

Article

maGENEgerZ: An Efficient Artificial Intelligence-Based Framework Can Extract More Expressed Genes and Biological Insights Underlying Breast Cancer Drug Response Mechanism

Turki Turki ^{1,*}  and Y-h. Taguchi ^{2,*} ¹ Department of Computer Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia² Department of Physics, Chuo University, Tokyo 112-8551, Japan

* Correspondence: tturki@kau.edu.sa (T.T.); tag@granular.com (Y.-h.T.)

Abstract: Understanding breast cancer drug response mechanisms can play a crucial role in improving treatment outcomes and survival rates. Existing bioinformatics-based approaches are far from perfect and do not adopt computational methods based on advanced artificial intelligence concepts. Therefore, we introduce a novel computational framework based on an efficient support vector machine (esvm) working as follows: First, we downloaded and processed three gene expression datasets related to breast cancer responding and non-responding to treatments from the gene expression omnibus (GEO) according to the following GEO accession numbers: GSE130787, GSE140494, and GSE196093. Our method esvm is formulated as a constrained optimization problem in its dual form as a function of λ . We recover the importance of each gene as a function of λ , y , and x . Then, we select p genes out of n , which are provided as input to enrichment analysis tools, Enrichr and Metascape. Compared to existing baseline methods, including deep learning, results demonstrate the superiority and efficiency of esvm, achieving high-performance results and having more expressed genes in well-established breast cancer cell lines, including MD-MB231, MCF7, and HS578T. Moreover, esvm is able to identify (1) various drugs, including clinically approved ones (e.g., tamoxifen and erlotinib); (2) seventy-four unique genes (including tumor suppression genes such as TP53 and BRCA1); and (3) thirty-six unique TFs (including SP1 and RELA). These results have been reported to be linked to breast cancer drug response mechanisms, progression, and metastasizing. Our method is available publicly on the maGENEgerZ web server.

Keywords: breast cancer; drug response; gene expression; machine learning; deep learning; AI application in cancer clinical trials

MSC: 92B05; 68T09



Citation: Turki, T.; Taguchi, Y.-h. maGENEgerZ: An Efficient Artificial Intelligence-Based Framework Can Extract More Expressed Genes and Biological Insights Underlying Breast Cancer Drug Response Mechanism. *Mathematics* **2024**, *12*, 1536. <https://doi.org/10.3390/math12101536>

Academic Editor: Tao Zhou

Received: 5 April 2024

Revised: 2 May 2024

Accepted: 12 May 2024

Published: 15 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The ability to elucidate various mechanisms underlying breast cancer drug response and resistance is a critical part of the clinical decision-making process, not only aiding in finding out the potential effectiveness of a drug compound but also spanning to (1) reducing the search space for candidate compounds; (2) having a greater awareness and management of probable adverse reactions before conducting clinical trials [1]; and (3) identifying potential drug targets associated with drug compounds. Studies have been conducted to analyze gene expression data obtained from biological experiments pertaining to breast cancer drug responses to unveil various molecular mechanisms. Du et al. [2] utilized a bioinformatics approach to identify important genes that play a key role in overcoming breast cancer drug resistance, working as follows: First, two gene expression datasets were downloaded from the gene expression omnibus (GEO) database based on GEO accession numbers GSE28694 (Miller and Payne grades 4 and 5) and GSE28826 (Miller and Payne grades 1 and 2). The GSE28694 dataset had 13 samples treated as the drug-sensitive group,

while the GSE28826 dataset had 28 samples treated as the drug-resistant group. Both processed datasets were provided as input to limma to identify 255 differentially expressed genes (DEGs) with $p < 0.05$, assigned to the enrichment analysis tool ClusterProfiler. The protein–protein interaction with the use of random walk identified three genes (i.e., PRC1, GGTL1, and IRS1) that are involved in immune pathways and involved in the breast cancer drug resistance mechanism. Further validation of the importance of these three genes was performed using additional datasets from the GEO and TCGA databases.

Wu et al. [3] utilized a bioinformatics approach to identify a gene signature that aids in predicting neoadjuvant chemotherapy response for breast cancer patients. They downloaded a gene expression dataset from the GEO database according to the GEO accession number GSE25066. The processed dataset had 508 samples, of which 16 were excluded because of missing data, resulting in 492 samples. To perform differential gene expression analysis, limma was applied to drug-resistant tumor samples and drug-sensitive tumor samples, identifying 347 DEGs. Then, they applied limma within the drug-resistant cell line samples against wild-type cell samples to identify 296 DEGs. Then, 36 genes were identified that were common between the 347 and 296 DEGs. The 36 genes were provided as input to enrichment analysis, finding out 12 hub genes considered as a gene signature (HJURP, IFI27, RAD51AP1, EZH2, DNMT3B, SLC7A5, DBF4, USP18, ELOVL5, PTGER3, KIAA1324, and CYBRD1) from the PPI that has been validated to assess its discriminative power using lasso, in which the same GSE25066 dataset was divided into training and validation sets.

Freitas et al. [4] performed a bioinformatics analysis to identify reliable biomarkers for adding carboplatin to the standard anthracycline/taxane treatment, which can aid in identifying triple-negative breast cancer (TNBC) patients achieving a pathologically complete response to neoadjuvant chemotherapy (NAC). Therefore, TNBC patients with expected poor clinical outcomes can be provided with other treatment options. The processed gene expression data were for 66 patients, of whom 33 were treated with carboplatin + paclitaxel (composed of 19 having RD and 14 achieving PCR), while the remaining 33 were treated with paclitaxel (composed of 23 having RD and 10 achieving PCR). In the 33 patients treated with carboplatin + paclitaxel, they applied limma to identify 37 DEGs between RD and PCR, while 27 DEGs were identified between RD and PCR in patients treated only with paclitaxel. Moreover, 24 DEGs were identified between RD and PCR patients among the 66 patients. Then, 10 statistically significant genes (BNIP3, ZBTB16, KCNB1, HAS1, HEMK1, TFF1, PLA2G4F, SNAI1, C5orf38, and GRIN2A) were selected out of the 37 and 27 DEGs, and 3 statistically significant genes (ALDH1A1, MCM2, and CXADR) out of the 27 DEGs. These 13 genes acted as gene signatures, and the reported results demonstrated their feasibility to discriminate between patients with RD and those achieving PCR.

Stevens et al. [5] aimed to unveil the molecular mechanism behind inducing chemotherapy resistance in inflammatory breast cancer (IBC) patients. They had a dataset of 131 samples between IBC and non-IBC patients derived from several profiling methods distributed based on 14 samples using ChIP-seq profiling, 84 samples using RNA-seq profiling, 3 samples using single-cell RNA-seq profiling, and 30 samples using RNA-seq II profiling. The dataset was deposited into the GEO database with accession number GSE163397. Bioinformatics analysis coupled with enrichment analysis revealed that JAK2/STAT3 signaling is a key player in driving chemoresistance in IBC. Therefore, inhibition of JAK2/STAT3 coupled with the use of paclitaxel can overcome therapeutic resistance in IBC patients. Debets et al. [6] performed a bioinformatics analysis coupled with enrichment analysis to identify a molecular signature (ER2, HER4, ER, IGF1R, and Kalirin) predictive of treatment response and resistance in HER2-positive breast cancer patients. Miri et al. [7] performed a bioinformatics analysis to identify critical genes and pathways that play a key role in doxorubicin resistance in breast cancer. They downloaded two gene expression datasets from GEO with GSE24460 and GSE76540 accession numbers. Then, limma was applied to normal and resistant samples of doxorubicin in which 1108 and 3207 DEGs were identified in GSE24460 and GSE76540, respectively. Pearson correlation was performed to select

36 and 406 significant genes in GSE24460 and GSE76540. Gene co-expression network (GCN) analysis was performed to identify 18 and 115 genes in GSE24460 and GSE76540, respectively. Nine genes (ABCB1, MMP1, TCEAL2, AKAP12, PLS3, LDHB, NEFH, CNN3, and VIM) were common between the two datasets and reported to play a key role in doxorubicin resistance. Other studies aimed to unveil various mechanisms leading to drug resistance in breast cancer patients [8–13].

Although current advances in breast cancer drug mechanisms within clinical testing mainly depend on bioinformatics-driven computational approaches with existing off-the-shelf tools, AI-driven computational frameworks are needed to unveil vast biological insights and to properly promote the use of AI in real clinical settings. The availability of such AI tools can help clinical oncologists avoid therapeutic targets associated with poor treatment responses early, thereby advancing clinical understanding and reducing the search space for potential drugs with adverse effects. The novelty of our study is attributed to the following major contributions:

1. We introduce an AI-driven computational approach consisting of efficient support vector machines (esvm) combined with enrichment analysis tools (Enrichr and Metascape), unveiling various molecular mechanisms pertaining to breast cancer drug response [14–16].
2. We downloaded and processed three gene expression datasets pertaining to breast cancer drug response according to the following GEO accession numbers: GSE130787, GSE140494, and GSE196093.
3. Performing an extensive experimental study from biological and classification perspectives, comparing our method against other bioinformatics-based methods (limma [17], sam, *t*-test [18,19], and lasso [20]) and adapted deep learning methods (DeepLIFT [21], DeepSHAP [22], and LRP [23]).
4. Compared to all methods, including deep learning-based methods, experimental results based on enrichment analysis demonstrate that our method (esvm) identified more expressed genes in three well-established breast cancer cell lines, including MD-MB231, MCF7, and HS578T. Moreover, we identified various drugs for breast cancer, including FDA-approved ones such as gemcitabine (Gemzar) and tamoxifen (Nolvadex). Moreover, 74 unique genes were identified, including tumor suppression genes such as TP53, PTEN, BRCA1, and RB1. A total of 36 unique TFs were reported, including SP1, NFKB1, and RELA. All of these have been reported to play a key role in breast cancer drug response and resistance mechanisms. In terms of the running time for learning-based methods, lasso was faster than our method, esvm, both of which were computationally faster than all other deep learning-based methods.
5. Results from a classification perspective demonstrated the superiority of the gene set obtained via our method when coupled with learning algorithms. Specifically, in Dataset1, when balanced accuracy (BAC) is considered, SVM coupled with a gene set from our method achieved a 32.4% performance improvement over the second-best for SVM with all genes (named None) (see Table S4 in the Supplementary Additional File). In Dataset2 using the BAC performance measure (see Table S5 in the Supplementary Additional File), ours, when coupled with SVM, had a 38.1% performance improvement when compared to the second-best for SVM coupled with the gene set from sam. In the last dataset (i.e., Dataset3), as shown in Table S5 of the Supplementary Additional File, SVM coupled with the gene set from our method had a 6.1% performance improvement over the second-best for SVM coupled with the gene set from DeepLIFT. The same holds true when we evaluated the classification performance using lasso as a learning algorithm coupled with a gene set from our method.
6. For reproducibility of the analysis in this study, we made a publicly available implementation of our method, esvm, within the maGENEgerZ web server at <https://aibio.shinyapps.io/maGENEgerZ/>. Moreover, we included the processed datasets within the Supplementary Datasets folder. We also provided a Supplementary maGENEgerZ_Screenshots.docx file to show the use of our web tool.

2. Materials and Methods

2.1. Gene Expression Profiles

In this work, we retrieved three datasets from different gene expression experiments with different GEO accession numbers [24].

2.1.1. GSE130787: Dataset1

For this dataset derived from the gene expression experiment at the GEO database, we had 89 samples and 5267 genes. As a result, we encoded the dataset as an 89×5268 matrix, including drug responses as a column vector. The 89 samples were distributed in terms of treatment into three groups, as follows: Twenty-six samples were used for patients treated with docetaxel, carboplatin, and trastuzumab (TCH). Thirty-eight samples were used for patients treated with docetaxel, carboplatin, trastuzumab, and lapatinib (TCHTy). Twenty-five samples were used for patients with docetaxel, carboplatin, and lapatinib (TCTy). In terms of the distribution of drug responses, thirty-eight BC patients achieved pathological complete response (PCR), while fifty-one BC patients achieved residual disease (RD). The gene expression experiment was performed using the microarray platform Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name version). This dataset is referred to as Dataset1.

2.1.2. GSE140494: Dataset2

For this second dataset derived from the performed gene expression experiment, we had 91 samples and 5313 genes, which were approved as protein-coding genes (PCGs) by domain experts from HUGO Gene Nomenclature Committee (HGNC) Biomart at <https://biomart.genenames.org/> (accessed on 17 March 2023) [25,26]. Thus, we encoded the dataset as a 91×5314 matrix, including a column vector for drug responses. The drug responses were distributed as follows: Nineteen BC patients achieved a resistant response, while seventy-two were sensitive to the treatment (i.e., docetaxel, followed by 5-fluorouracil, epirubicin, and cyclophosphamide (TFEC). The gene expression experiment was performed using the microarray platform of the Affymetrix Human Genome U133 Plus 2.0 Array. This dataset is referred to as Dataset2.

2.1.3. GSE196093: Dataset3

For this third dataset, we had 736 samples and 118 genes. Consequently, we encoded the dataset as a 736×119 matrix, including a column vector for drug responses, in which 256 BC patients achieved a complete response (CR) while 480 had a failed complete response (FCR). We had 11 treatments in which the 736 samples were distributed accordingly as follows: paclitaxel (169), paclitaxel + ABT 888 + carboplatin (63), paclitaxel + AMG-386 (110), paclitaxel + AMG-386 + trastuzumab (18), paclitaxel + MK-2206 (56), paclitaxel + MK-2206 + trastuzumab (31), paclitaxel + neratinib (105), paclitaxel + pembrolizumab (67), paclitaxel + pertuzumab + trastuzumab (43), paclitaxel + trastuzumab (25), and T-DM1 + pertuzumab (49). The gene expression data were performed using a reverse phase protein array (RPPA) microarray at George Mason University. This dataset is referred to as Dataset3. Table 1 provides an overview of the three-studied datasets.

