = 1402 in OS, = 1746 in DMFS and = 414 in PPS; for *PGR*, N = 1764 in RFS, = 626 in OS, = 664 in DMFS and = 173 in PPS) were split by median expression. NS, not significant. **D** Metastatic relapse-free survival of three DUE-A genes was stratified by the expression levels of each gene (low or high). Microarray expression data were analyzed by bc-GenExMiner v4.3 (<http://bcgenex.centregauducheau.fr/>). Patients (for *ABAT*, *LRIG1*, and *PGR*, N = 4434) were split by median expression. Statistical analyses were performed by pre-set analytic methods. HRs (hazardous ratios) and 95% CIs (confidence intervals) are indicated

KIF2C encodes a kinesin-like microtubule-dependent motor protein, which depolymerizes microtubules and promotes chromosomal segregation [48, 49]. Its overexpression has been observed in human breast cancer cases and cell lines [50, 51]. *ESR1* encodes estrogen receptor α , a hormone receptor whose transcription activity is regulated by estrogen binding. Patients with estrogen receptor α positive (ER α +) breast tumors have demonstrated better survival and later recurrence than those with ER α - breast tumors [52–54] (Table 2).

Overall, it is interesting to note that cell cycle-related genes (*CCNA2*, *CENPN*, *TTK*, and *KIF2C*) and hormone signaling-related genes (*ABAT*, *PGR*, and *ESR1*) were differentially over- and under-expressed in the metastatic breast cancers, respectively. They were also predominantly associated with poor and good clinical outcomes, respectively. The results suggest that targeting cell cycle regulators may but hormonal therapy may not be beneficial for metastatic breast cancer patients, in general, although an individual patient may respond differently. Indeed, cell cycle inhibitors such as CDK4/6i (inhibitor of the cyclin-dependent kinases 4 and 6) have been approved and used for metastatic breast cancer patients either alone or in a combinational therapy [55].

In addition, I attempted to identify functional, biological, molecular, and cellular processes specifically altered in metastatic human breast cancers (MBCs). Differentially expressed genes in MBCs are mostly involved in regulating cell death, epidermal growth factor receptor signaling, and membrane and cytoskeletal structures, and are also enriched in biological pathways such as progesterone- and estrogen-related signaling. In fact, EGF receptor inhibition often fails in the treatment of metastatic breast cancer potentially due to the “paradoxical” anti-proliferative and anti-metastatic function of EGF receptor signaling [56], which implicates that EGF receptor inhibitors should be used with caution in

metastatic breast cancer. Moreover, cancer metastasis and chemoresistance are demonstrated as a linked phenotype [57], which implies that chemotherapy-induced cell death signaling is fundamentally altered in metastatic breast cancer.

Although I demonstrated that the expression levels of potential biomarkers are positively/negatively correlated with the aggressive and metastatic nature of breast cancer and are associated with clinical outcomes of breast cancer patients, their molecular functions except for *CCNA2*, *PGR*, and *ESR1* have not been experimentally elucidated in breast carcinogenesis. Future functional validation is needed to warrant their potential values as breast cancer biomarkers as well as tumor-promoting or tumor-suppressing molecules. Also, the present study proves the usefulness of OncoPrint platform to identify enriched pathways and potential prognostic biomarkers to predict beneficial treatment options for and the clinical outcomes of breast cancer.

Conclusions

In the present study, I delineated biological functions and pathways specifically enriched in metastatic breast cancer and demonstrated that *CCNA2*, *CENPN*, *DEPDC1*, *TTK*, *ABAT*, *LRIG1*, *PGR*, *KIF2C*, and *ESR1* may serve as biomarkers to predict clinical outcomes of breast cancer patients. Pathway analysis suggests which therapeutic opportunities, in general, may or may not potentially be beneficial to the treatment of metastatic breast cancers. Additionally, the present study demonstrates the usefulness of OncoPrint data-mining platform. Further functional studies are needed to warrant validation of the roles of selected genes as functional tumor-promoting or tumor-suppressing molecules.

