Plasticity as an Underlying Mechanism of Tumor Heterogeneity in Breast Cancer

Authors: Muhammad Waqas Akbar[1], Murat Isbilen[2], Baris Kucukkaraduman[1], Secil Demirkol Canli[3], Ege Dedeoglu[1], Shila Azizolli[1], Isli Cela[1], Abbas Guven Akcay[1], Hasim Hakanoglu[1], Ali Osmay Gure[1,*]

1 Department of Molecular Biology and Genetics, Bilkent University, Ankara, Turkey

2 DNAFect Genetics Consulting R&D and Biotechnology Inc., Kocaeli, Turkey

3 Molecular Pathology Application and Research Center, Hacettepe University, Ankara, Turkey

* agure@bilkent.edu.tr

Abstract

Breast cancer shows plasticity in terms of classification. Upon drug treatment and metastasis some tumors switch to another subtype leading to loss of response to therapy. In this study, we ask the question which molecular subclasses of breast cancer are more switchable upon drug therapy and metastasis. We used in silico data to classify breast cancer tumors in PAM50 molecular classes before treatment and after treatment using gene expression data. Similar analysis was performed for primary tumors and their metastatic growth. Our analysis showed that in both scenarios some breast tumors shift from one class to another. This suggests that patients who underwent chemotherapy but resulted in relapse or/and metastasis should be retyped for molecular subclass so that treatment protocol should be adopted according to those subtypes. Additionally, 20 genes were identified as biomarkers for metastasis in breast cancer.

Introduction

Breast cancer (BC) is the second leading cause for cancer related mortality in women after lung cancer[1,2]. In clinics, breast cancer is classified using estrogen receptor (ER), progesterone receptor (PGR) and Erb-B2 Receptor Tyrosine Kinase 2 (ERBB2, previously known as HER2) statuses through immunohistochemistry (IHC classes). Moreover, BC is divided in Luminal A, Luminal B, HER2 enriched, Basal and Normal like subgroups by molecular classification [3]. Metastatic breast cancer (MBC) is the type of BC in which cancer cells leave primary cancer tissue and spread to other parts of the body. MBC remains an incurable disease as its median overall survival is about 3 years and 25% patients show only 5 year survival[4].

MBC consists of a series of sequential steps, including transfer of cells from primary tumor to circulation (intravasation), survival in circulation, extravasation in a new organ, colonization and metastatic growth. Metastatic cells go through epithelial to mesenchymal transition (EMT) to achieve metastasis[5]. This metastatic cascade results in switching the IHC status of BC in some samples, assigning metastatic growth a different class for which primary therapy might not be suitable and a different protocol should be opted to combat this growth[6]. Previously similar subtype switching is observed in PAM50 classes as well [7]. In this paper we address this switching as plasticity. Additionally, this subtype plasticity is also seen among pre- and post-treated tumors due to epithelial to mesenchymal transition[8].

In this paper we used public gene expression datasets to determine if there is a subtype switch between pre- and post- therapy patient tumor samples and also if this switch is present between primary and metastatic tissue. Moreover MBC biomarkers were identified using primary tumor expression data which can be used to determine if a breast cancer is capable of metastasis or not.

Methods

PAM50 plasticity

Datasets GSE10281, GSE32518, GSE32072 (patient samples before and after treatment), GSE32518 (patient samples collected with fine needle aspiration and core needle biopsy), GSE125989, GSE110590 (patient samples for primary and metastatic tissues) were downloaded from GEO NCBI database. All datasets were "justRMA" normalized using BRB Array Tools[9]. PAM50 signature was used to classify breast cancer into 5 molecular subgroups: Luminal A (LumA), Luminal B (LumB), HER2 enriched (HER2), Basal and Normal like (Normal) as described previously[10] for each dataset. Paired samples were matched for pre- and post- therapy, with fine needle aspiration and core needle biopsy, or primary and metastatic tissue to compare PAM50 plasticity in these samples.

Metastasis biomarkers

For metastasis-free survival analysis, GSE2603 (n=82, treatment not disclosed), GSE6532 ((n= 293, platform: Affymetrix Human Genome U133A and B Array, Tamoxifen treated), (n=87, Affymetrix Human Genome U133 Plus 2.0 Array)), GSE7390 (n=198, systemically untreated patients), GSE11121 (n=200, only surgery was performed), GSE12276 (n=204, adjuvant chemotherapy treated), GSE20685 (n=327, adjuvant chemotherapy treated) and GSE58812 (n=108, adjuvant chemotherapy treated) were downloaded. All datasets were "justRMA" normalized using BRB Array Tools. We performed cox proportional hazard analysis using "Survival" package built in R software for all datasets[11,12]. Genes that showed same direction in all datasets and were significant for 5/8 analysis were selected as metastasis biomarkers. 40 probesets were found significant and to select top 10 good and bad prognostic genes, genes were ranked by coxp value in each dataset individually for both directions and these ranks were summed. Top 10 genes were then selected from good prognostic genes based upon the least ranksum score and another 10 genes were selected from bad prognostic genes based on the same criteria. For genes which were significant with 2 or more probesets, only one probeset with the least ranksum score was selected.

Results

PAM50 plasticity

PAM50 is also prone to show different classes, as analyzed in dataset GSE32518, where authors collected fine needle aspiration and core needle biopsies from matching patients (N=37). When these samples were typed for PAM50, 78% of these samples (n=29) showed same phenotype in both biopsy methods while 22% samples (n=8) showed discordant subtypes. One HER2 sample by core needle biopsy was identified as Basal and another one was identified as LumB. One sample was identified as LumA (Core needle biopsy) and LumB (Fine needle aspiration). Most discordant results were for Normal subtype (Core needle biopsy) where 4 samples were identified Basal and 1 sample as LumA (Table 1). This PAM50 discordance can be attributed to tumor heterogeneity.