Table 1. Overview of the three studied breast cancer drug response datasets downloaded from the gene expression omnibus database.

Dataset	Samples	Responder	Non-Responder	Genes	Platform	Organism	Experiment Type
GSE130787	89	38	51	5267	GPL6480	Homo sapiens	Expression profiling by array
GSE140494	91	72	19	5313	GPL570	Homo sapiens	Expression profiling by array
GSE196093	736	256	480	118	GPL28470	Homo sapiens	Protein profiling by protein array

2.2. Computational Framework

In Figure 1, we outline the main steps pertaining to our computational framework. In terms of the preprocessing part, biopsy samples were obtained from breast cancer patients. Then, collected samples were prepared and provided to a biological technology, measuring the gene expression levels [27]. In the machine learning part, the input data correspond to a gene expression dataset, where x_i represents the i th sample and y_i is the associated drug response. In our study, y_i is a binary class label (e.g., {pathological complete response (PCR), residual disease (RD)}). The entire samples x_i (for $i = 1, \dots, m$) were encoded as an $m \times n$ matrix, in which m and n are the number of samples and genes, respectively. All drug responses y_i (where $i = 1, \dots, m$) are encoded as a $1 \times m$ column vector. To identify p important genes out of the n genes in which $p \ll n$, we used to find arguments (i.e., $\mathbf{w} = [w_1 \dots w_n]$ and $b \in \mathbb{R}$) that minimize the objective function in Equation (1) subject to the linear constraints as in [28,29]. After solving the optimization problem in Equation (1), weights in \mathbf{w} correspond to the importance of genes, with higher weights indicating how important these genes are. However, the main issue is as follows:

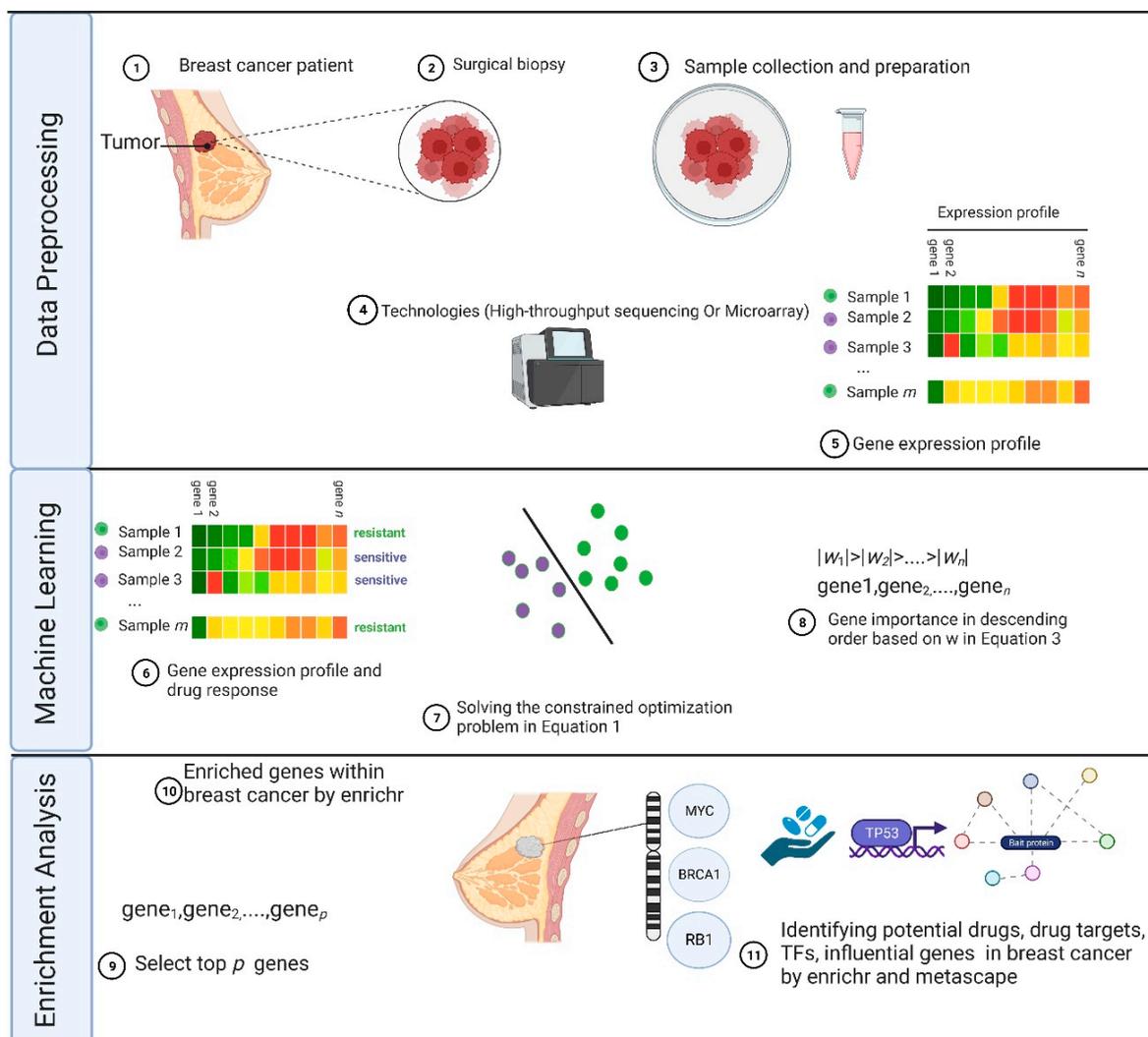


Figure 1. Flowchart of the introduced AI-based framework identifying drugs, drug targets, critical genes, and transcription factors in breast cancer. Figure created with BioRender.com.

The optimization problem in Equation (1) depends on \mathbf{w} and b , where $|\mathbf{w}|$ is equal to n , which is way larger than m (i.e., $m \ll n$) in genomic sciences [30]. That makes the

solution for the optimization problem in Equation (1) computationally expensive [31,32], where the number of genes (n) is typically larger than the number of samples (m).

$$\begin{aligned} & \min_{\mathbf{w}, b} \frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^m \zeta_i^k \\ & \text{subject to } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \zeta_i, \quad i = 1, 2, \dots, m \\ & \zeta_i \geq 0 \end{aligned} \tag{1}$$

Therefore, we seek to solve the dual form of SVM (see Equation (2)) [31]. It can be seen that the optimization problem now depends on finding the Lagrange multiplier (λ), in which each \mathbf{x}_i is associated with λ_i . This is way faster than finding \mathbf{w} in Equation (1) [31].

$$\begin{aligned} & \max_{\lambda} \sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ & \text{subject to } 0 \leq \lambda_i \leq C \text{ and } \sum_{i=1}^m \lambda_i y_i = 0, \quad i = 1, 2, \dots, m \end{aligned} \tag{2}$$

In Equation (3) [31], we recover \mathbf{w} from λ in Equation (2) as follows:

$$\mathbf{w} = \sum_{i=1}^m \lambda_i y_i \mathbf{x}_i \tag{3}$$

For any \mathbf{x}_j associated with $0 < \lambda_j < C$, b is recovered as

$$b = y_j - \sum_{i=1}^m \lambda_i y_i K(\mathbf{x}_i, \mathbf{x}_j) \tag{4}$$

where $K(\mathbf{x}_i, \mathbf{x}_j)$ is a similarity measure K (usually called kernel) of \mathbf{x}_i and \mathbf{x}_j .

A testing example \mathbf{z} is predicted as

$$y' = \text{sign}\left(\sum_{i=1}^m \lambda_i y_i K(\mathbf{x}_i, \mathbf{z}) + b\right) \tag{5}$$

where $\text{sign}()$ is an indicator function mapping to 1 (corresponding to RD) if its argument is greater than or equal to 0. Otherwise, it is mapped to -1 (corresponding to PCR). Equations (S1)–(S5) in the Supplementary Additional File show 5 popular kernels used with SVM [31]. When a linear kernel is used, the prediction model becomes as

$$y' = \text{sign}\left(\sum_{i=1}^m \lambda_i y_i (\mathbf{x}_i \cdot \mathbf{z}) + b\right) \tag{6}$$

We used CVXR in R to solve the dual form of SVM in Equation (2) and to find λ [33]. In terms of the enrichment analysis part, we uploaded the p genes as input to Enricher and Metascape, where these p genes are weighted with the top p weights ($|w_1| > |w_2| > \dots > |w_p|$). Then, we interpret and identify biologically related terms, including key expressed genes, drugs, drug targets, transcription factors, and others, which are provided in the next section.

3. Experiments and Results

3.1. Experimental Methodology

We compared our method, *esvm*, against the following baseline methods: linear models for microarray data (*limma*) [17], significance analysis of microarrays (*sam*), Student's *t*-test (*t*-test), and least absolute shrinkage and selection operator (*lasso*) [18]. The input to the five studied methods is labeled gene expression data. Because the prediction in our efficient SVM-based model (named *esvm*) is defined as the $\text{sign}(\sum_{i=1}^m \lambda_i K(\mathbf{x}_i, \mathbf{z}) + b)$, we had to recover \mathbf{w} as $\sum_{i=1}^m \lambda_i y_i \mathbf{x}_i$ and then select p genes associated with the top p weights (i.e., $\mathbf{w} = [|w_1| \dots |w_p|]$), which correspond to the top p important genes. For *lasso*, the

model is expressed as $\beta_0 + \beta x$, and we selected the p genes associated with top p coefficients (i.e., $\beta = [|\beta_1| \dots |\beta_p|]$) (excluding genes associated with zero coefficients). For limma, sam, and t -test baseline methods, genes were selected based on significantly adjusted p -values < 0.01 .

To perform enrichment analysis and evaluate the results from a biological perspective, we uploaded genes obtained from each method to Enrichr (<https://maayanlab.cloud/Enrichr/>, accessed on 3 October 2023) and Metascape (<https://metascape.org/gp/index.html>, accessed on 21 September 2023) [14,16]. When retrieved terms are related to breast cancer, a method that has terms associated with more genes is considered the superior method. Furthermore, we assessed the performance from the classification perspective against lasso as a baseline, reporting area under the ROC curve (AUC) as a performance measure, followed by conducting a statistical significance test and reporting the running time. In this study, we utilized R to run the experiments [34]. Specifically, we used the CVXR package in R to aid in solving the formulated optimization problem [33]. We employed the siggenes package to run sam [35], and we employed the limma package in R to run the limma using the two functions lmFit and eBayes [17]. In terms of the t -test, we employed the t -test function within the stats package [34]. To run lasso, we used the glmnet package [20], in which we set $\lambda = 0.05$ and also utilized cross-validation on the training set to find the optimal λ when using lasso for classification. For sam, limma, and t -test, to compute adjusted p -values, we employed the p.adjust function with the “BH”, setting $p < 0.01$ as in [29,36]. We used DeepLift, DeepSHAP, and LRP functions in the insight package in R to run DeepLIFT, DeepSHAP, and LRP, respectively [37].

For Dataset1, the gene expression experiment is to analyze and evaluate neoadjuvant docetaxel and carboplatin plus trastuzumab and/or lapatinib for patients with HER2+ breast cancer, obtained from the School of Medicine at the University of California, Los Angeles, USA. We retrieved gene expression profiles of patients responding to the treatment (labeled as pathological complete response (PCR)) and those not completely responding to the treatment (labeled as a residual disease (RD)) from the gene expression omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>, accessed on 17 March 2023) with GEO accession number GSE130787. We utilized the getGEO() function within the GEOquery package [38] to download and obtain the gene expression data. We utilized the fData() function from the Biobase package [39] to process and obtain gene names; the pData() function within the Biobase package was used to obtain drug responses (i.e., PCR and RD) associated with each sample; the exprs() function within the Biobase package to obtain expression values; and the missForest() function within the missForest package to impute missing values [40,41]. We selected 5267 protein-coding genes (PCGs) according to Biomart domain experts at the HUGO Gene Nomenclature Committee (see Section 2.1.1).

For Dataset2, the gene expression experiment is for predicting neoadjuvant chemotherapy response in early breast cancer, obtained from the Leibniz Research Centre for Working Environment and Human Factors, located in Dortmund, Germany. We retrieved gene expression profiles of breast cancer patients responding (labeled as PCR and pathological partial response (PPR)) and those not responding to treatment (labeled as pathological no change (PNC)) from the GEO database at NIH under GEO accession number GSE140494. As in Dataset1, we used the five functions (i.e., getGEO(), fData(), pData(), exprs(), and missForest()) to download, prepare, and impute Dataset2. Moreover, we selected 5313 PCGs according to BioMart domain experts at the HUGO Gene Nomenclature Committee (see Section 2.1.2).

For Dataset3, the gene expression experiment is for the I-SPY 2 neoadjuvant chemotherapy/targeted therapy trial for early-stage breast cancer patients with high risk, obtained from the University of California, San Francisco, CA, USA. Gene expression profiles pertaining to breast cancer patients responding to treatment (labeled as a complete response) against those not responding properly to the treatment (labeled as a failed complete response) were obtained from the GEO database under GEO Accession number GSE196093.

As in Dataset2, we used the five functions (i.e., `getGEO()`, `fData()`, `pData()`, `exprs()`, and `missForest()`) to download, prepare, and impute Dataset3.