Abbreviations

MBC: Metastatic breast cancer; GO: Gene ontology; KEGG : Kyoto Encyclopedia of Genes and Genomes; PPI: Protein-protein interaction; DOE-L

Table 2 Summary of nine selected genes and their roles as prognostic biomarkers in breast cancer.

Gene symbols	Gene names	Prognostic value in breast cancer with high expression	Cellular/biological roles
<i>CCNA2</i>	Cyclin A2	Poor	Cell cycle regulator
<i>CENPN</i>	Centromere protein N	Poor	Regulates kinetochore assembly
<i>DEPDC1</i>	DEP domain containing 1	Poor	Transcription regulator
<i>TTK</i>	TTK protein kinase	Poor	Regulates the spindle assembly checkpoint
<i>KIF2C</i>	Kinesin family member 2C	Poor	Microtubule-dependent motor protein
<i>ABAT</i>	4-Aminobutyrate aminotransferase	Good	Metabolizes GABA
<i>LRIG1</i>	Leucine-rich repeats and immunoglobulin-like domains 1	Good	Negatively regulates EGFR signaling
<i>PGR</i>	Progesterone receptor	Good	Hormone receptor
<i>ESR1</i>	Estrogen Receptor 1	Good	Hormone receptor

/ DUE-L: Differentially over- or under-expressed genesets from the four datasets with large patient numbers ($N > 200$); DOE-A / DUE-A: Differentially over- or under-expressed genesets from the entire twelve datasets with any patient numbers; *CCNA2*: Cyclin A2; *CENPN*: Centromere protein N; *DEPD C1*: DEP domain containing 1; *TTK*: TTK protein kinase (Thr/Tyr kinase); *ABAT*: 4-Aminobutyrate aminotransferase; *LRIG1*: Leucine-rich repeats and immunoglobulin like domains 1; *PGR*: Progesterone receptor; *IL6*: Interleukin 6; *CXCL8*: C-X-C motif chemokine ligand 8; *AURKA*: Aurora kinase A; *NOTCH1*: Notch receptor 1; *CDC20*: Cell division cycle 20; *APOE*: Apolipoprotein E; *CDKN2A*: Cyclin-dependent kinase inhibitor 2A; *KIF2C*: Kinesin family member 2C; *ESR1*: Estrogen receptor 1; *FOXA1*: Forkhead box A1; *GATA3*: GATA binding protein 3; *EEF2*: Eukaryotic translation elongation factor 2; *RPL7A*: Ribosomal protein L7a; *TFF1*: Trefoil factor 1; *RPL15*: Ribosomal protein L15; *AR*: Androgen receptor; *IGF1R*: Insulin-like growth factor 1 receptor; *STRING*: Search tool for the retrieval of interacting genes; *DC*: Degree of connectivity; *CCL*: Cancer Cell Line Encyclopedia; *BL*: Basal-like; *TNBC*: Triple-negative breast cancer; *PAM50*: Prediction analysis of microarray 50 (PAM50); *IHC*: Immunohistochemistry; *KM* plotter: Kaplan-Meier plotter; *RFS*: Relapse-free survival; *OS*: Overall survival; *DMFS*: Distant metastasis-free survival (DMFS); *PPS*: Post-progression survival; *MRFS*: Metastatic relapse-free survival; *HR*: Hazardous ratio; *CI*: Confidence interval; *BP*: Biological process; *MF*: Molecular function; *CC*: Cellular component; *hub_oe*: Hub genes co-expressed with DOE-A genes; *hub_ue*: Hub genes co-expressed with DUE-A genes; *TCGA*: The Cancer Genome Atlas

Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1186/s12957-021-02301-7>.