Plasticity exists between PAM50 classes, and tumor subtype can switch to another upon drug treatment through EMT and immunological status. To study PAM50 plasticity, we analyzed different datasets in which matched tumor samples gene expression data was available for both pre- and post-chemotherapy. When Letrozole pre- and post-treated samples (GSE10281) were compared for PAM50 class plasticity, 39% samples (7/18) retained their PAM50 class before and after Letrozole therapy and 61% (11/18) samples changed their class. Basal (n=2) and LumB (n=1) subtype was most stable in this dataset. One HER2, 2 LumA and 1 normal subtype samples retained their subtype while 3 LumA, 3 HER2 and 4

normal pre therapy samples switched to another subtype. Most of these samples (n=6) switched to LumB and 4 samples switched to Basal subtype (Table 2). Authors reported in their paper that upon chemotherapy treatment, residual tumors attain mesenchymal properties as expected[13]. Among patients treated with Anthracyclines and Taxanes based neoadjuvant chemotherapy (GSE32518), 2 basal samples, 1 HER2, 3 LumA, 3 lumB and 1 Normal retained their subtype and 2 basal, 3 HER2, 4 LumA, 8 LumB and 1 normal switched their phenotype (Table 3). Most samples switched their subtype to normal like tumors. Authors reported that genes related to extracellular metabolism, angiogenesis and developmental processes are related to chemoresistance in these patients[14]. In another dataset (GSE32072), where gene expression data for 21 cancer samples was available for pre- and post-neoadjuvant therapy (9 patients anthracycline based therapy, 8 patients anthracycline/taxane based therapy and 2 patients trastuzumab therapy). Additionally, 4 patients were also included which were not treated with any neoadjuvant therapy as control. Upon PAM50 subtyping, patients who did not receive any neoadjuvant chemotherapy, retained their classes (2 basal and 2 LumA). Interestingly, all basal (n=5), Her2 (n=4) and LumB (1) retained their subtype. Out of 6 pre-therapy lumA samples only one changed to LumB subtype. Normal subtype samples changed the most to other subtypes in this dataset (N=4) and only one sample retained its phenotype (Table 4). Authors reported enrichment of energy and metabolism related genes in residual samples while immune related genes were depleted[15].

Breast cancer metastasis can be molecularly distinct from their primary tissues[16]. Metastatic samples can switch to another PAM50 class when compared with primary tissue class. In dataset GSE125989, 16 matched tissues were available for both brain metastasis and primary site. These samples were PAM50 subtyped. 50% (8) of them retained their phenotype when primary site vs metastasis PAM50 classes were compared. One Basal changed to Her2 and 1 Her2 changed to basal. Major subtype switching was observed in Normal primary tissue phenotype, where 3 metastasized tissues became basal and 2 became HER2 (Table 5). Authors showed that when gene expression of metastasized tissue was compared with primary tissue, mesenchymal markers were up regulated while epithelial markers were downregulated. Additionally tumor infiltrating lymphocytes, B cells and dendritic cell gene signatures were downregulated in metastatic tissues. Dataset GSE110590 contains gene expression for 16 primary breast cancer tissues and their 67 matched metastasis. 55% (37/67) metastatic tissues retained the same PAM50 class when compared with primary site. Most stable classes were basal where 85% (23/27) metastatic sites retained basal phenotype and HER2 100% (4/4). While LumA showed 32% (10/31) and Normal showed 0% (0/5) consistecy. Most primary samples when metastasized switched to Basal and LumB class (Table 6).

In another study, authors compared 123 patients' primary tissues with metastatic tissues for PAM50 switching. Metastatic tissues included skin, lymph node, liver, bone, lung, peritoneum and others. Subtype concordance for Basal was highest (100%), followed by HER2 (76.92%) and LumB (70%). The most switchable subtype was LumA where 55.32% sample switched to luminal B and HER2[7].

Above analysis suggests PAM50 subtype switch between different classes both due to chemotherapy treatment and metastasis. Therfore subtyping for residual cancer or metastasis should be re-evaluated. Additionally, basal phenotype is the most stable phenotype and LumA is the most unstable one. Moreover PAM50 classification is subjected to sampling method as we show that different sampling techniques can result in different classification of the same samples.

Biomarkers for metastasis

To identify biomarkers for metastasis, we used 7 cohorts for which metastasis-free survival data was reported. Treatment protocol and patient numbers for each cohort are described in methods section. We

selected 10 good and 10 bad prognostic genes. All selected 20 genes and their respective probeset IDs are shown in figure 1.

Our selected genes can predict metastasis-free survival or distant metastasis-free survival in several cohorts. Next we used these genes to cluster patients in metastatic prone or resistant groups with the aim to identify enriched gene sets among these using GSEA[17]. Surprisingly, no common gene sets enriched in metastatic prone and resistant groups were found, suggesting that different mechanisms like EMT, inflammation, cell cycle and immune response are involved in metastasis (data not shown).

Good prognostic genes

Among the 20 genes selected, 10 were good prognostic. Epoxide hydrolase 2 (EPHX2) is largely involved in metabolism of epoxy fatty acids and plays an important part in cardiovascular biology, renal function, inflammation response and pain[18]. Previously, EPHX2 has been associated with metastasis suppression in breast cancer and its expression was four-fold higher in lymph node negative tumors as compared with lymph node positive tumors[19]. In another meta-analysis for breast cancer metastasis, EPHX2 was identified as metastasis suppressor gene and its biological functions were associated with drug metabolism, inflammatory response and reactive oxygen species metabolism[20]. So, it might be possible that EPHX2's diverse role in metabolism of carcinogens might play a role in metastasis of breast cancer. Previously high glucose levels were associated with increased breast cancer proliferation and metastasis[21] and interestingly specificity protein 1 (sp1) is shown to downregulate EPHX2 under increased glucose levels[22]. EPHX2 role has also been associated with recurrence in breast cancer[23].

Thymocyte nuclear protein 1 (THYN1), also known as THY28, is a highly conserved protein from bacteria to mammals and plays an important role in B cell development[24] and its downregulation has been associated with induction of apoptosis in Ramos B lymphoma cells[25]. However its role in breast cancer is not well studied. Previously, another study on breast cancer metastasis has associated its role as good prognosis marker as they used the same cohort GSE6532 (also used in this study). In another study, TYN1 role was associated with recurrence in breast cancer[23]. THYN1 is also negatively correlated with numerical chromosomal heterogeneity in NCI60 cancer cell lines. In this study, authors showed that chromosomal instability is associated with EMT, cancer invasiveness and metastasis[26].