3.2. Results

3.2.1. Dataset1

In Table 2, we report terms obtained from Enrichr based on input genes provided via each method. Terms in Table 2(a) show expressed genes within breast tissues according to each method. The more expressed genes, the better the computational method is. Table 2(a) demonstrates that `esvm` outperformed all compared methods, obtaining 28 expressed genes (`SLC22A3`, `FKBP10`, `GRP`, `STAB2`, `LPL`, `GLI3`, `ADCY5`, `OBP2A`, `SOSTDC1`, `APOD`, `HMGCS2`, `GABRE`, `CCL21`, `TMEM61`, `BBOX1`, `GFRA1`, `IGF1`, `BMP6`, `APLN`, `CAPN13`, `NR4A1`, `C1ORF116`, `ANGPTL7`, `KCNS3`, `SSPN`, `PGR`, `FGFR2`, and `LTF`) within the breast tissue. The second-best method is `limma` with 14 expressed genes (`TMEM86B`, `LRP1`, `MEOX1`, `PLCZ1`, `OBP2A`, `PPP1R1B`, `LRIG1`, `ECHDC3`, `MAFK`, `MZF1`, `A2M`, `RNF186`, `FMOD`, and `KANK3`) within the breast tissue. The worst-performing method is `lasso`, in which eight expressed genes (`DGAT1`, `RNASE7`, `FXYD1`, `VEGFB`, `ROR2`, `FGF1`, `HSPA12A`, and `PTPN14`) are expressed in the breast tissue. In Table 2(b), we show retrieved terms (i.e., cancer cell lines) and expressed genes within NCI-60 cancer cell lines. Our method, `esvm`, performed better than other methods, obtaining a total of eight expressed genes within the retrieved breast cancer cell lines. Specifically, three genes (`PNMT`, `COL13A1`, and `BCAS1`) were expressed within MD-MB231, three genes (`MYO5C`, `GFRA1`, and `FGFR2`) were expressed within MCF7, and two genes (`COL13A1` and `FAR2`) were expressed within HS578T. The second-best method is `limma` (tied with `esvm`), both having four expressed genes within two breast cancer cell lines. In terms of `limma`, one gene (`ECHDC3`) was expressed within MD-MB231, whereas three genes (`RBM8A`, `USP18`, and `PI4K2A`) were expressed within MCF7. For `sam`, one gene (`CLSPN`) was expressed within MD-MB231, and three genes (`TFAP2C`, `GFRA1`, and `CNNM3`) were expressed within MCF7. The worst-performing method is `lasso`, which has two expressed genes within breast cancer cell lines. These results demonstrate the superiority of our method (i.e., `esvm`) in identifying breast tissue as well as breast cancer cell lines used in cancer research. In Supplementary Table1_A and Table1_B, we include enrichment analysis results for ARCHS4 tissues and NCI-60 cancer cell lines, respectively.

In Figure 2a, we provide a visualization of intersected genes produced by all computational methods. It can be seen that `esvm` has 94 unique genes out of 100 when compared to all other methods. This implies that each method generates a different list of genes. For example, `limma`, `sam`, `t-test`, and `lasso` had unique genes of 96, 95, 94, and 38, respectively. It can also be seen that the number of common genes between any pair of methods is at most 2. This indicates that each method generated a different gene set, and the similarity among methods is minimal. In Supplementary DataSheet1_B, we list the genes according to the UpSet plot in Figure 2a.

As `esvm` demonstrated superiority over the other computational methods, we provided the 100 genes produced via `esvm` to Metascape for further enrichment analysis. Figure 2b displays the following 44 genes obtained from protein–protein interaction: `GRP`, `TRH`, `LTF`, `TTR`, `APOD`, `CSSC4`, `CAPN9`, `CAPN13`, `S100A6`, `S100A2`, `CARTPT`, `DUSP13B`, `FKBP10`, `GGH`, `ERP27`, `POF1B`, `SOSTDC1`, `CEACAM5`, `BMP6`, `CHRD2`, `GSTA1`, `QDPR`, `NDRG1`, `NR4A1`, `SOCs4`, `PPBP`, `CCL21`, `APLN`, `TRIM63`, `ADCY5`, `CALML4`, `MYO5C`, `RAB38`, `AQP3`, `TUBA3C`, `PTP4A3`, `FOS`, `MS4A2`, `MMP1`, `MMP3`, `MMP12`, `FGFR2`, `IGF1`, and `PGR`. The enrichment analysis showed that 10 genes (`BMP6`, `IGF1`, `MMP1`, `MMP3`, `MMP12`, `PPBP`, `S100A2`, `S100A6`, `CCL21`, `CHRD2`) were related to NABA MATRISOME ASSOCIATED, 7 genes (`BMP6`, `IGF1`, `PPBP`, `S100A2`, `S100A6`, `CCL21`, `CHRD2`) were related to NABA SECRETED FACTORS, and 6 genes (`BMP6`, `FGFR2`, `NR4A1`, `IGF1`, `MMP12`, `APLN`) were related to positive regulation of epithelial cell proliferation. It has been reported that the NABA MATRISOME ASSOCIATED and NABA SECRETED FACTORS pathways play a key role in breast cancer metastasis through their involvement with extracellular matrix

proteins [42,43]. Also, genes linked to the positive regulation of epithelial cell proliferation biological processes are related to the induction of metastasis and inhibition of breast cancer cell apoptosis through the promotion of epithelial cell proliferation via estrogen [44,45]. These results demonstrate the effectiveness of our computational method in unveiling important molecular mechanisms pertaining to breast cancer pathogenesis and metastasis.

Table 2. Enriched terms were obtained from (a) ARCHS4 Tissues and (b) NCI-60 Cancer Cell Lines via Enrichr according to the genes produced by each method when using Dataset1. The best result is shown in bold.

(a) ARCHS4 Tissues					
Method	Rank	Term	Overlap	p-Value	Adjusted p-Value
esvm	10	BREAST (BULK TISSUE)	28/2316	5.92×10^{-6}	5.81×10^{-5}
limma	18	BREAST (BULK TISSUE)	14/2316	2.65×10^{-1}	9.99×10^{-1}
sam	19	BREAST (BULK TISSUE)	13/2316	3.72×10^{-1}	9.99×10^{-1}
t-test	49	BREAST (BULK TISSUE)	9/2316	8.32×10^{-1}	9.99×10^{-1}
lasso	1	BREAST (BULK TISSUE)	8/2316	9.50×10^{-2}	9.87×10^{-1}
(b) NCI-60 Cancer Cell Lines					
Method	Rank	Term	Overlap	p-Value	Adjusted p-Value
esvm	12	MD-MB231	3/150	3.94×10^{-2}	2.50×10^{-1}
	33	MCF7	3/397	3.19×10^{-1}	7.21×10^{-1}
	24	HS578T	2/176	2.19×10^{-1}	6.63×10^{-1}
limma	45	MD-MB231	1/150	5.29×10^{-1}	7.74×10^{-1}
	29	MCF7	3/397	3.19×10^{-1}	6.74×10^{-1}
	-	HS578T	-	-	-
sam	46	MD-MB231	1/150	5.29×10^{-1}	8.85×10^{-1}
	29	MCF7	3/397	3.19×10^{-1}	7.83×10^{-1}
	-	HS578T	-	-	-
t-test	-	MD-MB231	-	-	-
	46	MCF7	2/397	5.93×10^{-1}	8.72×10^{-1}
	45	HS578T	1/176	5.87×10^{-1}	8.72×10^{-1}
lasso	26	MD-MB231	1/150	2.65×10^{-1}	5.11×10^{-1}
	-	MCF7	-	-	-
	29	HS578T	1/176	3.04×10^{-1}	5.24×10^{-1}

Figure 2c reports 12 transcription factors (TFs): NFKBIA, FOS, RELA, STAT3, JUN, BRCA1, SP1, USF1, ETS1, CREB1, NFKB1, and MYC. BRCA1 is known to play a key role in various biological processes in breast cancer [46–50]. TFs such as ETS1 and STAT3 have been reported as potential therapeutic targets in breast cancer [51]. Suppression of NFKBIA and CREB1 has been reported to be related to the inhibition of breast cancer progression [52,53]. These 12 TFs can (1) aid in understanding breast cancer molecular mechanisms and (2) act as potential therapeutic targets for breast cancer treatment. In Figure 2d, genes provided via esvm are related to various biological processes and pathways in breast cancer progression. The top-enriched term is the vitamin D receptor pathway, which ameliorates breast cancer by contributing to the growth regulation of breast cancer cells [54,55]. Other enriched terms, such as NABA MATRISOME ASSOCIATED, mammary gland development, and extracellular matrix organization, have been reported to play a key role in various biological processes related to breast cancer progression and treatment [56–60]. In Supplementary Enrichment_Dataset1, we list Metascape enrichment analysis results related to Dataset1.

In Table 3, we report terms (i.e., drugs) and expressed genes (i.e., drug targets) within IDG Drug Targets 2022. Tamoxifen and Fulvestrant are antiestrogen inhibitors (hormone therapy) that have been approved for breast cancer treatment (see Figure 3) [61–63]. Both drugs identified two drug targets, PGR and ATP1A2. Cisplatin is a chemotherapy used for breast cancer treatment [64,65]. ATP1A2 was a drug target reported in association with cisplatin. The obtained biological knowledge contributes to a better understanding of breast

cancer progression and treatment. In Supplementary Table1_C, we provide enrichment analysis results pertaining to IDG Drug Targets 2022.

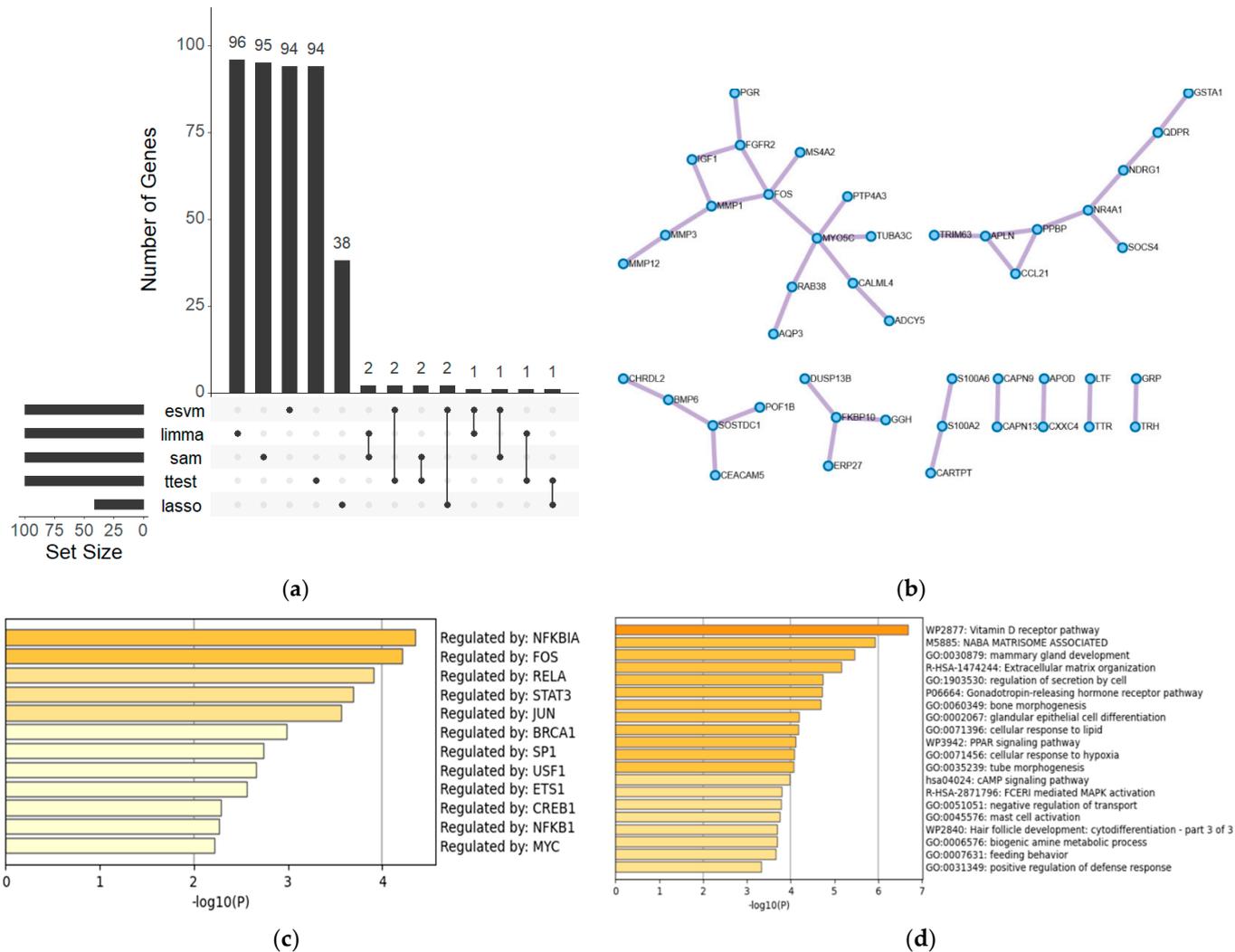


Figure 2. (a) UpSet plot of gene lists provided by the computational methods when using Dataset1. (b) Nine clusters in a protein–protein interaction based on genes of esvm coupled with Metascape. (c) Twelve transcription factors, according to Metascape, when coupled with genes from esvm. (d) Process and pathway enrichment analysis provided by Metascape according to the 100 genes of esvm.

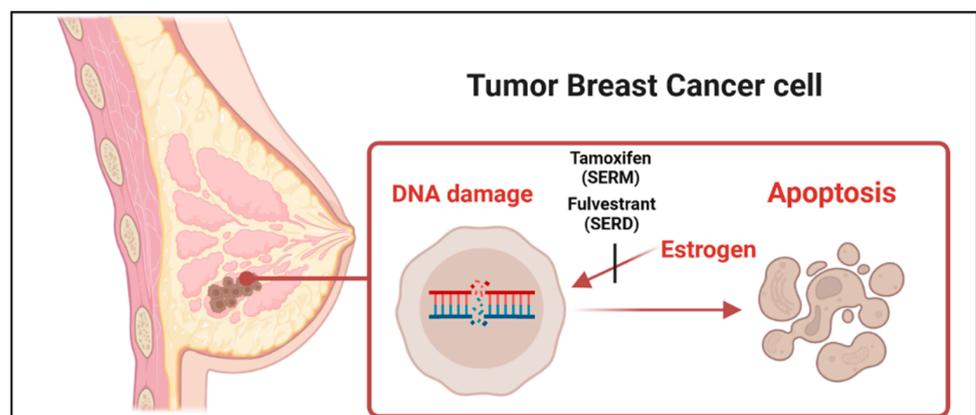


Figure 3. Tamoxifen and fulvestrant anticancer drugs inhibit estrogen, causing DNA damage and cell death. SERM is a selective estrogen receptor modulator. SERD is a selective estrogen degrader.