Additional file 1: Supplementary Figure S1. *KIF2C* and *ESR1* are hub genes the most significantly co-expressed with the potential biomarker candidate genes. (A and B) Shown are Pearson's pairwise correlation plots of RNA-seq gene expression between four DOE-A (A) or three DUE-A genes (B) and their most significantly co-expressed hub genes identified from PPI networks. Statistical analyses were performed by the pre-set analytic method of bc-GenExMiner. Supplementary Figure S2. Comparison of mRNA expression of the two most significantly co-expressed hub genes (*KIF2C* and *ESR1*) between basal-like or triple-negative breast cancer and other subtypes of breast cancer. (A and B) RNA-seq data of *KIF2C* and *ESR1* were obtained from the Cancer Cell Line Encyclopedia (CCLE) and analyzed. $N = 31$ in BL/TNBC and $N = 26$ in luminal type cell lines. (C and D) RNA-seq data of *KIF2C* and *ESR1* from The Cancer Genome Atlas (TCGA) [58] were analyzed at bc-GenExMiner v4.3. $N = 97$ in BL/TNBC and $N = 736$ in non-BL/TNBC type breast cancer patient samples. Statistical significance in A and B was determined by unpaired *t*-tests and those in C and D were determined by the pre-set analytic method of bc-GenExMiner. Supplementary Figure S3. Correlation between the expression levels of two co-expressed hub genes (*KIF2C* and *ESR1*) and patient survivals. (A and B) Relapse-free, overall, distant metastasis-free, and post-progression survival of two co-expressed hub genes (*KIF2C* in (A); *ESR1* in (B)) were stratified by the expression levels of each gene (low or high). Expression data were analyzed by KM plotter (<http://kmplot.com/>). JetSet best probes were selected and patients (for *KIF2C*, $N = 3951$ in RFS, $= 1402$ in OS, $= 1746$ in DMFS and $= 414$ in PPS; for *ESR1*, $N = 3951$ in RFS, $= 1402$ in OS, $= 1746$ in DMFS and $= 414$ in PPS) were split by median expression. (C) Metastatic relapse-free survival of *KIF2C* and *ESR1* was stratified by the expression levels of each gene (low or high). Microarray expression data were analyzed by bc-GenExMiner v4.3 (<http://bcgenex.centregauducheau.fr/>). Patients (*KIF2C*, $N = 4533$; *ESR1*, $N = 4785$) were split by median expression. Statistical analyses were performed by pre-set analytic methods. HRs (hazardous ratios) and 95% CIs (confidence intervals) are indicated.

Additional file 2: Table S1. Twelve raw datasets with over-expressed genes.

Additional file 3: Table S2. Twelve raw datasets with under-expressed genes.

Acknowledgements

I thank Qinglei Hang and Moon Jong Kim at UT MD Anderson Cancer Center, the USA, for their statistical assistance. I also thank Ashley Siverly at Methodist Hospital, the USA, for proofreading this manuscript.

Author's contributions

J.K. conceived and designed the study. J.K. analyzed and interpreted the data. J.K. wrote the manuscript. The author read and approved the final manuscript.

Funding

This work was supported by the Sogang University Research Grant of 2019 [201910004.01] and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) [No. 2019R1F1A1060705, 2020R1F1A1065643 and 2021R1F1A1062226].

Availability of data and materials

The gene expression datasets are available at [OncoPrint.org](https://www.oncoPrint.org).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The author declares no conflict of interest.