ADP ribosylation factor like GTPase 6 interacting protein 5 (ARL6IP5), also known as JWA, has been shown previously as microtubule associated protein and has been related to differentiation, apoptosis and response to stress stimulators and its expression is mediated by ER[27]. In NIH3T3 and HELF cell lines, increased H_2O_2 levels resulted in increased expression of ARL6IP5 expression indicating its role in oxidative stress[28]. In breast cancer, knockdown of ARL6IP5, resulted in increased invasion and migration along with decreased apoptosis in MDA-MB-231 cells[29]. In another study, authors showed that expression of CXCR4 (breast cancer metastatic marker) was inversely proportional to ARL6IP5 and ARL6IP5 decreases breast cancer migration and invasion by proteomic degradation of CXCR4[30]. Moreover, ARL6IP5 has been shown to suppress metastasis of melanoma cells as well as cancer cell adhesion, invasion and metastasis in gastric cells, similar to our analysis[31-33].

CAMP responsive element binding protein like 2 (CREBL2) is identified as a metabolic regulator[34]. CREBL2 was shown to be expressed by T47D cell line when compared with MDA-MB-435 (a highly metastatic cell line)[35]. CREBL2 expression is decreased in inflammatory breast cancer (an aggressive subtype of breast cancer)[36]. CREBL2 was found to be downregulated in OVCAR paclitaxel resistant cell line (ovarian cancer cell line) when compared with controls[37]. CREBL2 was also found downregulated among grade 3 breast cancer samples[38]. In another study, to identify genes differentially expressed

between patients who suffered from distant metastasis and ones who did not, CREBL2 was downregulated in metastatic group[39].

Purinergic receptor P2Y13 (P2RY13) is a G-protein-coupled receptor and it is downregulated upon EGF- and hypoxia- induced EMT in breast cancer cells[40,41]. Additionally, overexpression of P2RY13 is associated with better recurrence free survival in hepatocellular carcinoma[42].

Alpha-L-Fucosidase 1(FUCA1) is a lysosomal enzyme involved in cleavage of terminal fructose residue. Its lower expression has been associated with aggressive tumors such as breast, neuroblastoma and colorectal cancer. Moreover its reduced expression is related to shorter recurrence and cancer free survival in lymph node positive breast cancer luminal B patients[43]. Its lower expression is also related to metastasis in solid tumors[44]. One study showed that FUCA1 functions downstream of p53 and can contribute to EGFR signaling repression by removing fucose from EGFR and thus is a marker of lower survival in breast cancer[45]. Among triple negative breast cancer patients, FUCA1 downregulation is associated with decreased survival through modulation of glycosylation status[46].

Kinesin family member 13B (KIF13B) belongs to Kinesin 3 family. KIF13B is downregulated upon estrogen stimulation of breast cancer cell proliferation[47]. And its expression is upregulated in paclitaxel resistant MCF7 cells when compared with control cells[48]. Although KIF13B is not studied well in cancer, its role was emphasized as VEGFR2 trafficker in endothelial cells previously. Thus it plays an important role in angiogenesis and its depletion leads to reduced VEGF induced capillary tube formation, endothelial cells migration and neovascularization in mice[49].

Progesterone receptor (PGR) status has been described in previous sections as it goes through plasticity during metastasis. In one study, authors found 46% tumors shifting from PR positive status to negative status in metastatic tissue, although in the same study they also saw the opposite trend for 16% tumors as well[50]. In another study, 40.4% discordance was found between primary and metastatic tissue regarding PGR status and 74% discordance was due to loss of PGR status[51].

DIX domain containing 1 (DIXDC1) gene is a regulator of Wnt signaling pathway and its expression has been associated with good prognosis in breast cancer as well as metastasis repression in lung adenocarcinoma as expected[52,53]. Knockdown of DIXDC1 in glioma cells has been shown to inhibit proliferation and migration of these cells[54]. Interestingly lower expression of DIXDC1 has been associated with worse outcomes in hepatocellular carcinoma[55] and intestinal type gastric carcinoma[56].

Joining chain of multimeric IgA and IgM (JCHAIN), previous symbol IGJ, is downregulated in ductal carcinoma *in situ* and invasive ductal carcinoma when compared with normal stroma[57]. Additionally, JCHAIN was previously shown to be downregulated in both tumor stroma and tumor epithelium in inflammatory breast cancer (aggressive form of breast cancer related to metastasis) when compared with non-inflammatory breast cancer[36].

Bad prognostic genes

Serine dehydratase (SDS) converts L-serine to pyruvate and ammonia. It has not been studied extensively in breast cancer. In one study, its expression was shown to be higher in tumor samples when compared with normal samples[58].

N-Alpha-Acetyltransferase 10, NatA Catalytic Subunit (NAA10), also known as ARD1, is related to post-translational modifications and has been previously shown to increase breast cancer proliferation upon overexpression[59]. Its expression is increased in breast tumor cells when compared with normal adjacent

cells[60]. In one study, authors related its expression with lymph node metastasis in breast cancer[60]. Surprisingly in another study, it was reported as the inhibitor of metastasis by inhibiting cell motility[61].

Centromere protein E (CENPE), with chromosome misalignment being the hallmark of cancer, aids the proliferation of tumor cells via its involvement in capturing and positioning the chromosomes to mitotic spindle[62,63]. The inhibition of CENP-E causes the death and remediation of triple negative breast cancer cells[64,65].

Argonaute-2 (AGO2) plays an indispensable role in the post-transcriptional downregulation of gene expression through guided small-RNAs. It is one of the four members of the Argonaute protein family which is highly conserved amongst different species and is the only member ofto have catalytic activity and therefore, is the most important for gene silencing along with the RISC complex[66,67]. AGO2 has been linked to different types of cancers previously[68,69]. AGO2 expression was shown to increase in Her2 and basal subtype. Additionally increase in AGO2 expression showed reduced relapse free survival. In the same research it was indicated that cancer cell-lines showed weak AGO2 expression as opposed to the tissue samples, suggesting a role of the tumor microenvironment in AGO2 expression [70]. AGO2 has also been shown to be highly expressed in a subset of ERα+ breast tumors and is correlated with poor prognosis in patients. It was shown that the endocrine resistance of ER+/PGR- tumors were highly correlated with AGO2 overexpression[71]. Cells transfected with AGO2 were shown to have a more aggressive phenotype with higher proliferative capabilities and higher expression of pro-oncogenic microRNAs[72].

