International Journal for Research in Health Science and Nursing
(ISSN: 2208- 2670)
Vol. 12, Issue 02, June 2026
DOI: 10.65327/hsn.v12i2.2491

CROSS-CELL-LINE PHARMACOGENOMIC MAPPING OF CONSERVED PACLITAXEL RESISTANCE SIGNATURES IN BREAST CANCER

Prof. Raafat Abd El-Gawad El-Awady[1], Dr. Afsheen Raza[2]

[1]Department of Pharmacy Practice and Pharmacotherapeutics, College of Pharmacy, University of Sharjah, Sharjah, United Arab Emirates. Email: relawady@sharjah.ac.ae

[2]Department of Biomedical Sciences, College of Health Sciences, Abu Dhabi University; Director, Abu Dhabi University Cancer Research Institute, Abu Dhabi, United Arab Emirates. Email: raza.afsheen@adu.ac.ae

Received: 08/02/2026 **Revised:** 22/05/2026 **Acceptance:** 13/06/2026 **Published:** 27/06/2026

ABSTRACT-

Durable therapeutic results are difficult to achieve in breast cancer patients because of resistance to paclitaxel, which can develop via heterogeneous molecular mechanisms. This work created a publicly available pharmacogenomic dataset to map conserved resistance-associated gene-expression signatures in four breast cancer cell lines: BAS, HS578T, MCF7 and MDA. After removing duplications and consolidating genes at the level, 20,160 observations were analyzed. The candidates were defined as high-magnitude genes ($P < 0.05$ and $|logFC| \geq 1$) and cross-cell-line conservation was assessed using overlap, directional concordance, and Spearman correlation. Overall, the number of observations with high magnitudes ranged from 4.14% (HS578T) to 33.17% (MDA). Average pairwise Jaccard similarity was low (1.71–8.59%) and expression correlations were from −0.098 to 0.204, showing high cell-line specificity. Of the 215 genes shared in common across all models 28 consistently expressed in the same direction. CPE, USP18, FAT4, DUSP1, and TAPBPL were consistently found to be downregulated, while LYPD3, SDC1, MYBL2 and HK2 were consistently found to be upregulated. These results suggest that the signature can be used to define a focused functional signature for validation and highlight the importance of the directionality of the signatures when defining conserved paclitaxel-resistance biomarkers.

Keywords – Breast cancer, Paclitaxel resistance, Pharmacogenomics, Gene expression, Drug resistance

International Journal for Research in Health Science and Nursing
(ISSN: 2208- 2670)
Vol. 12, Issue 02, June 2026
DOI: 10.65327/hsn.v12i2.2491

1. Introduction

Breast cancer is very molecularly and therapeutically heterogeneous. Paclitaxel is a broad-spectrum drug used in breast cancer treatment as it stabilizes the microtubules, inhibits the progression of the mitoses and induces apoptosis of cancer cells. But intrinsic or acquired resistance can diminish the effectiveness of treatment, allow for the persistence of disease, and lead to disease recurrence. Resistance does not rely on a single molecular event; it can involve alterations in drug transport, microtubule regulation, escape from apoptosis, DNA-repair capacity, metabolic adaptation, and changes in the tumor microenvironment, all of which can contribute to resistance (Abouzeid et al., 2025). Therefore, knowledge of molecular signals shared among resistant models is crucial for the development of valid biomarkers and therapeutic strategies.

Another layer of complexity in breast cancer resistance is the differences between estrogen receptor (ER) positive, HER2 positive, and triple negative tumors. Specific cell lines have been chosen to reflect some of this diversity and provide systems in which to study drug response mechanisms. However, what is learned from one model might be a cell-line specific process of resistance not representative of a widely shared resistance mechanism. Recent literature suggests that paclitaxel treatment can trigger interferon-mediated effects and uncover cell cycle-related resistance mechanisms in triple-negative breast cancer (TNBC) cell models (Calistri et al., 2025). More comprehensive studies also define resistance as a developing phenotype that is driven by tumour plasticity, pathway redundancy, and treatment selection (Malik & Masood, 2022). These observations enable comparative analysis in different cellular backgrounds.

Gene-expression profiling offers a practical way to find out genes that are expressed differently in resistant phenotypes. Differential expression estimates can indicate activation or repression of genes and whether the extent and direction of the change is replicated in experimental models. Drug metabolism-related, tubulin functioning, proliferation and DNA repair-related candidate markers could have predictive and/or prognostic value in breast cancer (Tsyganov et al., 2022). In addition, functional screening has identified other non-coding regulators, such as micro RNAs, as being involved in paclitaxel resistance in triple-negative breast cancer (Nemours et al., 2024). Therefore, resistance signatures are molecular products to be interpreted as a multi-layered effect rather than as the effect of a single gene.

Pharmacogenomic analysis can correlate molecular variation with drug response, and can prioritize biomarkers for systems-pharmacology decisions (Kardamiliotis et al., 2022). Genes identified in multiple resistant models might be more reproducible than genes found in a single model in a cross-cell-line approach. Directional concordance is particularly significant as it has a more clear biological interpretation if one direction of the genes is upregulated or the other direction is upregulated. On the other hand, inconsistent directions might represent contextual regulation, subtype effects, or compensation. Pattern recognition methods using computational models can capture clinically relevant signals from genomic and mutational data, but these signals may be obscured by tumor heterogeneity, sample size, or model transferability (Patterson et al., 2023).