Table 3. Enriched terms from IDG Drug Targets 2022 via Enrichr were retrieved according to uploaded genes from esvm using Dataset1, showing genes (column: genes) associated with drugs (column: term). The rank column shows the order of terms when retrieved.

Rank	Term	Class	Genes
35	Tamoxifen	Antiestrogen	PGR and ATP1A2
37	Fulvestrant	Estrogen Receptor Antagonist	PGR
64	Cisplatin	Platinum Coordination Complex	ATP1A2

3.2.2. Dataset2

In Table 4(a), we report retrieved breast tissue terms associated with expressed genes, while Table 4(b) reports breast cancer cell line terms with expressed genes after uploading genes produced via all computational methods. It can be seen from Table 4(a) that our method (esvm) performed better than its competing baseline methods. Particularly, esvm had breast tissue terms with 20 expressed genes (COL15A1, PFKFB3, VCAM1, TAT, TNC, PLAT, LAMC2, ACTG2, NR4A1, CYP2A6, KRT19, KRT18, COL5A1, DUOXA1, SCNN1A, FOSB, KCNN4, CD300LG, LTF, and MPZL2) out of 2316. The second-best method was sam with 13 expressed genes (DSP, IGFBP5, ODF3, TACSTD2, KLK8, SLC5A6, EFEMP1, FABP4, PER3, SLPI, CRTAC1, STAB1, and DACT2) out of 2316. Limma and *t*-test were tied for the third-best performing methods, with both having 11 expressed genes within the breast tissue. Limma had the expressed genes, namely CAMSAP3, FABP4, JUP, TRAF4, SLFN11, TRIM29, WNT9A, A2M, PCDH1, LOXL1, and SLC9A1, while the *t*-test had the following genes: SLC22A23, MUCL1, GIPC3, PACS2, FZD7, PADI2, CLEC4F, CFB, KIAA0040, SYT7, and EGFR. The worst-performing method was lasso with 7 expressed genes (RECQL4, NR4A1, SLC44A4, GJD3, CLDN7, SHB, and LZTR1) out of 2316. Table 4(b) demonstrates enriched breast cancer cell line terms from NCI-60 cancer cell lines obtained via Enrichr. The best-performing method is esvm, with a total of seven expressed genes within MCF7 and HS578T breast cancer cell lines, distributed as follows: Six expressed genes (DCTN5, KRT19, DAAM1, KYNU, TRIM37, and MPZL2) out of 397 within the MCF7 breast cancer cell line, while one expressed gene (ACTG2) within the HS578T breast cancer cell line. In terms of the MD-MB231 breast cancer cell line, esvm had no retrieved results. Therefore, results were designated as “-.” The second best-performing method was sam, which resulted in five expressed genes within MD-MB231 and MCF7 breast cancer cell lines. Two expressed genes (RHEB and GARNL3) out of 150 were expressed within the MD-MB231 breast cancer cell line, whereas three genes (PIAS3, DCTN5, and CTCF) out of 397 were expressed within the MCF7 breast cancer cell line. For the HS578T breast cancer cell line, no retrieved results were reported for sam (see “-”). The worst-performing method was lasso with three expressed genes within MD-MB231, MCF7, and HS578T breast cancer cell lines. These results demonstrate the good performance of esvm in identifying breast tissues and cancer cell lines. In Supplementary DataSheet2_A, we include genes obtained from computational methods provided to Enrichr to derive enrichment analysis results. Additionally, enrichment analysis results related to Table 4 are included in Supplementary Table2_A and Table2_B. Figure 4a displays the UpSet plot in terms of intersection lists of genes from all computational methods. A total of 91 unique genes were attributed to esvm. sam, limma, *t*-test, and lasso had 94, 93, 92, and 33 unique genes, respectively. These results indicate that each method incorporated different computational steps, resulting in different lists of genes. The number of intersected genes between each pair of computational methods is upper-bounded by 4. In Supplementary DataSheet2_B, we include gene lists of computational methods related to Figure 4a.

Table 4. Enriched terms were obtained from (a) ARCHS4 Tissues and (b) NCI-60 Cancer Cell Lines via Enrichr according to the genes produced by each method when using Dataset2. The best result is shown in bold.

(a) ARCHS4 Tissues					
Method	Rank	Term	Overlap	<i>p</i> -Value	Adjusted <i>p</i> -Value
esvm	21	BREAST (BULK TISSUE)	20/2316	1.00×10^{-2}	4.91×10^{-2}
limma	26	BREAST (BULK TISSUE)	11/2316	6.18×10^{-1}	9.99×10^{-1}
sam	11	BREAST (BULK TISSUE)	13/2316	3.72×10^{-1}	9.99×10^{-1}
<i>t</i> -test	17	BREAST (BULK TISSUE)	11/2316	6.18×10^{-1}	9.99×10^{-1}
lasso	6	BREAST (BULK TISSUE)	7/2316	1.03×10^{-1}	8.76×10^{-1}
(b) NCI-60 Cancer Cell Lines					
Method	Rank	Term	Overlap	<i>p</i> -Value	Adjusted <i>p</i> -Value
esvm	-	MD-MB231	-	-	-
	4	MCF7	6/397	1.47×10^{-2}	2.75×10^{-1}
	58	HS578T	1/176	5.87×10^{-1}	7.72×10^{-1}
limma	-	MD-MB231	-	-	-
	43	MCF7	2/397	5.93×10^{-1}	8.36×10^{-1}
sam	-	HS578T	-	-	-
	32	MD-MB231	2/150	1.72×10^{-1}	4.48×10^{-1}
	41	MCF7	3/397	3.19×10^{-1}	6.46×10^{-1}
<i>t</i> -test	-	HS578T	-	-	-
	-	MD-MB231	-	-	-
	32	MCF7	3/397	3.19×10^{-1}	6.80×10^{-1}
lasso	49	HS578T	1/176	5.87×10^{-1}	8.43×10^{-1}
	14	MD-MB231	1/150	2.31×10^{-1}	6.54×10^{-1}
	33	MCF7	1/397	5.04×10^{-1}	6.54×10^{-1}
	15	HS578T	1/176	2.66×10^{-1}	6.54×10^{-1}

As esvm performed better than baseline computational methods, we provided genes obtained from esvm to Metascape to unveil biological insights within breast cancer drug responses. Figure 4b reports the following 12 genes obtained from the PPI network: ACTG2, PRKAR2B, ALDH2, PRKACB, KRT19, KRT18, WASF3, DLG5, TRIM37, GSTM3, GSTA1, and CYP2A6. Four genes (KRT18, KRT19, PLCB4, and PRKACB) were related to the estrogen signaling pathway, which is reported to play a key role in breast cancer progression and treatment [66–68]. Three genes (CYP2A6, GSTA1, and GSTM3) were linked to chemical carcinogenesis—the DNA adduct pathway—involved in cancer development [69,70]. In Figure 4c, we report the following 18 transcription factors: SP1, RELA, NFKB1, HIF1A, STAT3, ZEB1, FOXO3, NFE2L2, JUN, NRC1, SP3, EP300, E2F1, CEBPB, USF1, HDAC1, ETS1, and STAT1. STAT3 is involved in breast cancer progression [71]. E2F1 and EP300 have been reported to be involved in breast cancer development and metastasis [72–74]. These results demonstrate the importance of these TFs, and such mutations or alterations can affect gene regulation and thereby contribute to breast cancer development. In Figure 4d, the top-enriched term was the nuclear receptor meta-pathway. Ten genes (CYP2A6, GSTA1, GSTM3, HMOX1, ME1, S100P, SCNN1A, ABCC4, PLK2, and B3GNT5) were linked to the nuclear receptor meta-pathway, which has been related to breast cancer cell growth via nuclear receptors such as estrogen receptors [75]. Moreover, the 13 genes (AGT, BCL6, CDKN2C, FGFR3, GJA1, TNC, MT1X, S100A8, S100A9, SYT1, SOCS2, SEMA3C, and CHPT1) were related to the regulation of the growth process, which is linked to breast cancer cell growth. Eleven genes (AGT, BIRC5, CCND2, FGFR3, GSTA1, GSTM3, HMOX1, CXCL8, LAMC2, PLCB4, and PRKACB) were related to pathways in cancer, which are linked to breast cancer development and metastasis [76]. Six genes (BCL6, CCND2, CDKN2C, CXCL8, PLAT, and PROM1) were related to transcriptional misregulation in cancer, which is linked to mutations and altered gene expression in breast cancer [77]. Enrichment analysis results obtained from Metascape are provided in Supplementary Enrichment_Dataset2.

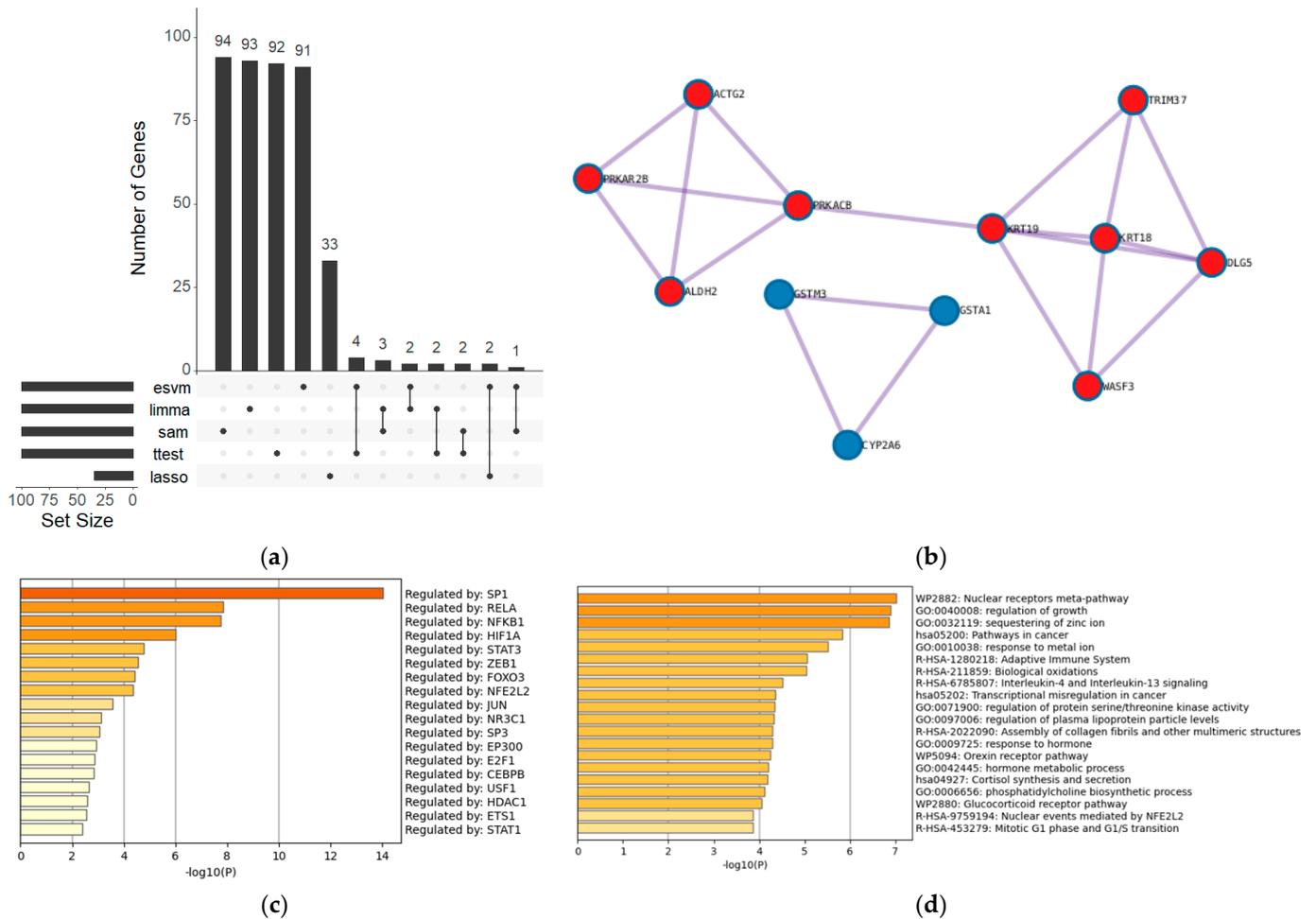


Figure 4. (a) UpSet plot of gene lists provided by the computational methods when using Dataset2. (b) Three clusters in a protein–protein interaction based on genes of esvm coupled with Metascape. (c) Eighteen transcription factors, according to Metascape, when coupled with genes from esvm. (d) Process and pathway enrichment analysis provided by Metascape according to the 100 genes of esvm.

Table 5 shows drugs and drug targets within the IDG Drug Targets 2022. Hydroxycarbamide and gemcitabine are ribonucleotide reductase enzyme (RNR) inhibitors that have been used for breast cancer treatment (see Figure 5) [78,79]. The two drugs are associated with RRM2 as a drug target [80]. Daunorubicin inhibits DNA replication and cyclophosphamide, causing damage to the DNA of cancer cells and thereby causing cancer cells to die. TOP2A was a drug target for daunorubicin, whereas RRM2 was a drug target for cyclophosphamide. In Supplementary Table2_C, we report enrichment analysis results for IDG Drug Targets 2022.