E2F8 is the eighth member of the E2F transcription factors, which are expressed in a variety of tissues and is involved in regulatory functions of cellular proliferation, differentiation, cell apoptosis, DNA repair & cell cycle [73,74]. E2F8 plays a major role in cell-cycle progression, especially to the G1 to S phase transition by binding to E2F1, repressing its expression and therefore inhibiting its effects [75]. This property has been reported to affect breast cancer tumorigenesis as well. It accompolishes this by transcriptionally activating CCNE1 and CCNE2 leading to G1/S phase transition resulting in increased proliferation [76]. E2F8 has also been shown to be overexpressed in patient data obtained from TCGA. Patients that had higher expression of E2F8 showed significantly poorer prognosis[77]. Additionally it has been recently shown that post-transcriptional regulation of E2F8 by RNA binding proteins in breast cancer cell lines, increases the proliferative capabilities of breast cancer[78]. It is clear that E2F8 overexpression increases the invasive properties of breast cancer and results in poorer prognosis of breast cancer patients when overexpressed.

BUB1 Mitotic Checkpoint Serine/Threonine Kinase B (BUB1B) has a significant relation with both stage and grade of the tumor. Overexpression of BUB1B is more often noticed in high grade breast cancer and it is correlated with shorter distant metastasis-free survival [79]. BUB1B's expression values have been shown to solely divide the breast cancer patients into metastatic and non-metastatic [80].

Transmembrane Protein 208 (TMEM208) codes for a transmembrane protein located on endoplasmic reticulum. TMEM208 overexpression has been shown to inhibit ER stress and autophagy[81]. Also, TMEM208 transcription is affected by oxygen levels, especially by hypoxic conditions under which HIF-1α binds to its promoter, resulting in an increase in transcript levels[82]. Hypoxic conditions have been shown to induce invasiveness of breast cancer cells under the orchestration of HIF-1α[83].

Proteasome 26S Subunit, Non-ATPase 2 (PSMD2) codes for a protein part of the proteasome subunit of the ubiquitin proteasome system (UPS).This gene is upregulated in breast cancer cells as compared with normal adjacent tissue by more than 3 fold[84]. High levels of PSMD2 correlate with shorter over all survival and distant metastasis free survival and the gene is also part of a metastatic signature that relates

to poorer prognosis[85,86]. Silencing of PSMD2 inhibits breast cancer cell proliferation, tumor growth and increases apoptosis[85].

Tubulin Alpha 1b (TUBA1B) is a member of tubulin family proteins, which are major components of microtubules. TUBA1B is involved in a diverse spectrum of biological activities including cell adhesion, movement, replication and division. Although its role is undetermined in breast cancer and metastatic progression, it was shown to be associated with poor survival in liver cancer [87].

Ubiquitin Conjugating Enzyme E2 A (UBE2A), also known as RAD6A, is a ubiquitin conjugating enzyme which is involved in protein metabolism. Studies showed that UBE2A is upregulated in aggressive breast cancers and its expression correlates with metastasis [88 89]. In one study, authors also revealed that increased RAD6 levels are associated with cancer stem cell phenotype in ovarian cancer resulting in disease recurrence and metastasis [90].

Conclusion

Our results suggest that PAM50 plasticity exists in response to treatment. Moreover, metastasized tissue shows different PAM50 class when compared with primary tissue. So, there is a need to reassess the classes in case of relapse and metastasis. For example patients with HER2 enriched phenotype are treated with targeted therapy, but if these tumors switch to a luminal phenotype upon relapse or metastasis, then a second line of treatment can be added to deal with this switch[7].

Additionally, our suggested metastatic biomarkers can be used to better predict metastatic events in breast cancer patients. Previously in literature several studies have identified metastatic biomarkers in breast cancer[91-93]. But here our markers show prognostic importance in 7 different cohorts.