Received: 12 April 2021 Accepted: 12 June 2021

Published online: 25 June 2021

References

- De Vita VT Jr. Breast cancer therapy: exercising all our options. *N Engl J Med*. 1989;320(8):527–9. <https://doi.org/10.1056/NEJM198902233200812>.
- Emens LA. Breast cancer immunotherapy: facts and hopes. *Clin Cancer Res*. 2018;24(3):511–20. <https://doi.org/10.1158/1078-0432.CCR-16-3001>.
- Youlten DR, Cramb SM, Yip CH, Baade PD. Incidence and mortality of female breast cancer in the Asia-Pacific region. *Cancer Biol Med*. 2014;11(2): 101–15. <https://doi.org/10.7497/j.issn.2095-3941.2014.02.005>.
- Kulkarni A, Stroup AM, Paddock LE, Hill SM, Plascak JJ, Llanos AAM. Breast cancer incidence and mortality by molecular subtype: statewide age and racial/ethnic disparities in New Jersey. *Cancer Health Disparities*. 2019;3:e1–e17. <https://doi.org/10.9777/chd.2019.1012>.
- Howlader N, Noone AM, Krapcho M, (editors). ea. Table 4.5: Cancer of the breast (invasive). Age-adjusted SEER incidence rates by year, race and sex. National Cancer Institute. Bethesda, MD. Accessed on April 27, 2020. https://seer.cancer.gov/csr/1975_2017/. Cancer Statistics Review, 1975–2017. 2020.
- Howlader N NA, Krapcho M, et al. (editors). Cancer Statistics Review, 1975–2017. Table 4.13: Cancer of the female breast (invasive): 5-year relative and period survival by race, diagnosis year, age and stage at diagnosis. National Cancer Institute. Bethesda, MD. Accessed on April 27, 2020. https://seer.cancer.gov/csr/1975_2017/, 2020.
- Ghafouri-Fard S, Oskooei VK, Azari I, Taheri M. Suppressor of cytokine signaling (SOCS) genes are downregulated in breast cancer. *World J Surg Oncol*. 2018;16(1):226. <https://doi.org/10.1186/s12957-018-1529-9>.
- Jia R, Li Z, Liang W, Ji Y, Weng Y, Liang Y, et al. Identification of key genes unique to the luminal a and basal-like breast cancer subtypes via bioinformatic analysis. *World J Surg Oncol*. 2020;18(1):268. <https://doi.org/10.1186/s12957-020-02042-z>.
- Liu X, Jin G, Qian J, Yang H, Tang H, Meng X, et al. Digital gene expression profiling analysis and its application in the identification of genes associated with improved response to neoadjuvant chemotherapy in breast cancer. *World J Surg Oncol*. 2018;16(1):82. <https://doi.org/10.1186/s12957-018-1380-z>.
- Mao XH, Ye Q, Zhang GB, Jiang JY, Zhao HY, Shao YF, et al. Identification of differentially methylated genes as diagnostic and prognostic biomarkers of breast cancer. *World J Surg Oncol*. 2021;19(1):29. <https://doi.org/10.1186/s12957-021-02124-6>.
- Mohamad-alizadeh-Hanjani Z, Shahbazi S, Geranpayeh L. Investigation of the SPAG5 gene expression and amplification related to the NuMA mRNA levels in breast ductal carcinoma. *World J Surg Oncol*. 2020;18(1):225. <https://doi.org/10.1186/s12957-020-02001-8>.
- Yuan Q, Zheng L, Liao Y, Wu G. Overexpression of CCNE1 confers a poorer prognosis in triple-negative breast cancer identified by bioinformatic