Table 5. Enriched terms from IDG Drug Targets 2022 via Enrichr were retrieved according to uploaded genes from esvm using Dataset2, showing genes (column: genes) associated with drugs (column: term). The rank column shows the order of terms when retrieved.

Rank	Term	Class	Genes
5	hydroxycarbamide	Antimetabolite	RRM2
13	daunorubicin	Anthracycline	TOP2A
19	cyclophosphamide	Alkylating Agent	RRM2
20	gemcitabine	Antimetabolite	RRM2

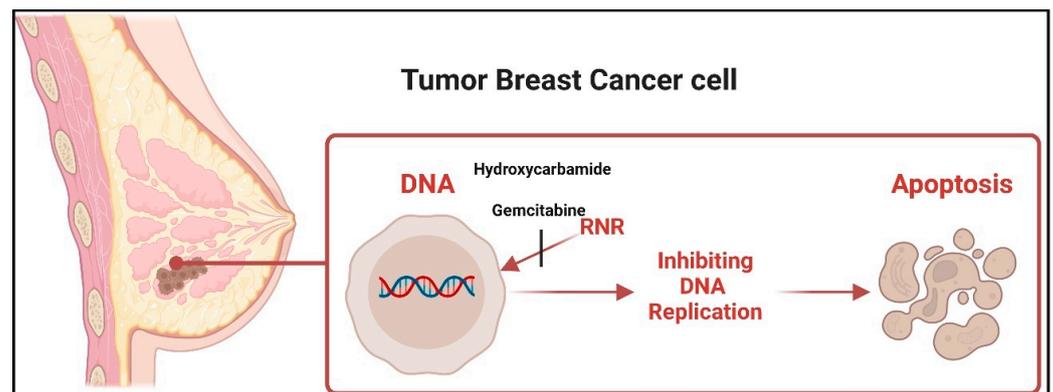


Figure 5. Hydroxycarbamide and gemcitabine inhibit the ribonucleotide reductase (RNR) enzyme, which inhibits DNA replication during the cell cycle and induces apoptosis.

3.2.3. Dataset3

After uploading produced genes via each method to Enrichr, we show retrieved breast tissue terms within ARCHS4 tissues (See Table 6(a)) and breast cancer cell lines terms within NCI-60 cancer cell lines (see Table 6(b)). From Table 6(a), we can see that esvm generated the best results. Particularly, esvm had six expressed genes (CSF1, CCND1, ERBB3, IRS1, ERBB2, and ESR1) out of 2316 within breast tissue, followed by sam and *t*-test, both having four common expressed genes (RET, CCND1, IRS1, and ERBB2) out of 2316 within breast tissue. Lasso had two (ESR1 and EGFR) expressed genes out of 2316 within breast tissue, while limma was the worst-performing method, having one expressed gene (ABL1) out of 2316 within the breast tissue. In Table 6(b), esvm is also the best-performing method by having 3 expressed genes out of 397 within the MCF7 breast cancer cell line. No results were associated with MD-MB231 and HS578T breast cancer cell lines. Therefore, we indicated results by “-”. Both the sam and *t*-test were tied by having two expressed genes out of 397 within the MCF7 breast cancer cell line. As for esvm, no reported results were found for the other two breast cancer cell lines (i.e., MD-MB231 and HS578T). The worst-performing method was limma, where no reported results were found for the three breast cancer cell lines. These results demonstrate the superiority of esvm when identifying breast cancer tissue and cell lines. We include enrichment analysis results regarding Table 6 in Supplementary Table3_A and Table3_B.

In Figure 6a, we show the UpSet plot showing the intersection of produced genes among all methods when Dataset3 is used. From the leftmost, it appears that esvm differs from all other methods by having 20 unique genes. Limma and lasso have 17 and 6 unique genes, respectively. esvm, sam, and *t*-test share 13 genes. Limma and esvm share 12 genes. sam and the *t*-test share eight genes. Lasso and esvm share three genes. sam, *t*-test, and lasso share two genes. It can also be seen that the number of common genes between the remaining intersections of methods is 1. These results demonstrate that our method is different from the remaining methods, attributed to the different computational steps involved in the computation of esvm. In Supplementary DataSheet3_B, we report genes related to the UpSet plot. Figure 6b reports the following 19 genes obtained from the PPI network: AR, BIRC5, CCND1, RB1, STAT1, STAT3, ESR1, IRS1, PTEN, ERBB2, ERBB3, ALK, AKT1, MET, JAK2, IGF1R, EGFR, TP53, and MTOR. Thirteen genes (AKT1, ARAF, CCND1, EGFR, ERBB2, ERBB3, MTOR, IGF1R, JAK2, MET, PTEN, RAF1, and STAT3) were linked to EGFR tyrosine kinase inhibitor resistance, involved in the resistance mechanism of EGFR inhibitors, and thereby breast cancer progression [81]. Twelve genes (AKT1, CCND1, ERBB2, ESR1, MTOR, IRS1, JAK2, NOS3, PTEN, RAF1, STAT1, and STAT3) were involved in the leptin signaling pathway, which is related to breast cancer malignancy [82,83]. Five genes (BIRC5, CCND1, ESR1, RB1, and CHEK2) were involved in the PID FOXM1 PATHWAY, which has been reported as a crucial oncogenic transcription factor that promotes breast cancer progression and growth [84,85].

Table 6. Enriched terms were obtained from (a) ARCHS4 Tissues and (b) NCI-60 Cancer Cell Lines via Enrichr according to the genes produced by each method when using Dataset3. The best result is shown in bold.

(a) ARCHS4 Tissues					
Method	Rank	Term	Overlap	p-Value	Adjusted p-Value
esvm	3	BREAST (BULK TISSUE)	6/2316	5.28×10^{-1}	9.97×10^{-1}
limma	40	BREAST (BULK TISSUE)	1/2316	9.78×10^{-1}	9.78×10^{-1}
sam	4	BREAST (BULK TISSUE)	4/2316	2.99×10^{-1}	9.47×10^{-1}
t-test	4	BREAST (BULK TISSUE)	4/2316	2.99×10^{-1}	9.47×10^{-1}
lasso	9	BREAST (BULK TISSUE)	2/2316	4.12×10^{-1}	7.71×10^{-1}
(b) NCI-60 Cancer Cell Lines					
Method	Rank	Term	Overlap	p-Value	Adjusted p-Value
esvm	-	MD-MB231	-	-	-
	3	MCF7	3/397	7.68×10^{-2}	5.88×10^{-1}
	-	HS578T	-	-	-
limma	-	MD-MB231	-	-	-
	-	MCF7	-	-	-
	-	HS578T	-	-	-
sam	-	MD-MB231	-	-	-
	3	MCF7	2/397	8.14×10^{-2}	4.41×10^{-1}
	-	HS578T	-	-	-
t-test	-	MD-MB231	-	-	-
	3	MCF7	2/397	8.14×10^{-2}	4.41×10^{-1}
	-	HS578T	-	-	-
lasso	-	MD-MB231	-	-	-
	18	MCF7	1/397	2.13×10^{-1}	3.08×10^{-1}
	-	HS578T	-	-	-

In Figure 6c, we report the following 20 transcription factors (TFs): TP53, BRCA1, HDAC1, RELA, STAT3, E2F1, SP1, NFKB1, YBX1, ESR1, NKX3-1, AR, VHL, PAX5, CTNBN1, PPARG, PGR, JUN, KDM4B, and DNMT1. Interestingly, the TP53 mutation has been reported to be the most frequently occurring in breast cancer [86,87]. Patients with BRCA1 mutations are at higher risk of developing breast cancer and are thereby considered an important biomarker [88,89]. HDAC1 has been reported to be related to breast cancer cell proliferation [90]. These results demonstrate the importance of these TFs and can aim to develop therapeutic strategies for breast cancer treatment. Figure 6d reports the top-enriched terms regarding processes and pathways. Twenty-one genes (AKT1, ALK, BIRC5, AR, ARAF, CCND1, CASP7, EGFR, ERBB2, ESR1, MTOR, IGF1R, JAK2, MET, PTEN, RAF1, RB1, STAT1, STAT3, TP53, and FADD) were linked to pathways in cancer (hsa05200), coinciding with recently reported results as one of the top-enriched pathways in breast cancer [91].

Thirteen genes (AKT1, ARAF, CCND1, EGFR, ERBB2, ERBB3, MTOR, IGF1R, JAK2, MET, PTEN, RAF1, and STAT3) were related to EGFR tyrosine kinase inhibitor resistance (WP4806), which was reported as a significant pathway associated with breast cancer [77,92]. Twelve genes (AKT1, CCND1, ERBB2, ESR1, MTOR, IRS1, JAK2, NOS3, PTEN, RAF1, STAT1, and STAT3) were linked to the leptin signaling pathway (WP2034) and reported to have a key role in breast cancer tumorigenesis [93]. Nine genes (AKT1, EGFR, ERBB2, ERBB3, ESR1, MTOR, IRS1, MET, and PTEN) were linked to PI3K/AKT signaling in cancer (R-HSA-2219528), whose inactivation suppressed the proliferation of breast cancer cells, thereby inducing apoptosis [94]. Seven genes (AKT1, BIRC5, ATM, CASP7, RAF1, TP53, and FADD) were related to apoptosis (hsa04210), which plays a key role in controlling the excessive proliferation of breast cancer cells [95]. In Supplementary Enrichment_Dataset3, we include enrichment analysis results obtained from Metascape pertaining to Dataset3.

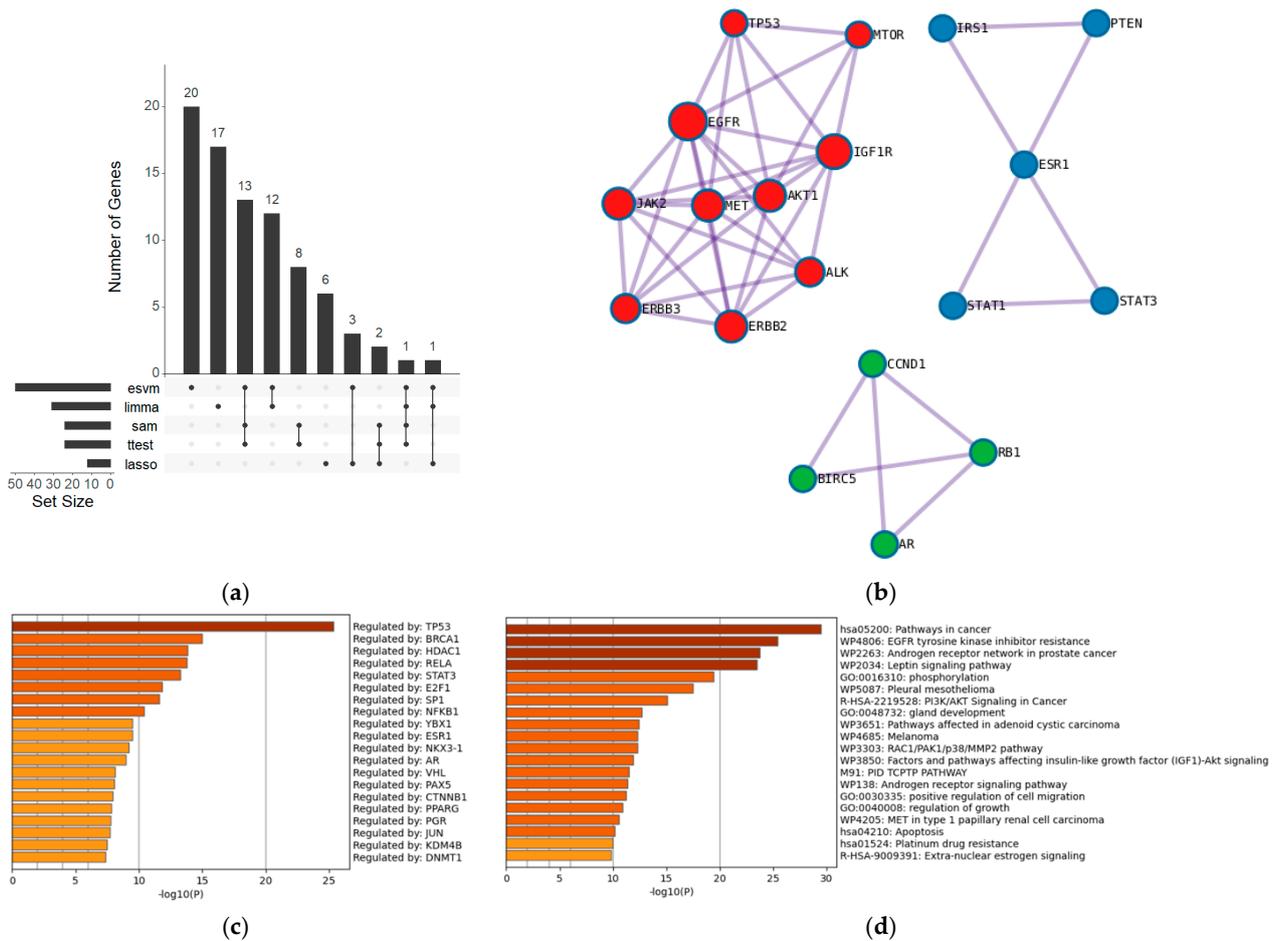


Figure 6. (a) UpSet plot of gene lists provided by the computational methods when using Dataset3. (b) Three clusters in a protein–protein interaction based on genes of esvm coupled with Metascape. (c) Twenty transcription factors, according to Metascape, when coupled with genes from esvm. (d) Process and pathway enrichment analysis provided by Metascape according to the genes of esvm.

Table 7 reports drug terms and drug targets within the IDG Drug Targets 2022. Ceritinib aims to inhibit the anaplastic lymphoma kinase (ALK) enzyme, thereby blocking the ability of tumors to grow and promoting apoptosis (see Figure 7) [96]. Erlotinib inhibits the effect of the tyrosine kinase enzyme on the epidermal growth factor receptor (EGFR), thereby preventing the proliferation of cancer cells and inducing apoptosis. In the Supplementary Table3_C, we report enrichment analysis results regarding drug terms within IDG Drug Targets 2022.