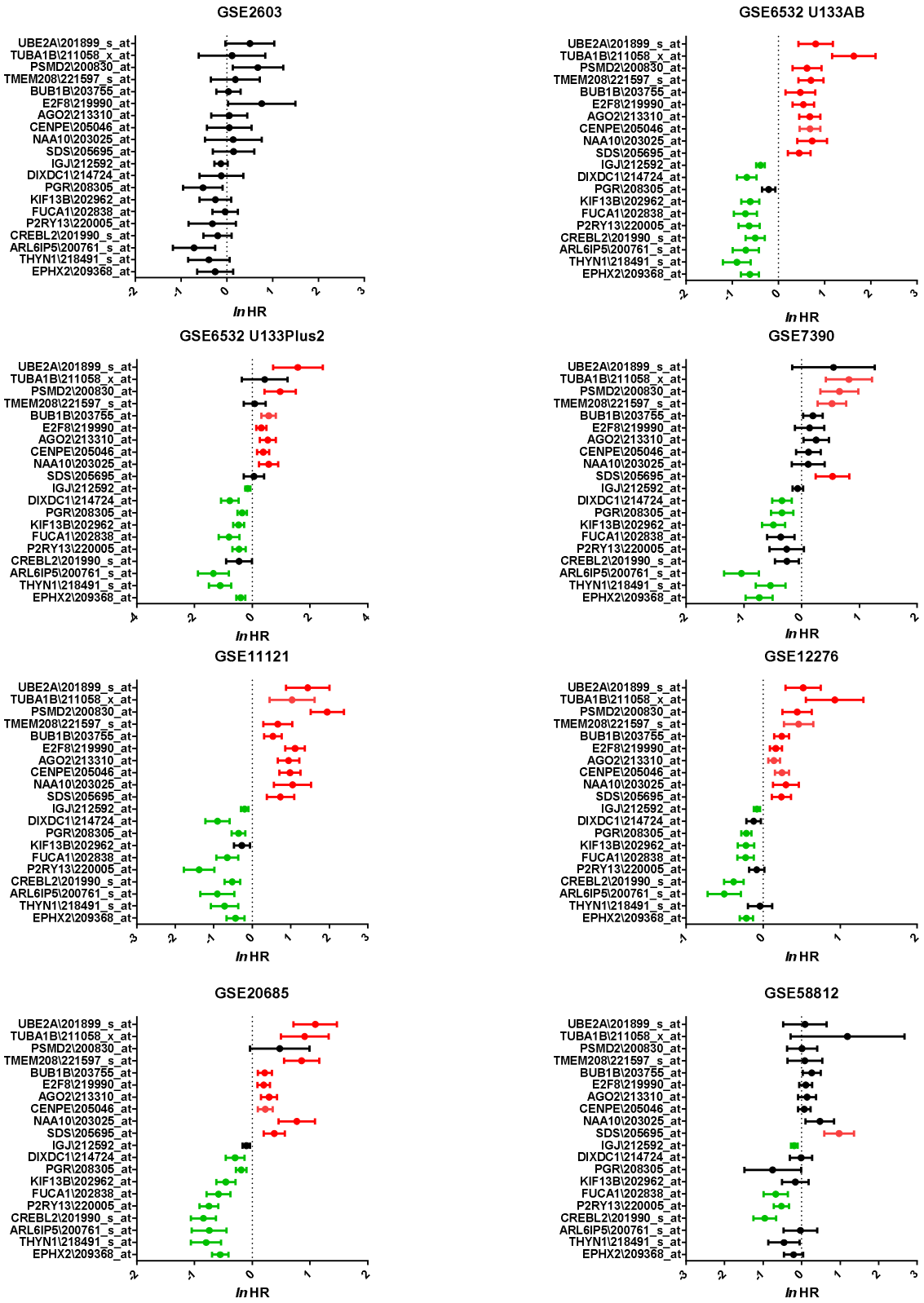
GSE2603
GSE6532 U133AB
GSE6532 U133Plus2
GSE7390
GSE11121
GSE12276
GSE20685
GSE58812
UBE2A\201899_s_at
TUBA1B\211058_x_at
PSMD2\200830_at
TMEM208\221597_s_at
BUB1B\203755_at
E2F8\219990_at
AGO2\213310_at
CENPE\205046_at
NAA10\203025_at
SDS\205695_at
IGJ\212592_at
DIXDC1\214724_at
PGR\208305_at
KIF13B\202962_at
FUCA1\202838_at
P2RY13\220005_at
CREBL2\201990_s_at
ARL6IP5\200761_s_at
THYN1\218491_s_at
EPHX2\209368_at
ln HR

Figure 1: Metastasis-free survival analysis: 20 genes (10 good prognostic, 10 bad prognostic) were selected as metastasis biomarkers. Gene names and probe IDs are shown on y axis. Y axis shows Hazard ratio (HR) in natural logarithm with 95% CI. Genes which are good prognostic and significant are shown in as green while bad prognostic and significant genes are shown in red color. If a gene was not significant it is shown in black.

Table 1: Comparison of PAM50 subtyping for samples collected through fine needle aspiration and core needle biopsy (GSE32518).

		Fine needle aspiration				
		Basal	HER2	LumA	LumB	Normal
Core needle biopsies	Basal	7	0	0	0	0
	HER2	1	5	0	1	0
	LumA	0	0	9	1	0
	LumB	0	0	0	7	0
	Normal	4	0	1	0	1

Table 2: Comparison of PAM50 subtypes pre- and post-Letrozole therapy (GSE10281).

		Post-therapy				
		Basal	HER2	LumA	LumB	Normal
Pre-therapy	Basal	2	0	0	0	0
	HER2	1	1	0	2	0
	LumA	0	0	2	3	0
	LumB	0	0	0	1	0
	Normal	3	0	1	1	1

Table 3: Comparison of PAM50 subtypes pre- and post-Anthracyclines and Taxanes based neoadjuvant chemotherapy (GSE32518).

		Post-therapy				
		Basal	HER2	LumA	LumB	Normal
Pre-therapy	Basal	2	0	0	0	2
	HER2	1	1	0	1	1
	LumA	0	0	3	0	4
	LumB	1	0	4	3	3
	Normal	0	0	1	0	1

Table 4: Comparison of PAM50 subtypes pre- and post- neoadjuvant chemotherapy (GSE32072).

		Post-therapy				
		Basal	HER2	LumA	LumB	Normal
Pre-therapy	Basal	5	0	0	0	0
	HER2	0	4	0	0	0
	LumA	0	0	5	1	0
	LumB	0	0	0	1	0
	Normal	2	1	1	0	1

Table 5: Comparison of PAM50 subtyping for primary tissues vs. brain metastatic tissues (GSE125989).

		Metastasized tissue				
		Basal	HER2	LumA	LumB	Normal
Primary tissue	Basal	2	1	0	0	0
	HER2	1	2	0	0	1
	LumA	0	0	4	0	0
	LumB	0	0	0	0	0
	Normal	3	2	0	0	0

Table 6: Comparison of PAM50 subtyping for primary tissues vs. metastatic tissues (GSE110590).

		PAM50 metastasis				
Sample name	Primary tissue	Basal	HER2	LumA	LumB	Normal
A11	Basal	5	0	0	0	0
A15	Basal	2	0	0	0	3
A1	Basal	4	0	0	0	1
A23	Basal	5	0	0	0	0
A5	Basal	2	0	0	0	0
A7	Basal	5	0	0	0	0
A8	HER2	0	4	0	0	0
A12	LumA	0	0	2	4	0
A17	LumA	2	0	0	0	0
A26	LumA	3	0	0	0	0
A28	LumA	0	0	3	3	0
A2	LumA	0	1	0	2	0
A30	LumA	5	0	0	0	0
A34	LumA	0	0	4	0	0
A4	LumA	0	1	1	0	0
A20	Normal	5	0	0	0	0