In the field of cancer, AI and data-driven approaches are being used for classification, biomarker identification, drug response prediction, and therapeutic decision-making (Murmu & Győrffy, 2024). They are only reliable if they have clearly defined preprocessing, easily understood analyses,

International Journal for Research in Health Science and Nursing
(ISSN: 2208- 2670)
Vol. 12, Issue 02, June 2026
DOI: 10.65327/hsn.v12i2.2491

and meaningful selection of biological features. AI has opened up new avenues for personalized cancer care and posed challenges in terms of data quality and validation (Ali et al., 2024). The same constraints have been identified in cancer-management applications, especially when combining different kinds of data sets without proper external validation (Ali et al., 2023). Thus, it is important to map the expression patterns of resistance before entering into complicated predictive modelling, as it will help create a map that is interpretable.

The prediction of effective anticancer combination drugs is a growing trend in computational drug research as it may be useful for overcoming anticancer drug resistance by targeting complementary pathways (Peng et al., 2024). AI has also been used in cancer diagnostics and therapeutic development, aiding the integration of molecular markers into clinical decision-making processes (Singh et al., 2024). Despite these advancements, there are still needs to establish reproducible input into oncology applications and to differentiate between the statistical association and actionable mechanisms (Vobugari et al., 2022). Such inputs and to direct functional experiments, combination screening or patient-level validation would be provided by expression-based signatures that are maintained consistently in various cellular contexts.

The search for molecules using transcriptomic data can be enhanced by orthogonal validation. As such, the need for specificity and reproducibility when developing expression panels is especially relevant in other biomedical applications, as shown by sensitive gene-signature strategies (Mucaki et al., 2022). Protein microarrays and other proteomic platforms can identify if a protein's abundance is changed, whether the pathway is altered, or whether it is a protein interaction that has been targeted by a drug (Chen et al., 2023). Additionally, the complex and context-dependent nature of both progression and therapeutic adaptation is reflected in cross-cancer molecular studies, which calls for cautious translation of model-derived signatures (Sinner, 2023).

Therefore, in the present study, data on gene expression in the paclitaxel-resistant cell lines of the breast cancer cell lines BAS, HS578T, MCF7, and MDA were analyzed. The analysis aims to separate conserved signals from cell-line-specific responses by comparing the magnitude of the signals, the direction of the response, the overlap of signals across the multiple gene sets, and the concordance of the response across models. This design fills an important gap between discovery of single models and reproducible interpretation of pharmacogenomics. This also serves as a basis for the identification of resistance candidates suitable for mechanistic investigation and therapeutic development in the future. The specific objectives were:

1. To characterize differential gene-expression patterns associated with paclitaxel resistance across four breast cancer cell lines
2. To identify conserved and directionally concordant resistance signatures through cross-cell-line pharmacogenomic overlap analysis
3. To prioritize reproducible candidate genes for mechanistic validation and future resistance-informed therapeutic development

2. Methodology

2.1 Dataset and Research Design

International Journal for Research in Health Science and Nursing
(ISSN: 2208- 2670)
Vol. 12, Issue 02, June 2026
DOI: 10.65327/hsn.v12i2.2491

This secondary bioinformatics study used the uploaded Kaggle-derived dataset containing processed differential-expression results from paclitaxel-resistant and corresponding control breast-cancer cells (Madiseh, 2023). Four cell-line groups were represented: BAS, HS578T, MCF7 and MDA, with MDA interpreted as MDA-MB-231 according to the source description. Each record contained a gene symbol, nominal P value, log fold change and cell-line label. The analytical design systematically compared transcriptomic resistance patterns among cell lines and searched for recurrent genes that could represent shared molecular signatures of paclitaxel resistance.

2.2 Data Preparation and Cleaning

The compressed archive was imported into Python and the CSV structure was verified before analysis. Gene symbols and cell-line labels were trimmed to remove unintended spaces, while P values and log fold changes were retained as numeric variables. Missingness, exact duplication and repeated gene–cell-line combinations were assessed. Exact duplicate rows were removed first. When multiple records represented the same gene within one cell line, the record with the smallest P value was retained together with its corresponding log fold change, producing one statistically strongest observation per gene and cell line.

2.3 Expression Classification and Candidate Selection

Genes were classified as upregulated when log fold change exceeded zero and downregulated when it was below zero. Because all deposited records already satisfied nominal significance below 0.05, an additional magnitude criterion of absolute log fold change at least 1.0 was applied to define high-magnitude candidate genes. A more restrictive sensitivity category required P below 0.01 and absolute log fold change at least 2.0. Counts, percentages, median P values and median absolute log fold changes were calculated separately for every breast-cancer cell line.

2.4 Cross-Cell-Line Overlap Analysis

Cell-line-specific candidate sets were compared using pairwise intersections and unions. The Jaccard percentage was calculated as the number of overlapping high-magnitude genes divided by the combined number of unique genes in each pair. Directional concordance represented the percentage of overlapping genes showing the same sign of log fold change in both cell lines. Genes appearing in all four cleaned datasets were examined separately. A conserved signature required presence across every cell line and a uniformly positive or uniformly negative expression direction, avoiding causal interpretation of the observed associations.