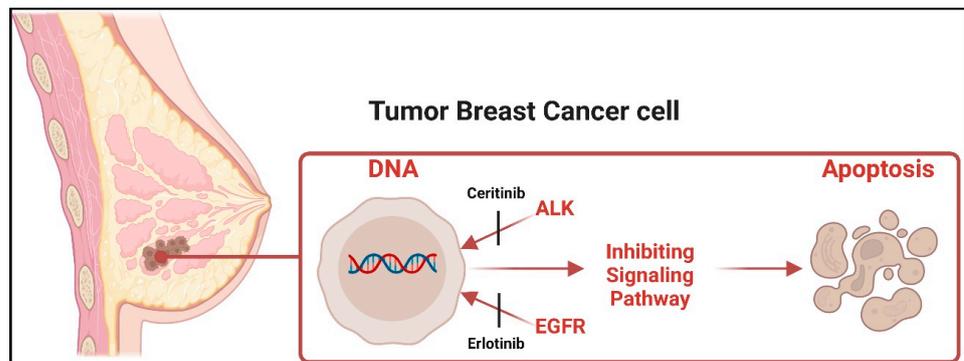


Figure 7. Ceritinib and erlotinib are anaplastic lymphoma kinase (ALK) and epidermal growth factor receptor (EGFR) inhibitors, respectively.

Table 7. Enriched terms from IDG Drug Targets 2022 via Enrichr were retrieved according to uploaded genes using Dataset3, showing genes (column: genes) associated with drugs (column: term). The rank column shows the order of terms when retrieved.

Rank	Term	Class	Genes
1	Ceritinib	Tyrosine Kinase Inhibitor	ALK; JAK2; MET; IGF1R
2	Erlotinib	Tyrosine Kinase Inhibitor	ALK; ERBB3; ERBB2; JAK2; MET
3	Entrectinib	Kinase Inhibitor	ALK; JAK2; IGF1R

3.3. Models Introspection

3.3.1. Dataset1

In Figure 8, we aim to obtain computational insights pertaining to learning-based models studied and applied in the Results Section. Figure 8a demonstrates that our method esvm leads to non-zero weights, while lasso in Figure 8b leads to a sparser representation and thereby many zero coefficients, attributed to the L1 penalty as in [28]. In Figure 8c, it can be seen that the lasso is $\sim 6.14 \times$ faster than the esvm. Figure 8d,e for esvm and lasso, respectively, demonstrate that prediction differences between breast cancer patients achieving pathological complete response against those having residual disease were statistically significant (p -value of all models $< 2.2 \times 10^{-16}$, obtained from a t -test).

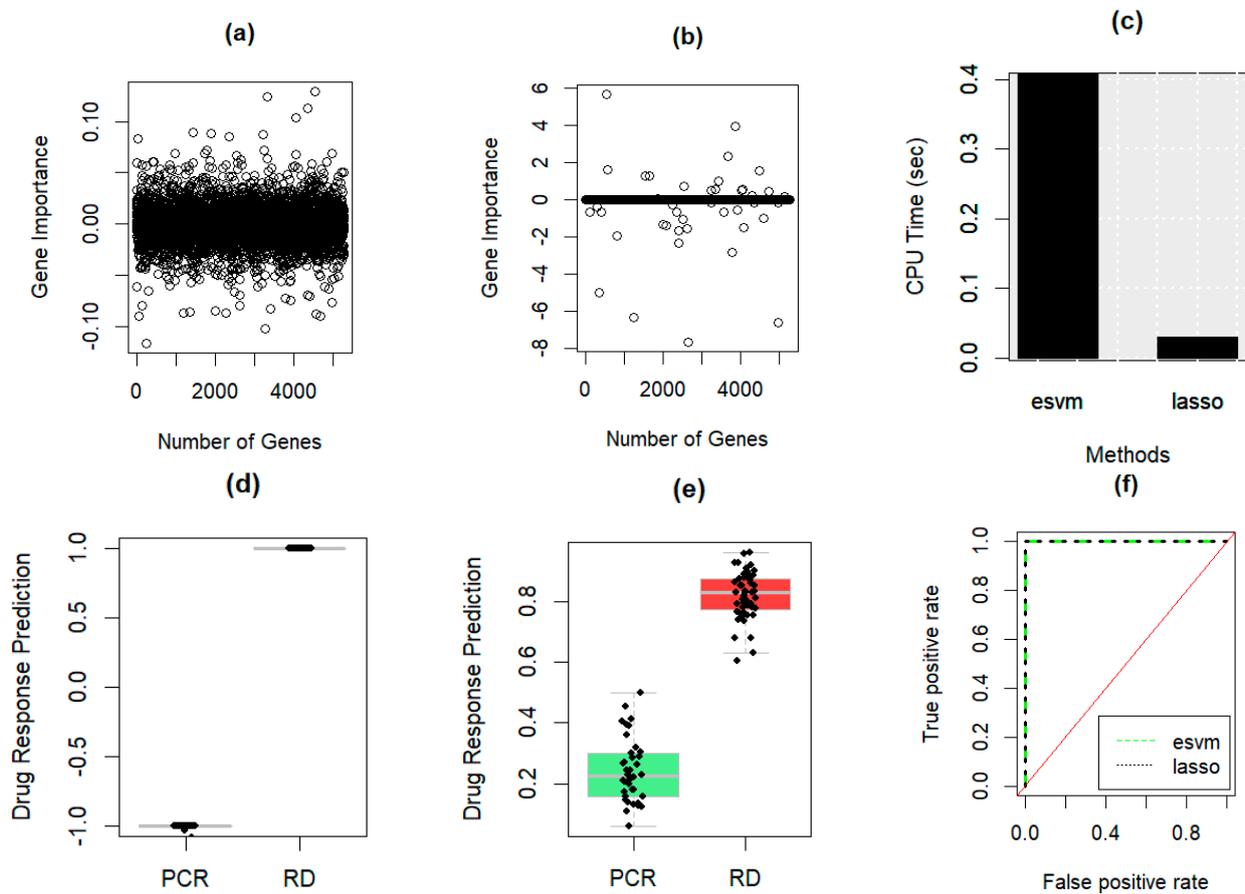


Figure 8. Classification models are compared for predicting the drug responses of TCH, TCHTy, and TCTy in breast cancer (BC) patients when Dataset1 is used. Gene importance when esvm (a) and lasso (b) are applied. Computational running time (c) for esvm and lasso. Boxplot and strip chart of drug sensitivity prediction for BC patients using esvm (d) and lasso (e). ROC curve (f) demonstrates the prediction performance. TCH is docetaxel, carboplatin, and trastuzumab. TCHTy is docetaxel, carboplatin, trastuzumab, and lapatinib. TCTy is docetaxel, carboplatin, and lapatinib. PCR is a pathological complete response. RD is a residual disease.

These results show that both induced models using Dataset1 are expected to be general predictors for drugs with PCR and RD responses. For the ROC curves of esvm and lasso in Figure 8f, both models achieved an area under the ROC curve (AUC) of 1.00.

3.3.2. Dataset2

As shown in Figure 9, different aspects related to models are reported as follows: Figure 9a shows that esvm leads to a non-sparse representation in which the weight vector w consists of non-zero weights. On the other hand, Figure 9b demonstrates that lasso leads to a sparse representation in which many of the coefficients β are zeros. Figure 9c displays that lasso is $44 \times$ faster than esvm.

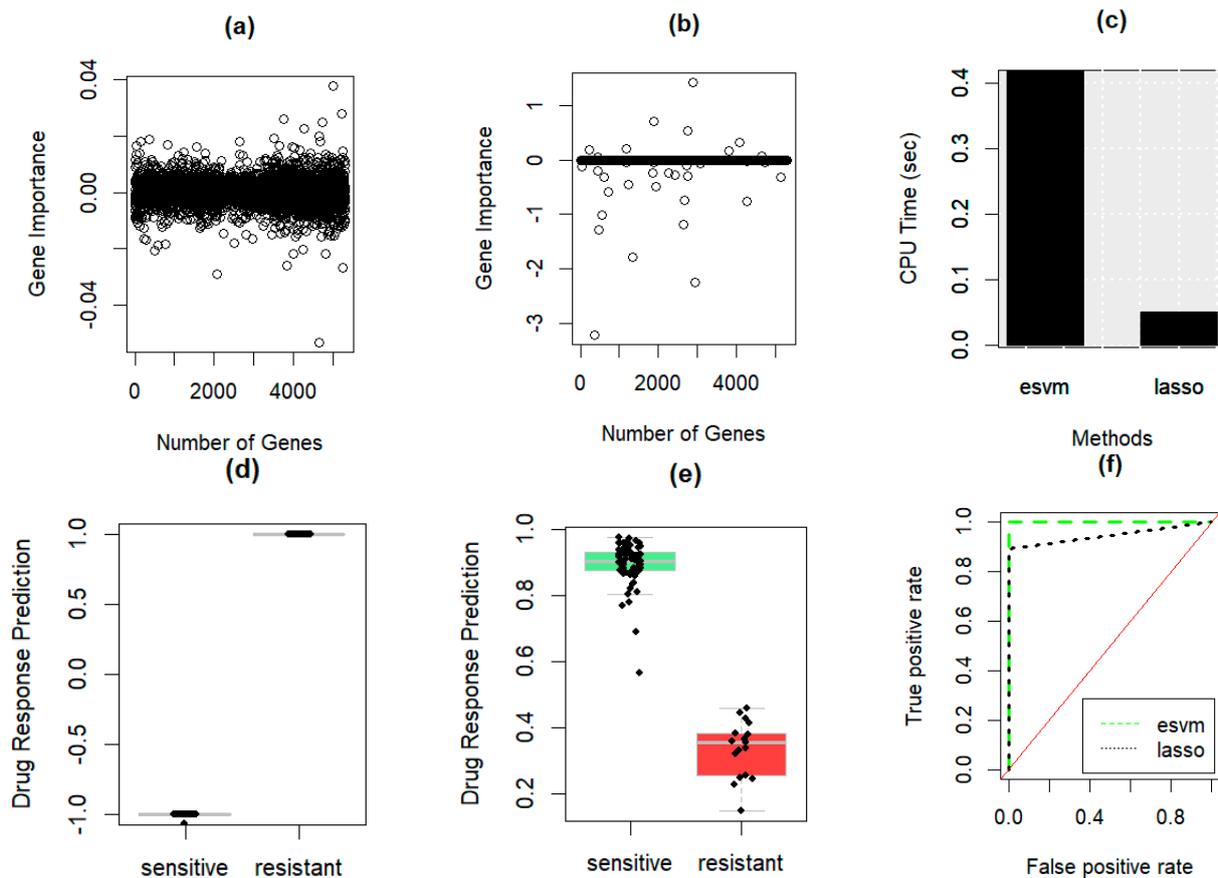


Figure 9. Classification models are compared for predicting TFEC drug responses in breast cancer (BC) patients when Dataset2 is used. Gene importance when esvm (a) and lasso (b) are applied. Computational running time (c). Boxplot and strip chart of drug sensitivity prediction for BC patients sensitive against those resistant to the drug treatment for esvm (d) and lasso (e). ROC curve (f) demonstrates the prediction performance. TFEC is docetaxel, 5-fluorouracil, epirubicin, and cyclophosphamide.

Figure 9d for esvm demonstrates that prediction differences between sensitive breast cancer patients and those with resistant drug responses were statistically significant (p -value $< 2.2 \times 10^{-16}$, obtained from a t -test). For lasso in Figure 9e, prediction differences between the two groups (i.e., BC patients sensitive to a drug against those having a resistant response) were not statistically significant (p -value = 1, obtained from a t -test). These results show that esvm is expected to be a general predictor of drug sensitivity and resistance. Lasso tends to be a specific predictor. Figure 9f displays the ROC curves for esvm and lasso, where the former has an AUC of 1.00 and the latter has an AUC of 0.947.

3.3.3. Dataset3

Figure 10 reports various computational aspects of the studied models from a classification perspective. It can be easily noticed that esvm does not shrink weights w to zero (Figure 10a), while lasso has a sparser representation attributed to the L1 regularization shrinking coefficients β to zero (see Figure 10b). In terms of efficiency, Figure 10c shows that lasso is $1974 \times$ faster than esvm. In terms of generalization, Figure 10d,e demonstrate that the prediction differences in esvm and lasso between breast cancer patients achieving complete response (CR) and those achieving failed complete response (FCR) were statistically significant (p -value $< 2.2 \times 10^{-16}$, obtained from a t -test). However, esvm had a better AUC of 0.722, while lasso had a lower AUC of 0.555 (see Figure 10f).