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 2018;68:394-424.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA: a cancer journal for clinicians* 2019;69:7-34.
3. Harbeck N, Penault-Llorca F, Cortes J, et al. Breast cancer. *Nature Reviews Disease Primers* 2019;5:66.
4. Cardoso F, Senkus E, Costa A, et al. 4th ESO–ESMO international consensus guidelines for advanced breast cancer (ABC 4). *Annals of Oncology* 2018;29:1634-1657.
5. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging biological principles of metastasis. *Cell* 2017;168:670-691.
6. Sari E, Guler G, Hayran M, et al. Comparative study of the immunohistochemical detection of hormone receptor status and HER-2 expression in primary and paired recurrent/metastatic lesions of patients with breast cancer. *Medical Oncology* 2011;28:57-63.
7. Cejalvo JM, de Dueñas EM, Galván P, et al. Intrinsic subtypes and gene expression profiles in primary and metastatic breast cancer. *Cancer research* 2017;77:2213-2221.
8. Akbar MW, Isbilen M, Belder N, et al. A Stemness and EMT Based Gene Expression Signature Identifies Phenotypic Plasticity and is A Predictive but Not Prognostic Biomarker for Breast Cancer. *Journal of Cancer* 2020;11:949.
9. Simon R, Lam A, Li M-C, et al. Analysis of gene expression data using BRB-array tools. *Cancer informatics* 2007;3:117693510700300022.
10. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology* 2009;27:1160.
11. Team RC. R Foundation for Statistical Computing; Vienna, Austria: 2015. *R: A language and environment for statistical computing* 2018:2013.
12. Therneau T, Grambsch P. Modeling survival data: Extending the Cox model (pp. 69–75): New York, NY: Springer Science & Business Media, 2013.
13. Creighton CJ, Li X, Landis M, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proceedings of the National Academy of Sciences* 2009;106:13820-13825.
14. Vera-Ramirez L, Sanchez-Rovira P, Ramirez-Tortosa CL, et al. Transcriptional shift identifies a set of genes driving breast cancer chemoresistance. *PLoS One* 2013;8.
15. Gonzalez-Angulo AM, Iwamoto T, Liu S, et al. Gene expression, molecular class changes, and pathway analysis after neoadjuvant systemic therapy for breast cancer. *Clinical Cancer Research* 2012;18:1109-1119.
16. Vecchi M, Confalonieri S, Nuciforo P, et al. Breast cancer metastases are molecularly distinct from their primary tumors. *Oncogene* 2008;27:2148-2158.
17. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 2005;102:15545-15550.
18. Tanaka H, Kamita SG, Wolf NM, et al. Transcriptional regulation of the human soluble epoxide hydrolase gene EPHX2. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 2008;1779:17-27.
19. Ellsworth R, Valente A, Field L, et al. Abstract P4-06-09: Genetic Signature Discriminating Metastatic from Non-Metastatic Small Tumors: AACR, 2010.

20. Thomassen M, Tan Q, Kruse TA. Gene expression meta-analysis identifies chromosomal regions and candidate genes involved in breast cancer metastasis. *Breast cancer research and treatment* 2009;113:239-249.

21. Sun S, Sun Y, Rong X, et al. High glucose promotes breast cancer proliferation and metastasis by impairing angiotensinogen expression. *Bioscience reports* 2019;39.

22. Oguro A, Oida S, Imaoka S. Down-regulation of EPHX2 gene transcription by Sp1 under high-glucose conditions. *Biochemical Journal* 2015;470:281-291.

23. Naoi Y, Kishi K, Tanei T, et al. Development of 95-gene classifier as a powerful predictor of recurrences in node-negative and ER-positive breast cancer patients. *Breast cancer research and treatment* 2011;128:633-641.

24. Fujita T, Kitaura F, Fujii H. A critical role of the Thy28-MYH9 axis in B cell-specific expression of the Pax5 gene in chicken B cells. *PLoS One* 2015;10.

25. Jiang X, Toyota H, Yoshimoto T, et al. Anti-IgM-induced down-regulation of nuclear Thy28 protein expression in Ramos B lymphoma cells. *Apoptosis* 2003;8:509-519.

26. Roschke AV, Glebov OK, Lababidi S, et al. Chromosomal instability is associated with higher expression of genes implicated in epithelial-mesenchymal transition, cancer invasiveness, and metastasis and with lower expression of genes involved in cell cycle checkpoints, DNA repair, and chromatin maintenance. *Neoplasia (New York, NY)* 2008;10:1222.

27. Chen R, Li A, Zhu T, et al. JWA—A novel environmental-responsive gene, involved in estrogen receptor-associated signal pathway in MCF-7 and MDA-MB-231 breast carcinoma cells. *Journal of Toxicology and Environmental Health, Part A* 2005;68:445-456.

28. Chen R, Qiu W, Liu Z, et al. Identification of JWA as a novel functional gene responsive to environmental oxidative stress induced by benzo [a] pyrene and hydrogen peroxide. *Free Radical Biology and Medicine* 2007;42:1704-1714.

29. Chen X, Feng J, Ge Z, et al. Effects of the JWA gene in the regulation of human breast cancer cells. *Molecular medicine reports* 2015;11:3848-3853.

30. Xu L, Cheng L, Yang F, et al. JWA suppresses the invasion of human breast carcinoma cells by downregulating the expression of CXCR4. *Molecular medicine reports* 2018;17:8137-8144.

31. Lu J, Tang Y, Cheng Y, et al. ING4 regulates JWA in angiogenesis and their prognostic value in melanoma patients. *British journal of cancer* 2013;109:2842-2852.

32. Chen Y, Xia X, Wang S, et al. High FAK combined with low JWA expression: clinical prognostic and predictive role for adjuvant fluorouracil–leucovorin–oxaliplatin treatment in resectable gastric cancer patients. *Journal of gastroenterology* 2013;48:1034-1044.

33. Bai J, Zhang J, Wu J, et al. JWA regulates melanoma metastasis by integrin α V β 3 signaling. *Oncogene* 2010;29:1227-1237.

34. Tiebe M, Lutz M, Tiebe DS, et al. Crebl2 regulates cell metabolism in muscle and liver cells. *Scientific Reports* 2019;9:1-12.

35. Mellick AS, Day CJ, Weinstein SR, et al. Differential gene expression in breast cancer cell lines and stroma–tumor differences in microdissected breast cancer biopsies revealed by display array analysis. *International journal of cancer* 2002;100:172-180.