2.5 Statistical Analysis

Descriptive analyses were performed in Python using pandas, NumPy and SciPy. Spearman rank correlations quantified similarity in log fold-change patterns among genes shared by each cell-line pair, with two-sided P values reported for correlation testing. Analyses emphasized effect magnitude, overlap and directional consistency because adjusted P values and raw sample-level expression matrices were unavailable. Numerical precision was limited to meaningful decimal places. No machine-learning model was fitted because the dataset contained gene-level summary statistics rather than independent biological samples with resistant and sensitive class labels.

2.6 Ethical Considerations

International Journal for Research in Health Science and Nursing
(ISSN: 2208- 2670)
Vol. 12, Issue 02, June 2026
DOI: 10.65327/hsn.v12i2.2491

The investigation exclusively reanalysed publicly accessible, de-identified gene-expression summaries and involved no recruitment, intervention or access to individual patient information. Consequently, additional institutional ethics committee approval and informed consent were not required for this secondary analysis. The original dataset sources and repository should nevertheless be cited in the manuscript. Results were interpreted as exploratory transcriptomic associations rather than experimentally validated resistance mechanisms. Candidate genes identified through cross-cell-line consistency require confirmation using independent expression cohorts, functional assays and prospectively generated paclitaxel-response data before any translational application.

3. Results

3.1 Dataset Characteristics and Data Quality

The uploaded file contained 21,312 records describing 12,951 distinct gene symbols across four paclitaxel-resistant breast-cancer cell lines. No values were missing in any variable. Cleaning removed 59 completely duplicated rows and consolidated 1,093 additional repeated gene–cell-line observations, leaving 20,160 unique analytical records. MCF7 contributed the largest cleaned set with 6,147 genes, followed by BAS with 6,241, HS578T with 3,961 and MDA with 3,811. Median P values ranged from 0.00154 in BAS to 0.00904 in MDA. The final dataset retained 94.59% of original records (Table 1) (Figure 1).

Table 1. Dataset characteristics following gene-level harmonization

Cell line	Original records	Retained records	Removed records	Missing values	Median P value	Median logFC	Median absolute logFC
BAS	6,250	6,241	9	0	0.00154	−0.374	0.574
HS578T	3,964	3,961	3	0	0.00734	0.351	0.414
MCF7	6,424	6,147	277	0	0.00270	0.219	0.377
MDA	4,674	3,811	863	0	0.00904	0.400	0.749
Total	21,312	20,160	1,152	0	0.00405	0.267	0.496

International Journal for Research in Health Science and Nursing
(ISSN: 2208- 2670)
Vol. 12, Issue 02, June 2026
DOI: 10.65327/hsn.v12i2.2491

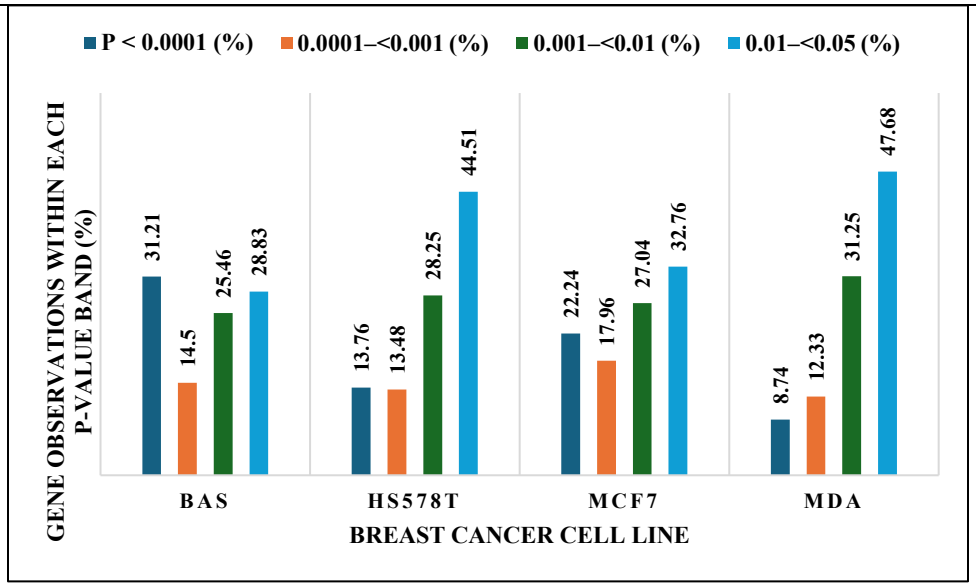


Figure 1. Distribution of statistical-evidence bands across cell lines

3.2 Directional Expression Patterns

Expression direction differed substantially among the four cellular models. BAS displayed predominantly negative changes, with 3,614 downregulated and 2,627 upregulated genes. Conversely, HS578T contained 2,819 upregulated genes compared with 1,142 downregulated genes. MDA included 2,084 upregulated and 1,727 downregulated genes, while MCF7 included 3,409 and 2,738, respectively. The corresponding upregulated-to-downregulated ratios were 0.73, 2.47, 1.24 and 1.21. Median log fold change ranged from −0.374 in BAS to 0.400 in MDA. These contrasts indicate marked cell-line heterogeneity within the deposited paclitaxel-resistance profiles (Figure 2).