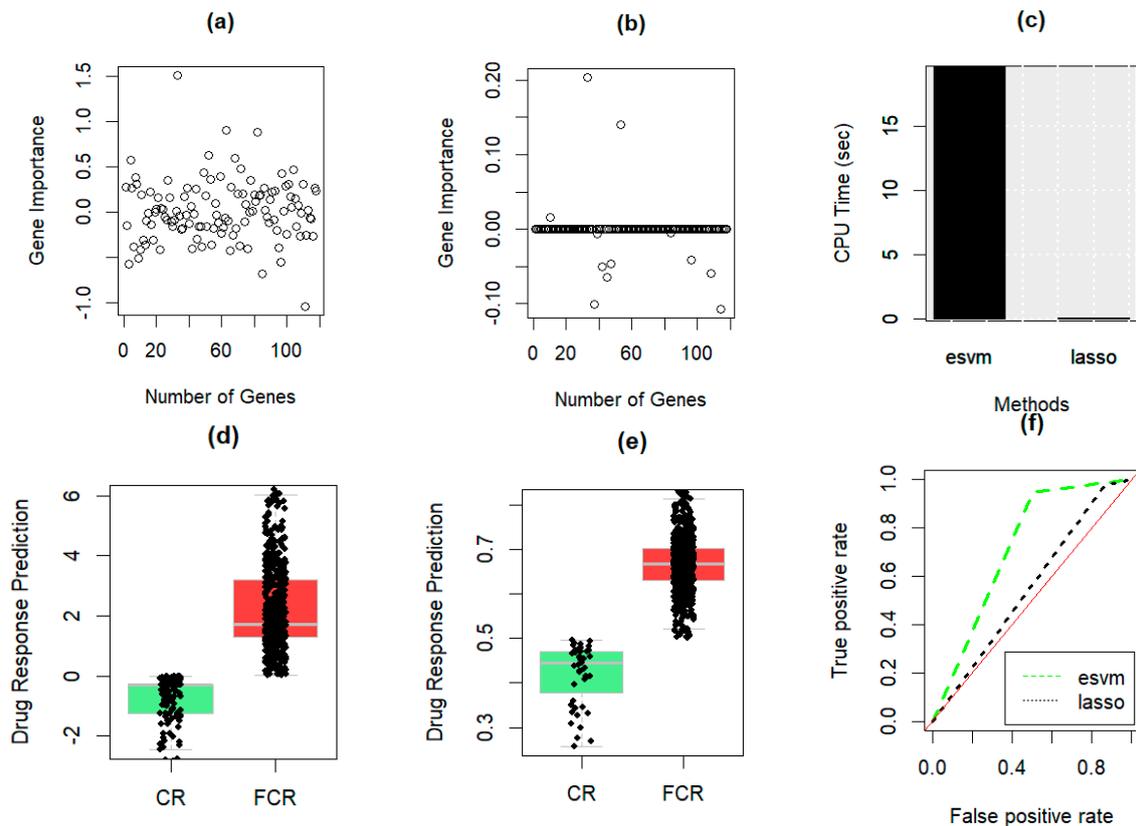


Figure 10. Classification models are compared for predicting multiple drug combination responses in breast cancer (BC) patients when Dataset3 is used. Gene importance when esvm (a) and lasso (b) are applied. Computational running time (c). Boxplot and strip chart of drug sensitivity prediction for BC patients sensitive against those resistant to the drug treatment for esvm (d) and lasso (e). ROC curve (f) demonstrates the prediction performance.

These performance results indicate that the prediction performance difference between esvm and lasso is 16.7% when AUC is considered.

3.3.4. Scalability

In Figure 11a–j, we report the computational running time for increased dimensionality, starting from 50,000 dimensions to 500,000 dimensions and fixing the number of rows to 100. The generation of x was conducted according to the uniform distribution $U(0, 1)$, and y was generated in which the class distribution was balanced. When the number of dimensions is 50,000 (see Figure 11a), the lasso was 27 and $121.5 \times$ faster than esvm and SVM, respectively. Our method, esvm, was 4.5 times faster than the baseline SVM. The average running times for esvm, SVM, and lasso spanning over all results in Figure 11a–j are 3.54 s, 23.88, and 0.275, respectively. That means our method, esvm, on average, is $6.74 \times$ faster than SVM,

while lasso was 12.87 and $86.83 \times$ faster than esvm and SVM, respectively. These results demonstrate the computational efficiency of our method over the SVM implementation using the e1071 package in R. In Supplementary Running_Time, we include all running time results related to Figure 11a–j.

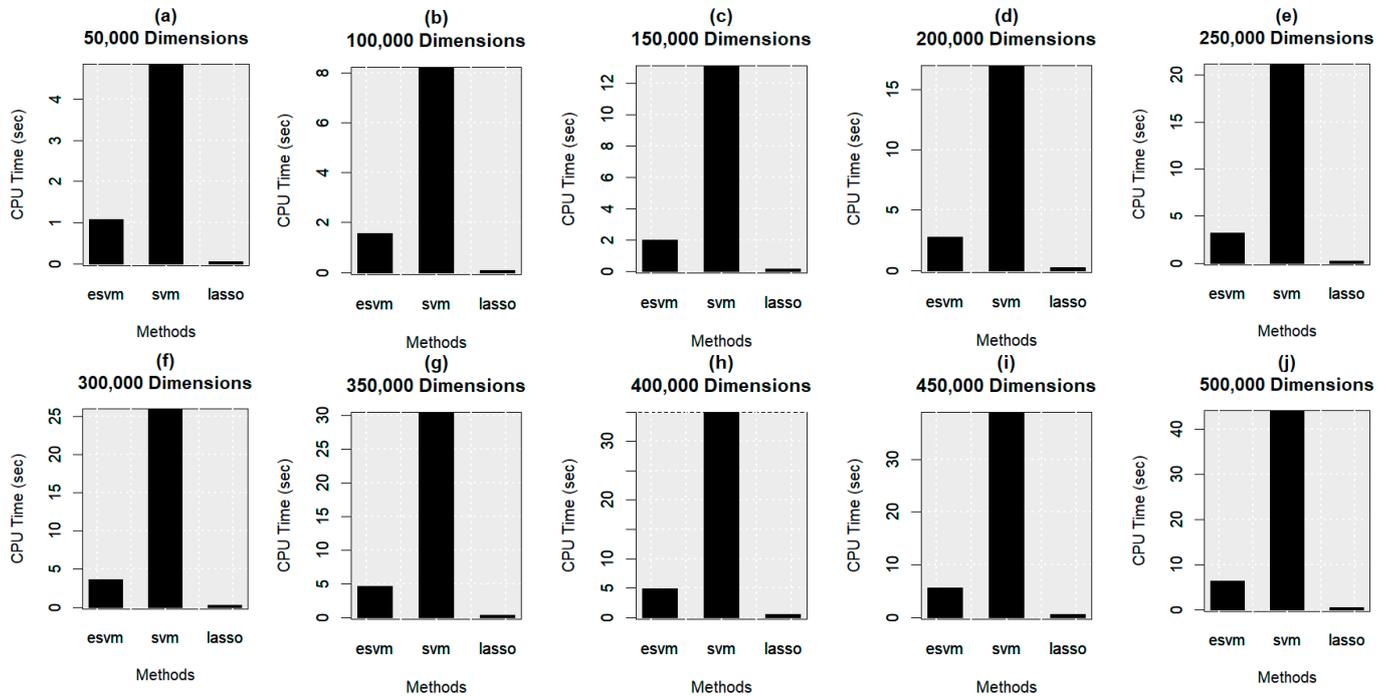


Figure 11. Computational running time for our model esvm against baseline methods (SVM and lasso) using simulated data of increased dimensionality.

4. Discussion

Identifying critical genes, drugs, drug targets, and transcription factors plays a key role in unveiling the underlying drug response mechanism of breast cancer. Therefore, we introduced an AI-based computational framework that functions as follows: First, because gene expression datasets consist of many genes compared to the number of samples, the optimization problem formulation of SVM in terms of finding weight vector \mathbf{w} and bias term b is impractical (see Equation (1)). Therefore, we employ the dual form of the SVM optimization problem formulated in terms of finding lambda λ as shown in Equation (2). Then, we recover the weight vector \mathbf{w} using Equation (3). Our method, esvm, takes as input gene expression data along with breast cancer drug responses obtained from the GEO database. The output is a list of the selected 100 genes according to the top 100 corresponding weights in \mathbf{w} . Then, we performed enrichment analysis by providing the output genes to two enrichment analysis tools, Enrichr and Metascape. When Enrichr is considered, our method outperformed the baseline methods by having more expressed genes in breast cancer cell lines, including MD-MB231, MCF7, and HS578T. Moreover, our method, esvm, had more expressed genes in breast tissue. For Metascape, our method identified important genes, including tumor suppressor genes (e.g., TP53 and BRCA1), TFs, drugs, drug targets, pathways, and biological processes that play a key role in understanding breast cancer drug response mechanisms.

As computational running time plays a key role in the gene selection process, our method, esvm, was way faster than the baseline SVM. Although lasso was the fastest method, esvm had more expressed genes in breast cancer cell lines as well as breast tissues when compared to lasso. Therefore, the improvements in esvm are attributed to (1) the superiority of computational efficiency when compared to the baseline SVM and (2) high-performance results measured using the AUC when compared to lasso. Another advantage

of esvm attributed to its computational efficiency is its computational feasibility to explicitly change the data representation and apply esvm to identify important genes. As a result, esvm can aid in analyzing gene expression data coupled with clinical data and other profiling datasets.

Drug responses in this study were classified as follows: For Dataset1, the treated groups were classified during the Phase II trial as (1) pathological complete response, referring to the absence of invasive cancer in the axilla and breast, and (2) residual disease, referring to the presence of invasive cancer in the breast and axilla. In terms of Dataset2, treated groups after neoadjuvant therapy within the Phase II multicenter trial were classified as (1) sensitive, referring to patients completely responding to the treatment, and (2) resistant, referring to patients not completely responding to the treatment. Regarding Dataset3, the treated groups were categorized during the neoadjuvant I-SPY2 trial as (1) complete response when responding to the treatment and (2) failed complete response when not completely responding to the treatment.

We specify the parameters of the methods in our study for gene selection as follows. esvm ($C = 2$), lasso ($\lambda = 0.05$), and SVM (setting parameters associated with the linear kernel to their default values). Each method produces a gene list, and the probability of each method coinciding with our method, esvm, is estimated as follows: For Dataset1, Dataset2, and Dataset3, the probabilities are equal to $\frac{1}{\binom{5267}{100}}$, $\frac{1}{\binom{5313}{100}}$, and $\frac{1}{\binom{118}{50}}$, respectively; $\binom{n}{k}$ indicates the binomial coefficient, in which n is the total number of genes in the considered dataset and k is the total number of genes produced by esvm. Therefore, the odds of having a method producing results like ours are unlikely to occur.

We evaluated the performance of selected genes from a biological perspective against deep learning methods, including DeepLIFT, DeepSHAP, and LRP. In the three datasets (see Supplementary Additional File: Tables S1(b), S2(b), and S3(b)), our method had more expressed genes in breast cancer cell lines (MD-MB231, MCF7, and HS578T). Specifically, in Dataset1, esvm had a total of eight expressed genes compared to six (two and four) for DeepLIFT (DeepSHAP and LRP). For Dataset2, esvm had a total of seven expressed genes, while DeepLIFT, DeepSHAP, and LRP had a total of five, five, and six, respectively. In terms of Dataset3, esvm had a total of three expressed genes, while each deep learning method had a total of two expressed genes. In our study, we identified genes enriched in terms related to breast cancer cell lines, and our method was the best. On the other hand, when considering genes in breast tissue irrespective of cell types, DeepLIFT was the best (see Supplementary Additional File: Tables S1(a), S2(a), and S3(a)). As a fraction of genes are expressed in each breast cancer cell line, our method had more expressed genes related to the studied breast cancer cell lines. Therefore, these results demonstrate the superiority of esvm. Additional details for the studied deep learning models are in the Supplementary Additional File (Figures S1 and S2) and Supplementary DLModels.

To further demonstrate the effectiveness of esvm as a gene selection method, we assessed the performance from a classification perspective of our method against adapted deep learning methods for gene selection, including Deep Learning Important Features (DeepLIFT) [8], Deep SHapley Additive exPlanations (DeepSHAP) [9], and Layer-wise relevance propagation (LRP) [10]. Tables S4–S6 in the Supplementary Additional File demonstrate the performance results (and standard deviation) when SVM is coupled with each gene set produced by each method for the three datasets. It can be seen from Table S4 in the Supplementary Additional File that when SVM is coupled with Dataset1 of selected genes via esvm, it generates the highest accuracy (ACC) of 0.921, the highest balanced accuracy (BAC) of 0.935, and the highest Matthews correlation coefficient (MCC) of 0.849. For Dataset2 (see Table S5 in Supplementary Additional File), SVM, when coupled with the gene set via esvm in Dataset2, generated the highest ACC of 0.978, the highest BAC of 0.960, the highest F1 of 0.950, and the highest MCC of 0.944. The same holds true for Dataset3, in which SVM, when coupled with a dataset of genes selected from esvm, achieved the highest performance results based on the four performance measures (see Table S6 in the Supplementary Additional File).

When evaluating the performance from a classification perspective for the three datasets using lasso (rather than SVM) as a learning algorithm (see Supplementary Additional File: Tables S7–S9), Table S7 in Supplementary Additional File, demonstrates that lasso when coupled with Dataset1 of genes selected via esvm generated the highest ACC of 0.662, the highest BAC of 0.660, the highest F1 of 0.738. In terms of Dataset2 (see Table S8 in Supplementary Additional File), lasso, when coupled with Dataset2 with selected genes from esvm, generated the highest BAC of 0.537, while other methods had marginal improvements over our method (esvm) when the ACC performance measure was considered. The same holds true for Dataset3, in which lasso, when coupled with Dataset3 of genes from esvm, achieved the highest performance results using two performance metrics in imbalanced classification. Particularly, lasso with genes from esvm achieved the highest BAC of 0.627 and the highest MCC of 0.280. These results demonstrate the effectiveness of our method in exploring discriminatory genes. Combined confusion matrices for performance results using SVM and lasso are provided in the Supplementary Additional File, displayed in Figures S3–S8, and the sum of entries in each confusion matrix is equal to the number of samples in the corresponding dataset.