36. Boersma BJ, Reimers M, Yi M, et al. A stromal gene signature associated with inflammatory breast cancer. *International journal of cancer* 2008;122:1324-1332.

37. Duan Z, Lamendola DE, Duan Y, et al. Description of paclitaxel resistance-associated genes in ovarian and breast cancer cell lines. *Cancer chemotherapy and pharmacology* 2005;55:277-285.

38. Cava C, Bertoli G, Ripamonti M, et al. Integration of mRNA expression profile, copy number alterations, and microRNA expression levels in breast cancer to improve grade definition. *PloS one* 2014;9.

39. Myhre S, Mohammed H, Tramm T, et al. In silico ascription of gene expression differences to tumor and stromal cells in a model to study impact on breast cancer outcome. *PLoS One* 2010;5.

40. Davis FM, Kenny PA, Soo ET, et al. Remodeling of purinergic receptor-mediated Ca2+ signaling as a consequence of EGF-induced epithelial-mesenchymal transition in breast cancer cells. *PloS one* 2011;6.

41. Azimi I, Beilby H, Davis FM, et al. Altered purinergic receptor-Ca2+ signaling associated with hypoxia-induced epithelial-mesenchymal transition in breast cancer cells. *Molecular oncology* 2016;10:166-178.

42. Maynard JP, Lee J-S, Sohn BH, et al. P2X3 purinergic receptor overexpression is associated with poor recurrence-free survival in hepatocellular carcinoma patients. *Oncotarget* 2015;6:41162.

43. Bonin S, Parascandolo A, Aversa C, et al. Reduced expression of α-L-Fucosidase-1 (FUCA-1) predicts recurrence and shorter cancer specific survival in luminal B LN+ breast cancer patients. *Oncotarget* 2018;9:15228.

44. Daves MH, Hilsenbeck SG, Lau CC, et al. Meta-analysis of multiple microarray datasets reveals a common gene signature of metastasis in solid tumors. *BMC medical genomics* 2011;4:56.

45. Ezawa I, Sawai Y, Kawase T, et al. Novel p53 target gene FUCA 1 encodes a fucosidase and regulates growth and survival of cancer cells. *Cancer science* 2016;107:734-745.

46. Cheng T-C, Tu S-H, Chen L-C, et al. Down-regulation of α-L-fucosidase 1 expression confers inferior survival for triple-negative breast cancer patients by modulating the glycosylation status of the tumor cell surface. *Oncotarget* 2015;6:21283.

47. Zou JX, Duan Z, Wang J, et al. Kinesin family deregulation coordinated by bromodomain protein ANCCA and histone methyltransferase MLL for breast cancer cell growth, survival, and tamoxifen resistance. *Molecular Cancer Research* 2014;12:539-549.

48. Park S, Kim H, Ji HW, et al. Cold Atmospheric Plasma Restores Paclitaxel Sensitivity to Paclitaxel-Resistant Breast Cancer Cells by Reversing Expression of Resistance-Related Genes. *Cancers* 2019;11:2011.

49. Yamada KH, Nakajima Y, Geyer M, et al. KIF13B regulates angiogenesis through Golgi to plasma membrane trafficking of VEGFR2. *J Cell Sci* 2014;127:4518-4530.

50. Aurilio G, Disalvatore D, Pruneri G, et al. A meta-analysis of oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 discordance between primary breast cancer and metastases. *European journal of cancer* 2014;50:277-289.

51. Amir E, Miller N, Geddie W, et al. Prospective study evaluating the impact of tissue confirmation of metastatic disease in patients with breast cancer. *Journal of clinical oncology* 2012;30:587.

52. Mao XY, Lee MJ, Zhu J, et al. Genome-wide screen identifies a novel prognostic signature for breast cancer survival. *Oncotarget* 2017;8:14003.

53. Goodwin JM, Svensson RU, Lou HJ, et al. An AMPK-independent signaling pathway downstream of the LKB1 tumor suppressor controls Snail1 and metastatic potential. *Molecular cell* 2014;55:436-450.

54. Chen J, Shen C, Shi J, et al. Knockdown of DIXDC1 inhibits the proliferation and migration of human glioma cells. *Cellular and molecular neurobiology* 2017;37:1009-1019.

55. Zhou S, Shen J, Lin S, et al. Downregulated expression of DIXDC1 in hepatocellular carcinoma and its correlation with prognosis. *Tumour Biol* 2016;37:13607-13616.

56. Wang L, Tan C, Qiao F, et al. Upregulated expression of DIXDC1 in intestinal-type gastric carcinoma: co-localization with beta-catenin and correlation with poor prognosis. *Cancer Cell Int* 2015;15:120.

57. Ma XJ, Dahiya S, Richardson E, et al. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Res* 2009;11:R7.

58. Mehrmohamadi M, Liu X, Shestov AA, et al. Characterization of the usage of the serine metabolic network in human cancer. *Cell Rep* 2014;9:1507-1519.

59. Yu M, Ma M, Huang C, et al. Correlation of expression of human arrest-defective-1 (hARD1) protein with breast cancer. *Cancer Invest* 2009;27:978-983.

60. Yu M, Gong J, Ma M, et al. Immunohistochemical analysis of human arrest-defective-1 expressed in cancers in vivo. *Oncol Rep* 2009;21:909-915.

61. Zeng Y, Min L, Han Y, et al. Inhibition of STAT5a by Naa10p contributes to decreased breast cancer metastasis. *Carcinogenesis* 2014;35:2244-2253.

62. El-Arabey AA, Salama SA, Abd-Allah AR. CENP-E as a target for cancer therapy: Where are we now? *Life Sci* 2018;208:192-200.

63. Yardimci H, van Duffelen M, Mao Y, et al. The mitotic kinesin CENP-E is a processive transport motor. *Proc Natl Acad Sci U S A* 2008;105:6016-6021.

64. Kung PP, Martinez R, Zhu Z, et al. Chemogenetic evaluation of the mitotic kinesin CENP-E reveals a critical role in triple-negative breast cancer. *Mol Cancer Ther* 2014;13:2104-2115.