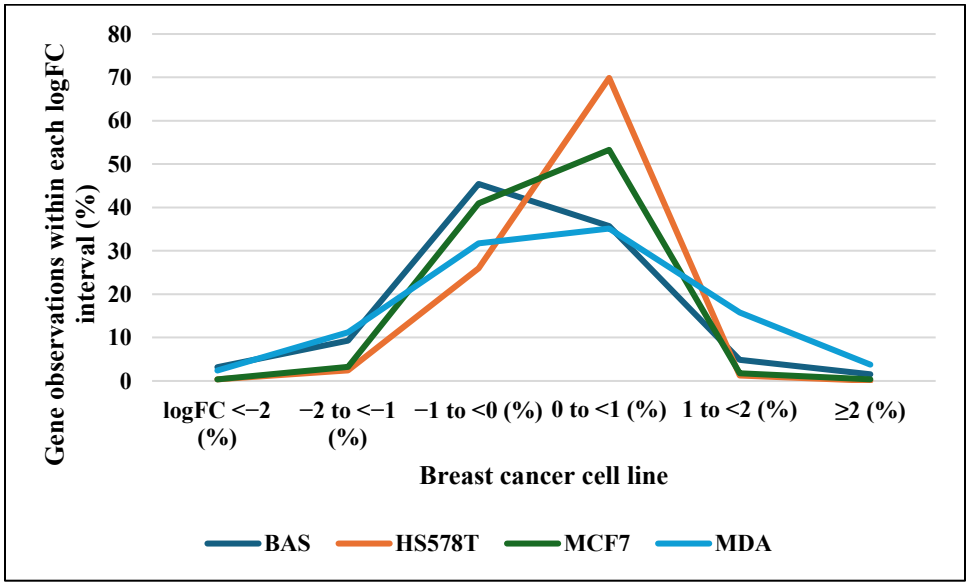


Figure 2. Distribution of logFC magnitude and direction

International Journal for Research in Health Science and Nursing
(ISSN: 2208- 2670)
Vol. 12, Issue 02, June 2026
DOI: 10.65327/hsn.v12i2.2491

3.3 High-Magnitude Expression Signatures

Applying nominal P below 0.05 and absolute log fold change of at least 1 identified 2,958 high-magnitude gene–cell-line records. MDA showed the largest proportional response, with 1,264 candidates representing 33.17% of its cleaned genes. BAS contributed 1,177 candidates, or 18.86%, whereas MCF7 and HS578T contributed 353 and 164, respectively. Downregulation predominated in BAS, HS578T and MCF7, while MDA contained more upregulated than downregulated candidates. The restrictive category included 296 BAS, 18 HS578T, 46 MCF7 and 203 MDA records in total (Table 2) (Figure 3).

Table 2. Distribution of candidate resistance-associated genes

Cell line	Analysed genes	Upregulated genes	Downregulated genes	High-magnitude upregulated	High-magnitude downregulated	Total high-magnitude	High-magnitude (%)	Restrictive candidates
BAS	6,241	2,627	3,614	398	779	1,177	18.86	296
HS578T	3,961	2,819	1,142	51	113	164	4.14	18
MCF7	6,147	3,409	2,738	133	220	353	5.74	46
MDA	3,811	2,084	1,727	745	519	1,264	33.17	203
Total	20,160	10,939	9,221	1,327	1,631	2,958	14.67	563

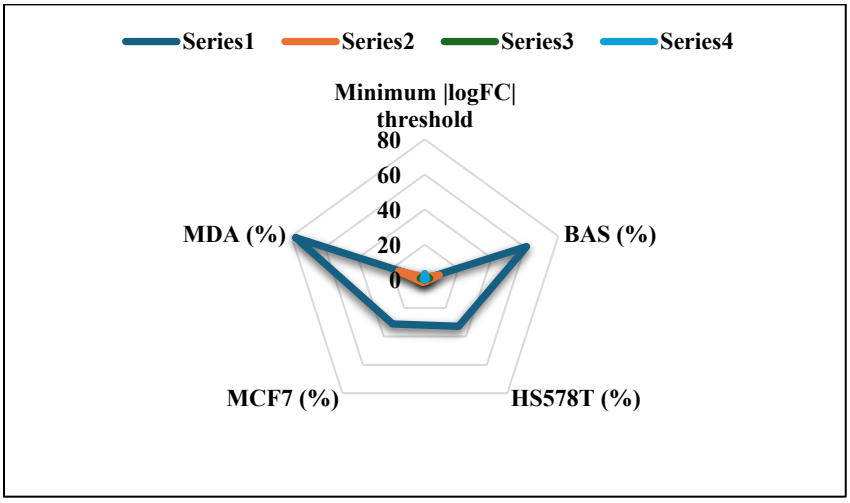


Figure 3. Candidate prevalence across increasing effect-size thresholds

3.4 Pairwise Gene-Set Overlap and Concordance

Pairwise overlap among high-magnitude candidate sets was generally limited. BAS and MDA shared the largest number of genes, with 193 overlapping candidates and a Jaccard percentage of

International Journal for Research in Health Science and Nursing
(ISSN: 2208- 2670)
Vol. 12, Issue 02, June 2026
DOI: 10.65327/hsn.v12i2.2491

8.59%. Their directional agreement reached 66.32%. Other Jaccard values ranged from 1.71% for HS578T–MDA to 3.73% for BAS–MCF7. BAS–HS578T shared 45 candidates, and MCF7–MDA shared 36. The highest directional concordance was observed between BAS and MCF7 at 70.91%, whereas HS578T and MDA showed only 29.17%. These findings demonstrate predominantly cell-line-specific resistance responses overall (Table 3) (Figure 4).