- analysis. *World J Surg Oncol*. 2021;19(1):86. <https://doi.org/10.1186/s12957-021-02200-x>.
13. Zhou X, Xiao C, Han T, Qiu S, Wang M, Chu J, et al. Prognostic biomarkers related to breast cancer recurrence identified based on Logit model analysis. *World J Surg Oncol*. 2020;18(1):254. <https://doi.org/10.1186/s12957-020-02026-z>.
 14. Zhu C, Hu H, Li J, Wang J, Wang K, Sun J. Identification of key differentially expressed genes and gene mutations in breast ductal carcinoma in situ using RNA-seq analysis. *World J Surg Oncol*. 2020;18(1):52. <https://doi.org/10.1186/s12957-020-01820-z>.
 15. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004;6(1):1–6. [https://doi.org/10.1016/S1476-5586\(04\)80047-2](https://doi.org/10.1016/S1476-5586(04)80047-2).
 16. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014;8(Suppl 4):S11.
 17. Dai X, Cheng H, Bai Z, Li J. Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. *J Cancer*. 2017;8(16):3131–41. <https://doi.org/10.7150/jca.18457>.
 18. Jiang G, Zhang S, Yazdanparast A, Li M, Pawar AV, Liu Y, et al. Comprehensive comparison of molecular portraits between cell lines and tumors in breast cancer. *BMC Genomics*. 2016;17 Suppl 7:525.
 19. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*. 2006;10(6):515–27. <https://doi.org/10.1016/j.ccr.2006.10.008>.
 20. Ghadie M, Xia Y. Estimating dispensable content in the human interactome. *Nat Commun*. 2019;10(1):3205. <https://doi.org/10.1038/s41467-019-11180-2>.
 21. Wodak SJ, Pu S, Vlasblom J, Seraphin B. Challenges and rewards of interaction proteomics. *Mol Cell Proteomics*. 2009;8(1):3–18. <https://doi.org/10.1074/mcp.R800014-MCP200>.
 22. Aysola K, Desai A, Welch C, Xu J, Qin Y, Reddy V, et al. Triple negative breast cancer - an overview. *Hereditary Genet*. 2013;2013(Suppl 2).
 23. Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, et al. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin Cancer Res*. 2008;14(5):1368–76. <https://doi.org/10.1158/1078-0432.CCR-07-1658>.
 24. Oner G, Altintas S, Canturk Z, Tjalma W, Verhoeven Y, Van Berckelaer C, et al. Triple-negative breast cancer-role of immunology: a systemic review. *Breast J*. 2019.
 25. Toft DJ, Cryns VL. Minireview: Basal-like breast cancer: from molecular profiles to targeted therapies. *Mol Endocrinol*. 2011;25(2):199–211. <https://doi.org/10.1210/me.2010-0164>.
 26. Garrido-Castro AC, Lin NU, Polyak K. Insights into molecular classifications of triple-negative breast cancer: improving patient selection for treatment. *Cancer Discov*. 2019;9(2):176–98. <https://doi.org/10.1158/2159-8290.CD-18-1177>.
 27. Yin L, Duan JJ, Bian XW, Yu SC. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res*. 2020;22(1):61. <https://doi.org/10.1186/s13058-020-01296-5>.
 28. Hortobagyi GN, Connolly JL, D'Orsi CJ, Edge SB, Mittendorf EA, Rugo HS, et al. *Breast. Eighth Edition: AJCC Cancer Staging Manual*; 2017.
 29. Li JQ, Miki H, Wu F, Sáo K, Nishioka M, Ohmori M, et al. Cyclin A correlates with carcinogenesis and metastasis, and p27(kip1) correlates with lymphatic invasion, in colorectal neoplasms. *Hum Pathol*. 2002;33(10):1006–15. <https://doi.org/10.1053/hupa.2002.125774>.
 30. Yang L, Zeng W, Sun H, Huang F, Yang C, Cai X, et al. Bioinformatical analysis of Gene Expression Omnibus database associates TAF7/CCNB1, TAF7/CCNA2, and GTF2E2/CDC20 pathways with glioblastoma development and prognosis. *World Neurosurg*. 2020;138:e492–514. <https://doi.org/10.1016/j.wneu.2020.02.159>.
 31. Yasmeen A, Berdel WE, Serve H, Muller-Tidow C. E- and A-type cyclins as markers for cancer diagnosis and prognosis. *Expert Rev Mol Diagn*. 2003; 3(5):617–33. <https://doi.org/10.1586/14737159.3.5.617>.
 32. Mellone B, Erhardt S, Karpén GH. The ABCs of centromeres. *Nat Cell Biol*. 2006;8(5):427–9. <https://doi.org/10.1038/ncb0506-427>.
 33. Harada Y, Kanehira M, Fujisawa Y, Takata R, Shuin T, Miki T, et al. Cell-permeable peptide DEPDC1-ZNF224 interferes with transcriptional repression and oncogenicity in bladder cancer cells. *Cancer Res*. 2010;70(14): 5829–39. <https://doi.org/10.1158/0008-5472.CAN-10-0255>.
 34. Lauze E, Stoelcker B, Luca FC, Weiss E, Schutz AR, Winey M. Yeast spindle pole body duplication gene MPS1 encodes an essential dual specificity protein kinase. *EMBO J*. 1995;14(8):1655–63.
 35. Mason JM, Wei X, Fletcher GC, Kiarash R, Broxk R, Hodgson R, et al. Functional characterization of CFI-402257, a potent and selective Mps1/TTK kinase inhibitor, for the treatment of cancer. *Proc Natl Acad Sci U S A*. 2017; 114(12):3127–32. <https://doi.org/10.1073/pnas.1700234114>.