65. Ohashi A, Ohori M, Iwai K, et al. A Novel Time-Dependent CENP-E Inhibitor with Potent Antitumor Activity. *PLoS One* 2015;10:e0144675.

66. Meister G. Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* 2013;14:447-459.

67. Fischer KR, Durrans A, Lee S, et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* 2015;527:472-476.

68. Voller D, Reinders J, Meister G, et al. Strong reduction of AGO2 expression in melanoma and cellular consequences. *Br J Cancer* 2013;109:3116-3124.

69. Zhang H, Wang Y, Dou J, et al. Acetylation of AGO2 promotes cancer progression by increasing oncogenic miR-19b biogenesis. *Oncogene* 2019;38:1410-1431.

70. Casey MC, Prakash A, Holian E, et al. Quantifying Argonaute 2 (Ago2) expression to stratify breast cancer. *BMC Cancer* 2019;19:712.

71. Conger AK, Martin EC, Yan TJ, et al. Argonaute 2 Expression Correlates with a Luminal B Breast Cancer Subtype and Induces Estrogen Receptor Alpha Isoform Variation. *Noncoding RNA* 2016;2.

72. Adams BD, Claffey KP, White BA. Argonaute-2 expression is regulated by epidermal growth factor receptor and mitogen-activated protein kinase signaling and correlates with a transformed phenotype in breast cancer cells. *Endocrinology* 2009;150:14-23.

73. Logan N, Graham A, Zhao X, et al. E2F-8: an E2F family member with a similar organization of DNA-binding domains to E2F-7. *Oncogene* 2005;24:5000-5004.

74. Lv Y, Xiao J, Liu J, et al. E2F8 is a Potential Therapeutic Target for Hepatocellular Carcinoma. *J Cancer* 2017;8:1205-1213.

75. Kim LK, Park SA, Eoh KJ, et al. E2F8 regulates the proliferation and invasion through epithelial-mesenchymal transition in cervical cancer. *Int J Biol Sci* 2020;16:320-329.

76. Ye L, Guo L, He Z, et al. Upregulation of E2F8 promotes cell proliferation and tumorigenicity in breast cancer by modulating G1/S phase transition. *Oncotarget* 2016;7:23757-23771.

77. Li Y, Huang J, Yang D, et al. Expression patterns of E2F transcription factors and their potential prognostic roles in breast cancer. *Oncol Lett* 2018;15:9216-9230.

78. Iino K, Mitobe Y, Ikeda K, et al. RNA-binding protein NONO promotes breast cancer proliferation by post-transcriptional regulation of SKP2 and E2F8. *Cancer Sci* 2020;111:148-159.

79. Cai Y, Mei J, Xiao Z, et al. Identification of five hub genes as monitoring biomarkers for breast cancer metastasis in silico. *Hereditas* 2019;156:20.

80. Mansouri N, Movafagh A, Sayad A, et al. Targeting of BUB1b Gene Expression in Sentinel Lymph Node Biopsies of Invasive Breast Cancer in Iranian Female Patients. *Asian Pac J Cancer Prev* 2016;17:317-321.

81. Zhao Y, Hu J, Miao G, et al. Transmembrane protein 208: a novel ER-localized protein that regulates autophagy and ER stress. *PLoS One* 2013;8:e64228.

82. Lei P, Xiao Y, Li P, et al. hSnd2/TMEM208 is an HIF-1alpha-targeted gene and contains a WH2 motif. *Acta Biochim Biophys Sin (Shanghai)* 2020.

83. Ren J, Guo H, Wu H, et al. GPER in CAFs regulates hypoxia-driven breast cancer invasion in a CTGF-dependent manner. *Oncol Rep* 2015;33:1929-1937.

84. Deng S, Zhou H, Xiong R, et al. Over-expression of genes and proteins of ubiquitin specific peptidases (USPs) and proteasome subunits (PSs) in breast cancer tissue observed by the methods of RFDD-PCR and proteomics. *Breast Cancer Res Treat* 2007;104:21-30.

85. Li Y, Huang J, Zeng B, et al. PSMD2 regulates breast cancer cell proliferation and cell cycle progression by modulating p21 and p27 proteasomal degradation. *Cancer Lett* 2018;430:109-122.

86. Tomida S, Yanagisawa K, Koshikawa K, et al. Identification of a metastasis signature and the DLX4 homeobox protein as a regulator of metastasis by combined transcriptome approach. *Oncogene* 2007;26:4600-4608.

87. Lu C, Zhang J, He S, et al. Increased alpha-tubulin1b expression indicates poor prognosis and resistance to chemotherapy in hepatocellular carcinoma. *Dig Dis Sci* 2013;58:2713-2720.

88. Shekhar MP, Lyakhovich A, Visscher DW, et al. Rad6 overexpression induces multinucleation, centrosome amplification, abnormal mitosis, aneuploidy, and transformation. *Cancer Res* 2002;62:2115-2124.

89. Haynes B, Zhang Y, Liu F, et al. Gold nanoparticle conjugated Rad6 inhibitor induces cell death in triple negative breast cancer cells by inducing mitochondrial dysfunction and PARP-1 hyperactivation: Synthesis and characterization. *Nanomedicine* 2016;12:745-757.

90. Somasagara RR, Spencer SM, Tripathi K, et al. RAD6 promotes DNA repair and stem cell signaling in ovarian cancer and is a promising therapeutic target to prevent and treat acquired chemoresistance. *Oncogene* 2017;36:6680-6690.

91. Cabioglu N, Yazici MS, Arun B, et al. CCR7 and CXCR4 as novel biomarkers predicting axillary lymph node metastasis in T1 breast cancer. *Clinical Cancer Research* 2005;11:5686-5693.

92. Shaashua L, Shabat-Simon M, Haldar R, et al. Perioperative COX-2 and β-adrenergic blockade improves metastatic biomarkers in breast cancer patients in a phase-II randomized trial. *Clinical Cancer Research* 2017;23:4651-4661.

93. Millar E, Graham PH, O'Toole SA, et al. Prediction of local recurrence, distant metastases, and death after breast-conserving therapy in early-stage invasive breast cancer using a five-biomarker panel. *J Clin Oncol* 2009;27:4701-4708.