Table 3. Pairwise overlap and expression concordance between cell lines

Cell-line comparison	Shared analysed genes	Spearman r	Correlation P value	High-magnitude overlap	Jaccard (%)	Directionally concordant genes	Directional concordance (%)
BAS–HS578T	1,919	0.062	0.0065	45	3.47	27	60.00
BAS–MCF7	2,204	0.204	<0.001	55	3.73	39	70.91
BAS–MDA	1,630	0.145	<0.001	193	8.59	128	66.32
HS578T–MCF7	1,332	−0.049	0.0714	13	2.58	7	53.85
HS578T–MDA	942	−0.098	0.0025	24	1.71	7	29.17
MCF7–MDA	1,207	0.041	0.1515	36	2.28	19	52.78

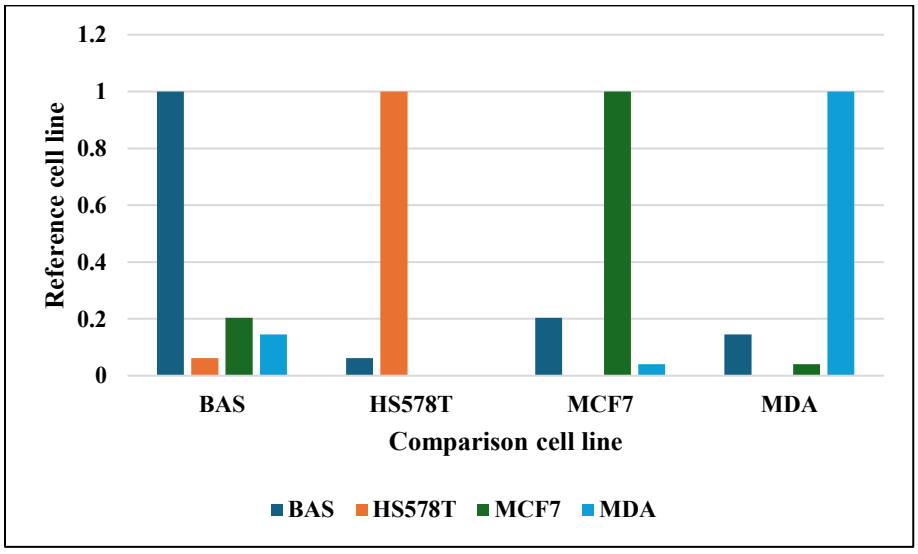


Figure 4. Cross-cell-line expression correlation heatmap

3.5 Conserved Paclitaxel-Resistance Signatures

A total of 215 genes were represented in all four cleaned cell-line datasets. Only 28 maintained a common direction: 18 were consistently upregulated and 10 consistently downregulated. CPE

International Journal for Research in Health Science and Nursing
(ISSN: 2208- 2670)
Vol. 12, Issue 02, June 2026
DOI: 10.65327/hsn.v12i2.2491

showed the largest direction-consistent average absolute log fold change at 1.523, followed by USP18 at 1.182 and FAT4 at 1.154. Only PLXDC2 met the absolute log fold-change threshold in every cell line, but its direction reversed in MDA. Pairwise Spearman coefficients ranged from −0.098 to 0.204, confirming weak overall concordance between resistance profiles (Table 4) (Figure 5).

Table 4. Leading direction-consistent genes shared across all four cell lines

Gene	BAS logFC	HS578T logFC	MCF7 logFC	MDA logFC	Mean absolute logFC	Median logFC	Maximum P value	Direction
CPE	−1.527	−0.702	−0.807	−3.057	1.523	−1.167	0.00007	Down
USP18	−0.919	−0.413	−1.020	−2.376	1.182	−0.970	0.00901	Down
FAT4	−2.057	−0.373	−1.448	−0.737	1.154	−1.092	0.03310	Down
DUSP1	−1.902	−0.589	−0.367	−1.488	1.086	−1.039	0.00073	Down
TAPBPL	−1.767	−0.427	−0.618	−1.427	1.060	−1.023	0.00721	Down
LYPD3	1.545	0.387	1.253	0.939	1.031	1.096	0.01030	Up
SDC1	0.896	0.996	1.118	0.679	0.922	0.946	0.01240	Up
MYBL2	0.990	0.692	1.131	0.800	0.903	0.895	0.00612	Up
HK2	1.369	0.705	0.490	0.993	0.889	0.849	0.00065	Up
CAMK1D	−0.869	−0.324	−0.726	−1.534	0.863	−0.797	0.03310	Down