 36. Chandler BC, Moubadder L, Ritter CL, Liu M, Cameron M, Wilder-Romans K, et al. TTK inhibition radiosensitizes basal-like breast cancer through impaired homologous recombination. *J Clin Invest*. 2020;130(2):958–73. <https://doi.org/10.1172/JCI130435>.
 37. Jansen MP, Sas L, Siewewerts AM, Van Cauwenbergh C, Ramirez-Ardila D, Look M, et al. Decreased expression of ABAT and STC2 hallmarks ER-positive inflammatory breast cancer and endocrine therapy resistance in advanced disease. *Mol Oncol*. 2015;9(6):1218–33. <https://doi.org/10.1016/j.molonc.2015.02.006>.
 38. Chen X, Cao Q, Liao R, Wu X, Xun S, Huang J, et al. Loss of ABAT-mediated GABAergic system promotes basal-like breast cancer progression by activating Ca(2+)-NFAT1 axis. *Theranostics*. 2019;9(1):34–47. <https://doi.org/10.7150/thno.29407>.
 39. Gur G, Rubin C, Katz M, Amit I, Citri A, Nilsson J, et al. LRIG1 restricts growth factor signaling by enhancing receptor ubiquitylation and degradation. *EMBO J*. 2004;23(16):3270–81. <https://doi.org/10.1038/sj.emboj.7600342>.
 40. Ji Y, Kumar R, Gokhale A, Chao HP, Rycak K, Chen X, et al. LRIG1, a regulator of stem cell quiescence and a pleiotropic feedback tumor suppressor. *Semin Cancer Biol*. 2021. <https://doi.org/10.1016/j.semcancer.2020.12.016>.
 41. Li Q, Liu B, Chao HP, Ji Y, Lu Y, Mehmood R, et al. LRIG1 is a pleiotropic androgen receptor-regulated feedback tumor suppressor in prostate cancer. *Nat Commun*. 2019;10(1):5494. <https://doi.org/10.1038/s41467-019-13532-4>.
 42. Morrison MM, Williams MM, Vaught DB, Hicks D, Lim J, McKernan C, et al. Decreased LRIG1 in fulvestrant-treated luminal breast cancer cells permits ErbB3 upregulation and increased growth. *Oncogene*. 2016;35(9):1206. <https://doi.org/10.1038/ncr.2015.418>.
 43. Torigoe H, Yamamoto H, Sakaguchi M, Youyi C, Namba K, Sato H, et al. Tumor-suppressive effect of LRIG1, a negative regulator of ErbB, in non-small cell lung cancer harboring mutant EGFR. *Carcinogenesis*. 2018;39(5): 719–27. <https://doi.org/10.1093/carcin/bgy044>.
 44. Braun L, Mietzsch F, Seibold P, Schneeweiss A, Schirmacher P, Chang-Claude J, et al. Intrinsic breast cancer subtypes defined by estrogen receptor signalling-prognostic relevance of progesterone receptor loss. *Mod Pathol*. 2013;26(9):1161–71. <https://doi.org/10.1038/modpathol.2013.60>.
 45. Purdie CA, Quinlan P, Jordan LB, Ashfield A, Ogston S, Dewar JA, et al. Progesterone receptor expression is an independent prognostic variable in early breast cancer: a population-based study. *Br J Cancer*. 2014;110(3):565–72. <https://doi.org/10.1038/bjc.2013.756>.
 46. Ueno T, Saji S, Chiba T, Kamma H, Isaka H, Itoh H, et al. Progesterone receptor expression in proliferating cancer cells of hormone-receptor-positive breast cancer. *Tumour Biol*. 2018;40(10):1010428318811025. <https://doi.org/10.1177/1010428318811025>.
 47. Van Belle V, Van Calster B, Brouckaert O, Vanden Bempt I, Pintens S, Harvey V, et al. Qualitative assessment of the progesterone receptor and HER2 improves the Nottingham Prognostic Index up to 5 years after breast cancer diagnosis. *J Clin Oncol*. 2010;28(27):4129–34. <https://doi.org/10.1200/JCO.2009.26.4200>.
 48. Moore AT, Rankin KE, von Dassow G, Peris L, Wagenbach M, Ovechkina Y, et al. MCAK associates with the tips of polymerizing microtubules. *J Cell Biol*. 2005;169(3):391–7. <https://doi.org/10.1083/jcb.200411089>.
 49. Shao H, Huang Y, Zhang L, Yuan K, Chu Y, Dou Z, et al. Spatiotemporal dynamics of Aurora B-PLK1-MCAK signaling axis orchestrates kinetochore bi-orientation and faithful chromosome segregation. *Sci Rep*. 2015;5(1):12204. <https://doi.org/10.1038/srep12204>.
 50. Li TF, Zeng HJ, Shan Z, Ye RY, Cheang TY, Zhang YJ, et al. Overexpression of kinesin superfamily members as prognostic biomarkers of breast cancer. *Cancer Cell Int*. 2020;20(1):123. <https://doi.org/10.1186/s12935-020-01191-1>.
 51. Shimo A, Tanikawa C, Nishidate T, Lin ML, Matsuda K, Park JH, et al. Involvement of kinesin family member 2C/mitotic centromere-associated kinesin overexpression in mammary carcinogenesis. *Cancer Sci*. 2008;99(1): 62–70. <https://doi.org/10.1111/j.1349-7006.2007.00635.x>.
 52. Bentzon N, Durning M, Rasmussen BB, Mouridsen H, Kroman N. Prognostic effect of estrogen receptor status across age in primary breast cancer. *Int J Cancer*. 2008;122(5):1089–94. <https://doi.org/10.1002/ijc.22892>.
 53. Fisher B, Redmond C, Fisher ER, Caplan R. Relative worth of estrogen or progesterone receptor and pathologic characteristics of differentiation as indicators of prognosis in node negative breast cancer patients: findings