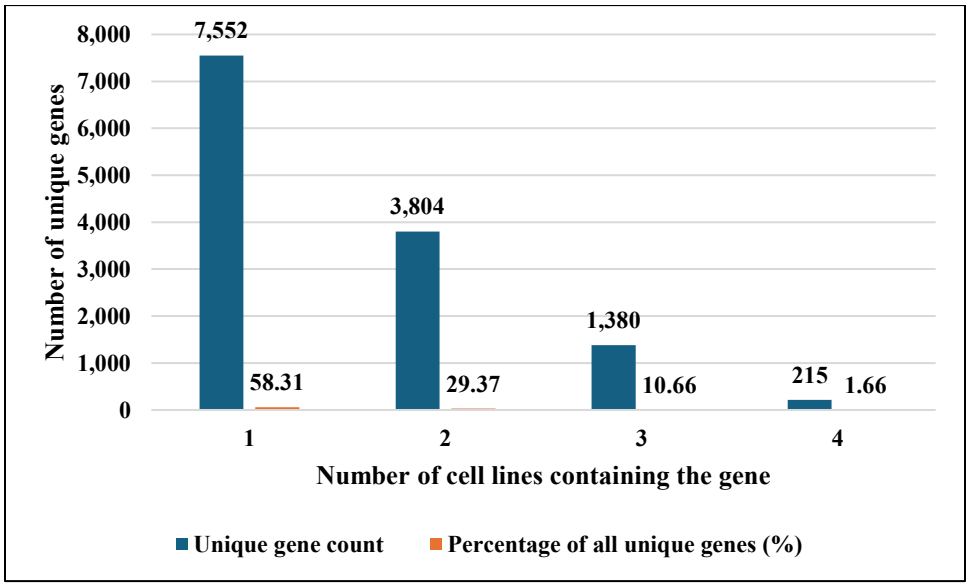


Figure 5. Distribution of genes according to cross-cell-line coverage

International Journal for Research in Health Science and Nursing
(ISSN: 2208- 2670)
Vol. 12, Issue 02, June 2026
DOI: 10.65327/hsn.v12i2.2491

4. Discussion

A cross-cell-line analysis found significant variation in the transcriptional response to paclitaxel resistance. Of the 20,160 gene–cell-line combinations that were cleaned, 2,958 were considered to have high magnitude, $P < 0.05$ and $|\log FC| \geq 1$. The proportion varied sharply, from 4.14% in HS578T and 5.74% in MCF7 to 18.86% in BAS and 33.17% in MDA. The differences indicate that the resistant phenotypes are defined by the molecular programs of the cells, instead of by a common transcriptional response. Similarly, in tamoxifen resistant breast cancer cells, remodelling of the phenotype was apparent from changes in lectin reactivity that went beyond known markers of resistance (Hlaing et al., 2022). Tamoxifen and paclitaxel have different mechanisms of action, but both show the importance of the biological diversity of acquired drug resistance.

There was also variation in the balance of up- and downregulation of the candidates across the four models. There were more genes with high magnitude changes that were downregulated in BAS than there were upregulated genes, while more genes with high magnitude changes were upregulated in MDA. This directional change could be attributed to differences in the molecular subtype, basal transcription levels, resistance-selection history, and/or activation of compensatory pathways. Similar analysis of breast cancer cells that are resistant to three chemotherapeutic drugs revealed a number of proteins that were common to the resistant cells and others that were unique on each drug, which again showed that adaptation to chemotherapy can occur in different molecular combinations (Cha et al., 2022). The present transcriptomic findings further illustrate this: even resistance to the same agent can result in different expression architectures among breast cancer cell lines.

The overlap between the sets of genes of high magnitude was minimal. The Jaccard similarity values for the overlapping candidates ranged from 1.71% to 8.59% and the value of directional concordance ranged from 29.17% to 70.91% between the overlapping candidates. Spearman correlations were also low ranging from −0.098 to 0.204. These findings suggest that the majority of the changes in expression were not conserved and that conservation is not simply a matter of presence of the gene. Large-scale machine-learning analysis of pancreatic cancer treated with the drug gemcitabine has also revealed that drug response is not a single universal marker, but rather complex gene-expression and DNA-methylation signatures (Ogunleye et al., 2024). But that study, though it used a different tumor type and drug, backs the notion of incorporating multiple dimensions of molecules when analyzing resistance signatures.

The four sets of cleaned cell lines had 215 genes in common and only 28 of these had the same direction of expression. Of all genes, only PLXDC2 was detected in all models with $|\log FC| \geq 1$ and had the opposite direction in MDA, which made it not a conserved directional marker. Conversely, the expression of CPE, USP18, FAT4, DUSP1 and TAPBPL was consistently decreased, whereas LYPD3, SDC1, MYBL2 and HK2 were consistently increased. These genes were thus more defensible candidates to validate their functionality even if they showed varying effects in different models. The small set of concordant genes is in line with the fact that microRNAs are part of different anthracycline-resistance pathways in breast cancer (Si et al., 2022). Non-coding RNAs are also emerging as general cancer specificity regulators and therapeutic targets in the broader field of cancer research (Beňačka et al., 2024).

International Journal for Research in Health Science and Nursing
(ISSN: 2208- 2670)
Vol. 12, Issue 02, June 2026
DOI: 10.65327/hsn.v12i2.2491

Part of this heterogeneity could be explained by epigenetic plasticity. In one recent preprint it was reported that extensive DNA-methylation changes occur during the acquisition of paclitaxel resistance in triple-negative breast cancer (TNBC) with some of these changes partially re-sensitized to decitabine (Trnkova et al., 2025). The lack of cross-cell-line correlation seen in the present analysis may thus represent alternative epigenetic pathways to a similar resistant phenotype. Epigenetic regulation of pathways governing cell death mechanisms has also been linked to cancer-cell survival and therapeutic escape (Beato et al., 2022). Epigenetic responses are generally determined by the interaction of multiple biological and environmental exposures, and expression signatures are therefore context dependent (Zhang et al., 2024).

The results also point to the need for an examination of conserving coding-gene signals in the future in combination with the regulatory networks of RNA. In gastric cancer, the activity of long non-coding RNA families may affect proliferation, apoptosis, invasion, and treatment response, with regulatory mechanisms that have the potential to be common to other tumor types that need to be validated for the specific disease (Liu et al., 2024). The methylation data, non-coding RNA data, proteomic data and functional perturbation data can be integrated together to help clarify whether the prioritized genes are actively promoting paclitaxel resistance or simply following resistant states. This validation is especially relevant for the genes CPE, USP18, FAT4, DUSP1, LYPD3, SDC1, MYBL2 and HK2 that consistently showed a direction, but different strength.

There are a few caveats to be taken into account. The analysis was performed on processed gene-level statistics, not on raw expression matrices, so no batch correction normalization and re-evaluation nor no multivariable modelling. The dataset was not accompanied with detailed experimental metadata and there was no patient-level validation, and nominal P values may allow for the false-positive retention of candidates. However, duplicate removal, thresholds for effect sizes, sensitivity analysis, assessment of overlaps, and directional comparison ensured that the candidates were prioritized in a transparent manner. To validate the conserved signature in independent paclitaxel-resistant models, false discovery-rate control should be applied to raw data, candidate genes tested experimentally, and the ability of combined molecular panels to accurately predict resistance compared to individual markers should be assessed.

5. Conclusion

Cross-cell-line pharmacogenomic mapping revealed that paclitaxel resistance in breast cancer is based primarily on the signature of heterogeneous rather than homogeneous transcriptional responses. Pairwise gene-set overlap and expression correlations were generally small and the number of high-magnitude expression changes was highly variable in the four different types of cells (BAS, HS578T, MCF7, MDA). These results suggest that there are different molecular pathways leading to the resistance to paclitaxel in individual breast cancer models. However, a small conserved component was determined. There were 28 genes that were consistently co-regulated between all four cell types, consistently downregulated being CPE, USP18, FAT4, DUSP1, and TAPBPL and consistently upregulated being LYPD3, SDC1, MYBL2, and HK2. These candidates are candidates to be validated to a molecular panel for further validation. While the overlap is necessary for defining conserved signatures, PLXDC2 did demonstrate high-magnitude expression in all models and changed direction in MDA, demonstrating the need for a directional consistency along with overlap. Overall, the analysis provides a clear framework to differentiate signals of broad reproducibility from alterations that are specific to the cell-line. The

prioritized genes need to be confirmed through independent transcriptomic datasets, false-discovery-rate-controlled analyses, protein-level measurements and functional experiments to determine if these genes are directly involved in the regulation of resistance. Ultimately, validated multi-gene signatures can assist in the development of biomarkers, in resistance-informed drug selection, and in rational combination therapy to improve the response to paclitaxel in breast cancer.

References

1. Abouzeid, H. A., Kassem, L., Liu, X., & Abuelhana, A. (2025). Paclitaxel resistance in breast cancer: Current challenges and recent advanced therapeutic strategies. Cancer treatment and research communications, 43, 100918.
2. Ali, M., Wani, S. U. D., Dey, T., & Mehdi, S. (2024). Artificial intelligence in the treatment of cancer: changing patterns, constraints, and prospects. Health and Technology, 14(3), 417-432.
3. Ali, M., Wani, S. U. D., Khan, A. A., Dey, T., & Salahuddin, M. (2023). Current status of artificial intelligence in cancer management: emerging trends, limitations and future outlook.
4. Beato, A., Della Torre, L., Capone, V., Carannante, D., Favale, G., Verrilli, G., ... & Carafa, V. (2022). Epi-regulation of cell death in cancer. In Cell Death and Disease. IntechOpen.
5. Beňačka, R., Szabóová, D., Guľašová, Z., & Hertelyová, Z. (2024). Non-Coding RNAs in breast cancer: Diagnostic and therapeutic implications. International Journal of Molecular Sciences, 26(1), 127.
6. Calistri, N. L., Liby, T. A., Hu, Z., Zhang, H., Dane, M. A., Gross, S. M., & Heiser, L. M. (2025). TNBC response to paclitaxel phenocopies interferon response which reveals cell cycle-associated resistance mechanisms. Scientific Reports, 15(1), 4294.
7. Cha, H. K., Cheon, S., Kim, H., Lee, K. M., Ryu, H. S., & Han, D. (2022). Discovery of proteins responsible for resistance to three chemotherapy drugs in breast cancer cells using proteomics and bioinformatics analysis. Molecules, 27(6), 1762.
8. Chen, G., Yang, L., Liu, G., Zhu, Y., Yang, F., Dong, X., ... & Li, B. (2023). Research progress in protein microarrays: Focussing on cancer research. PROTEOMICS–Clinical Applications, 17(1), 2200036.
9. Hlaing, M. T., Horimoto, Y., Denda-Nagai, K., Fujihira, H., Noji, M., Kaji, H., ... & Irimura, T. (2022). Tamoxifen-resistant breast cancer cells exhibit reactivity with Wisteria floribunda agglutinin. Plos one, 17(8), e0273513.
10. Kardamiliotis, K., Karanatsiou, E., Aslanidou, I., Stergiou, E., Vizirianakis, I. S., & Malousi, A. (2022). Unraveling drug response from pharmacogenomic data to advance systems pharmacology decisions in tumor therapeutics. Future Pharmacology, 2(1), 31-44.
11. Liu, S., Rong, Y., Tang, M., Zhao, Q., Li, C., Gao, W., & Yang, X. (2024). The functions and mechanisms of long non-coding RNA SNHGs in gastric cancer. Combinatorial chemistry & high throughput screening, 27(18), 2639-2653.