- from National Surgical Adjuvant Breast and Bowel Project Protocol B-06. *J Clin Oncol.* 1988;6(7):1076–87. <https://doi.org/10.1200/JCO.1988.6.7.1076>.
54. Hua H, Zhang H, Kong Q, Jiang Y. Mechanisms for estrogen receptor expression in human cancer. *Exp Hematol Oncol.* 2018;7(1):24. <https://doi.org/10.1186/s40164-018-0116-7>.
 55. Piezzo M, Cocco S, Caputo R, Cianniello D, Gioia GD, Lauro VD, et al. Targeting Cell Cycle in Breast Cancer: CDK4/6 Inhibitors. *Int J Mol Sci.* 2020; 21(18).
 56. Ali R, Wendt MK. The paradoxical functions of EGFR during breast cancer progression. *Signal Transduct Target Ther.* 2017;2(1). <https://doi.org/10.1038/sigtrans.2016.42>.
 57. Acharyya S, Oskarsson T, Vanharanta S, Malladi S, Kim J, Morris PG, et al. A CXCL1 paracrine network links cancer chemoresistance and metastasis. *Cell.* 2012;150(1):165–78. <https://doi.org/10.1016/j.cell.2012.04.042>.
 58. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490(7418):61–70. <https://doi.org/10.1038/nature11412>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