12. Madiseh, A. A. (2023). Taxol drug resistance cell lines in breast cancer [Data set]. Kaggle. https://www.kaggle.com/datasets/aliabedimadiseh/taxol-drug-resistance-cell-lines-in-breast-cancer
13. Malik, S. S., & Masood, N. (2022). Emerging Paradigm of Breast Cancer Resistance and Treatment. In Breast Cancer: From Bench to Personalized Medicine (pp. 519-536). Singapore: Springer Nature Singapore.
14. Mucaki, E. J., Shirley, B. C., & Rogan, P. K. (2022). Improved radiation expression profiling in blood by sequential application of sensitive and specific gene signatures. International Journal of Radiation Biology, 98(5), 924-941.
15. Murmu, A., & Győrffy, B. (2024). Artificial intelligence methods available for cancer research. Frontiers of Medicine, 18(5), 778-797.
16. Nemours, S., Solé, C., Goicoechea, I., Armesto, M., Arestin, M., Urruticoechea, A., ... & Lawrie, C. H. (2024). Use of Gain-of-Function Screening to Identify miRNAs Involved in Paclitaxel Resistance in Triple-Negative Breast Cancer. International Journal of Molecular Sciences, 25(24), 13630.
17. Ogunleye, A., Piyawajanusorn, C., Ghislat, G., & Ballester, P. J. (2024). Large-scale machine learning analysis reveals DNA methylation and gene expression response signatures for gemcitabine-treated pancreatic cancer. Health data science, 4, 0108.
18. Patterson, A., Elbasir, A., Tian, B., & Auslander, N. (2023). Computational methods summarizing mutational patterns in cancer: promise and limitations for clinical applications. Cancers, 15(7), 1958.
19. Peng, Z., Ding, Y., Zhang, P., Lv, X., Li, Z., Zhou, X., & Huang, S. (2024). Artificial intelligence application for anti-tumor drug synergy prediction. Current Medicinal Chemistry, 31(40), 6572-6585.
20. Si, Z., Zhong, Y., Lao, S., Wu, Y., Zhong, G., & Zeng, W. (2022). The role of miRNAs in the resistance of anthracyclines in breast cancer: a systematic review. Frontiers in Oncology, 12, 899145.
21. Singh, R. S., Masih, G. D., Joshi, R., Sharma, S., Singh, A., & Medhi, B. (2024). Role of artificial intelligence in cancer diagnostics and therapeutics. In Biomarkers in Cancer Detection and Monitoring of Therapeutics (pp. 83-97). Academic Press.
22. Sinner, J. (2023). Deciphering the molecular landscape of metastatic progression in small cell lung cancer: implications for targeted therapies.
23. Trnkova, L., Burikova, M., Soltysova, A., Ficek, A., Plava, J., Cumova, A., ... & Smolkova, B. (2025). Development of Paclitaxel Resistance in Triple-Negative Breast Cancer Is Associated with Extensive DNA Methylation Changes That Are Partially Reversed by Decitabine. bioRxiv, 2025-05.
24. Tsyganov, M. M., Ibragimova, M. K., Garbukov, E. Y., Tsydenova, I. A., Gaptulbarova, K. A., Dolgasheva, D. S., ... & Litviakov, N. V. (2022). Predictive and prognostic significance of mRNA expression and DNA copies aberrations of ERCC1, RRM1, TOP1,

TOP2A, TUBB3, TYMS, and GSTP1 genes in patients with breast cancer. Diagnostics, 12(2), 405.

25. Vobugari, N., Raja, V., Sethi, U., Gandhi, K., Raja, K., & Surani, S. R. (2022). Advancements in oncology with artificial intelligence—a review article. Cancers, 14(5), 1349.
26. Zhang, M., Hu, T., Ma, T., Huang, W., & Wang, Y. (2024). Epigenetics and environmental health. Frontiers of Medicine, 18(4), 571-596.