In terms of the number of common genes in Dataset1, it can be seen from Figure 2a that the number of common genes (if they exist) between any pair of methods does not exceed 2. For example, limma and sam had two common genes (ATP2B1 and RNF186), esvm and *t*-test had two common genes (RAB38 and ABHD1), and sam and *t*-test had two common genes (MFN2 and RPS24). The rest of the intersections are provided in Supplementary DataSheet1_B. For Dataset2, Figure 4a shows that esvm and *t*-test had four common genes (PTCHD1, CTSB, ALDH2, and CHGB), while limma and sam had three common genes (FABP4, SNX32, and GBP6). In Supplementary DataSheet2_B, we include the rest of the common genes. For Dataset3, Figure 6a shows that esvm, sam, and *t*-tests had 13 common genes (CCND1, EGFR.3, EIF4GI, ERBB2, ERBB3.1, ESR1.1, IGF1R, IRS1, JAK2, KS6B1.2, PD1L1.3, STAT1, and STA5A). The rest of the intersections are provided in Supplementary DataSheet3_B. In Dataset1, Figure 2b–d are presented in tabular format in Tables S10–S12 in the Supplementary Additional File. For Dataset2, the corresponding tables for Figure 4b–d are Tables S13–S15 in the Supplementary Additional File. For Dataset3, Tables S16–S18 correspond to Figure 6b–d in the Supplementary Additional File.

It is worth noting that Tables 2, 4 and 6 were populated based on results from an enrichment analysis tool, Enrichr. For a set of terms in a category, testing the null hypothesis can tell if a user's gene list enriched in a given term (i.e., overlap column) is more than a random chance (or not) compared to that term's gene list in the human genome background [18]. The terms in a category are ranked by *p*-values, which were derived using Fisher's exact test, and the adjusted *p*-values were corrected based on the Benjamini–Hochberg procedure [15]. In our study, we had two categories in Enrichr: ARCHS4 tissues and NCI-60 cancer cell lines, reporting one term (BREAST (BULK TISSUE)) in the former and three terms (MD-MB231, MCF7, and HS578T) in the latter. These breast cancer cell lines are well-established in biology and medicine when analyzing breast cancer, as in [97–102].

5. Conclusions and Future Work

In this paper, we present a computational framework based on machine learning and enrichment analysis, unveiling critical genes, drugs, drug targets, and other biological knowledge underlying breast cancer drug response mechanisms. Our framework receives as an input a gene expression dataset pertaining to breast cancer patients responding and not responding to a treatment; we downloaded three different gene expression datasets from the GEO database according to the following accession numbers: GSE130787, GSE140494, and GSE196093. As $n \gg m$ arises challenge in computational genomics in which n and m are the numbers of genes and samples, respectively, we formulate our method according to the dual form in which we solve the optimization problem as a function of λ associated with m (instead of the formulation as a function of \mathbf{w} associated with n). Then, we perform mathematical calculations to efficiently recover \mathbf{w} as a function of three

inputs ($\lambda, \mathbf{y}, \mathbf{x}$) and then identify important p genes out of n , provided to enrichment analysis tools Enrichr and Metascape. In addition to (1) significantly achieving the highest performance results in terms of the area under the curve and (2) being more computationally efficient than the baseline SVM, results demonstrate that our method esvm outperformed existing baseline methods, including deep learning in (1) breast cancer cell line identification, showing more expressed genes; and (2) achieving the highest performance results from a classification perspective when coupled with SVM. Moreover, we reported several drugs (including tamoxifen, cisplatin, and erlotinib), 36 unique TFs (e.g., SP1, NFKB1, RELA), and 74 unique genes (including tumor suppression genes such as TP53, BRCA1, and RB1) that have been reported to be connected to drug response and resistance mechanisms, progression, and metastasis of breast cancer. We made our computational method available publicly on the maGENEgerZ web server at <https://aibio.shinyapps.io/maGENEgerZ/>.

Future work includes (1) utilizing our framework to unveil various biological knowledge behind drug response mechanisms related to different cancer types such as pancreatic cancer, liver cancer, and multiple myeloma; (2) collaborating with clinical research physicians to apply our tool to analyze drug response mechanisms in the neoadjuvant setting; and (3) integrating different profiling data related to cancer drug response and performing an assessment from a biological perspective.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/math12101536/s1>.

Author Contributions: T.T.: conceptualization, formal analysis, methodology, data curation, software, supervision, writing—original draft preparation, visualization, and investigation. Y.-h.T.: conceptualization, formal analysis, methodology, data curation, validation, and writing—original draft preparation. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported in part by funds from Chuo University (TOKUTEI KADAI KENKYU).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tang, J.; Ravikumar, B.; Alam, Z.; Rebane, A.; Vähä-Koskela, M.; Peddinti, G.; van Adrichem, A.J.; Wakkinen, J.; Jaiswal, A.; Karjalainen, E. Drug target commons: A community effort to build a consensus knowledge base for drug-target interactions. *Cell Chem. Biol.* **2018**, *25*, 224–229. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Du, Y.; Han, Y.; Wang, X.; Wang, H.; Qu, Y.; Guo, K.; Ma, W.; Fu, L. Identification of Immune-Related Breast Cancer Chemotherapy Resistance Genes via Bioinformatics Approaches. *Front. Oncol.* **2022**, *12*, 772723. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Wu, J.; Tian, Y.; Liu, W.; Zheng, H.; Xi, Y.; Yan, Y.; Hu, Y.; Liao, B.; Wang, M.; Tang, P. A novel twelve-gene signature to predict neoadjuvant chemotherapy response and prognosis in breast cancer. *Front. Immunol.* **2022**, *13*, 1035667. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Freitas, A.J.A.; Nunes, C.R.; Mano, M.S.; Causin, R.L.; Santana, I.V.V.; de Oliveira, M.A.; Calfa, S.; Silveira, H.C.S.; de Pádua Souza, C.; Marques, M.M.C. Gene expression alterations predict the pathological complete response in triple-negative breast cancer exploratory analysis of the NACATRINE trial. *Sci. Rep.* **2023**, *13*, 21411. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Stevens, L.E.; Peluffo, G.; Qiu, X.; Temko, D.; Fassl, A.; Li, Z.; Trinh, A.; Seehawer, M.; Jovanović, B.; Alečković, M. JAK–STAT Signaling in Inflammatory Breast Cancer Enables Chemotherapy-Resistant Cell States. *Cancer Res.* **2023**, *83*, 264–284. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Debets, D.O.; Stecker, K.E.; Piskopou, A.; Liefwaard, M.C.; Wesseling, J.; Sonke, G.S.; Lips, E.H.; Altelaar, M. Deep (phospho) proteomics profiling of pre-treatment needle biopsies identifies signatures of treatment resistance in HER²⁺ breast cancer. *Cell Rep. Med.* **2023**, *4*, 101203. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Miri, A.; Gharechahi, J.; Samiei Mosleh, I.; Sharifi, K.; Jajarmi, V. Identification of co-regulated genes associated with doxorubicin resistance in the MCF-7/ADR cancer cell line. *Front. Oncol.* **2023**, *13*, 1135836. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Jin, X.; Wang, J.; Wang, Z.; Pang, W.; Chen, Y.; Yang, L. Chromatin-modifying protein 4C (CHMP4C) affects breast cancer cell growth and doxorubicin resistance as a potential breast cancer therapeutic target. *J. Antibiot.* **2024**, *77*, 93–101. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Das, S.; Kundu, M.; Hassan, A.; Parekh, A.; Jena, B.C.; Mundre, S.; Banerjee, I.; Yetirajam, R.; Das, C.K.; Pradhan, A.K. A novel computational predictive biological approach distinguishes Integrin β 1 as a salient biomarker for breast cancer chemoresistance. *Biochim. Biophys. Acta (BBA)-Mol. Basis Dis.* **2023**, *1869*, 166702. [\[CrossRef\]](#)

10. Wu, M.; Zhao, Y.; Peng, N.; Tao, Z.; Chen, B. Identification of chemoresistance-associated microRNAs and hub genes in breast cancer using bioinformatics analysis. *Investig. New Drugs* **2021**, *39*, 705–712. [[CrossRef](#)]
11. Hoogstraat, M.; Lips, E.H.; Mayayo-Peralta, I.; Mulder, L.; Kristel, P.; van der Heijden, I.; Annunziato, S.; van Seijen, M.; Nederlof, P.M.; Sonke, G.S. Comprehensive characterization of pre-and post-treatment samples of breast cancer reveal potential mechanisms of chemotherapy resistance. *NPJ Breast Cancer* **2022**, *8*, 60. [[CrossRef](#)]
12. Kim, M.W.; Lee, H.; Lee, S.; Moon, S.; Kim, Y.; Kim, J.Y.; Kim, S.I.; Kim, J.Y. Drug-resistant profiles of extracellular vesicles predict therapeutic response in TNBC patients receiving neoadjuvant chemotherapy. *BMC Cancer* **2024**, *24*, 185. [[CrossRef](#)]
13. Raju, B.; Narendra, G.; Verma, H.; Silakari, O. Identification of chemoresistance associated key genes-miRNAs-TFs in docetaxel resistant breast cancer by bioinformatics analysis. *3 Biotech* **2024**, *14*, 128. [[CrossRef](#)]
14. Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **2016**, *44*, W90–W97. [[CrossRef](#)]
15. Xie, Z.; Bailey, A.; Kuleshov, M.V.; Clarke, D.J.; Evangelista, J.E.; Jenkins, S.L.; Lachmann, A.; Wojciechowicz, M.L.; Kropiwnicki, E.; Jagodnik, K.M. Gene set knowledge discovery with enrichr. *Curr. Protoc.* **2021**, *1*, e90. [[CrossRef](#)]
16. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A.H.; Tanaseichuk, O.; Benner, C.; Chanda, S.K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **2019**, *10*, 1523. [[CrossRef](#)]
17. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [[CrossRef](#)]
18. Tusher, V.G.; Tibshirani, R.; Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5116–5121. [[CrossRef](#)]
19. Team, R.C.; Team, M.R.C.; Suggests, M.; Matrix, S. *Package Stats*, R Stats Package. 2018.
20. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1. [[CrossRef](#)]
21. Shrikumar, A.; Greenside, P.; Kundaje, A. Learning important features through propagating activation differences. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 3145–3153.
22. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
23. Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.-R.; Samek, W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE* **2015**, *10*, e0130140. [[CrossRef](#)] [[PubMed](#)]
24. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2012**, *41*, D991–D995. [[CrossRef](#)] [[PubMed](#)]
25. Povey, S.; Lovering, R.; Bruford, E.; Wright, M.; Lush, M.; Wain, H. The HUGO gene nomenclature committee (HGNC). *Hum. Genet.* **2001**, *109*, 678–680. [[CrossRef](#)] [[PubMed](#)]
26. Durinck, S.; Spellman, P.T.; Birney, E.; Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **2009**, *4*, 1184–1191. [[CrossRef](#)] [[PubMed](#)]
27. Bi, M.; Zhang, Z.; Jiang, Y.-Z.; Xue, P.; Wang, H.; Lai, Z.; Fu, X.; De Angelis, C.; Gong, Y.; Gao, Z. Enhancer reprogramming driven by high-order assemblies of transcription factors promotes phenotypic plasticity and breast cancer endocrine resistance. *Nat. Cell Biol.* **2020**, *22*, 701–715. [[CrossRef](#)] [[PubMed](#)]
28. Turki, T.; Taguchi, Y.H. GENEvaRX: A novel AI-driven method and web tool can identify critical genes and effective drugs for Lichen Planus. *Eng. Appl. Artif. Intell.* **2023**, *124*, 106607. [[CrossRef](#)]
29. Turki, T.; Taguchi, Y. A new machine learning based computational framework identifies therapeutic targets and unveils influential genes in pancreatic islet cells. *Gene* **2022**, *853*, 147038. [[CrossRef](#)] [[PubMed](#)]
30. Taguchi, Y.; Turki, T. Adapted tensor decomposition and PCA based unsupervised feature extraction select more biologically reasonable differentially expressed genes than conventional methods. *Sci. Rep.* **2022**, *12*, 17438. [[CrossRef](#)]
31. Mohri, M.; Rostamizadeh, A.; Talwalkar, A. *Foundations of Machine Learning*; MIT Press: Cambridge, MA, USA, 2018.
32. Tan, P.-N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*; Pearson Education India: Noida, India, 2016.
33. Fu, A.; Narasimhan, B.; Boyd, S. CVXR: An R Package for Disciplined Convex Optimization. *J. Stat. Softw.* **2020**, *94*, 1–34. [[CrossRef](#)]
34. Team, R.C. R: A Language and Environment for Statistical Computing. *J. Stat. Softw.* **2008**, *25*. Available online: <https://www.r-project.org/> (accessed on 4 April 2024).
35. Schwender, H. siggenes: Multiple testing using SAM and Efron’s empirical Bayes approaches. *R Package Version* **2012**, *1*, 1–70.
36. Taguchi, Y.; Turki, T. A new advanced in silico drug discovery method for novel coronavirus (SARS-CoV-2) with tensor decomposition-based unsupervised feature extraction. *PLoS ONE* **2020**, *15*, e0238907. [[CrossRef](#)] [[PubMed](#)]
37. Koenen, N.; Wright, M.N. Interpreting Deep Neural Networks with the Package insight. *arXiv* **2023**, arXiv:2306.10822.
38. Davis, S.; Meltzer, P.S. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **2007**, *23*, 1846–1847. [[CrossRef](#)] [[PubMed](#)]
39. Huber, W.; Carey, V.J.; Gentleman, R.; Anders, S.; Carlson, M.; Carvalho, B.S.; Bravo, H.C.; Davis, S.; Gatto, L.; Girke, T. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **2015**, *12*, 115–121. [[CrossRef](#)] [[PubMed](#)]
40. Stekhoven, D.J.; Stekhoven, M.D.J. Package ‘missForest’. *R Package Version* **2013**, *1*, 21.

41. Stekhoven, D.J.; Bühlmann, P. MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics* **2011**, *28*, 112–118. [[CrossRef](#)] [[PubMed](#)]
42. Zhao, W.; Chang, Y.; Wu, Z.; Jiang, X.; Li, Y.; Xie, R.; Fu, D.; Sun, C.; Gao, J. Identification of PIMREG as a novel prognostic signature in breast cancer via integrated bioinformatics analysis and experimental validation. *PeerJ* **2023**, *11*, e15703. [[CrossRef](#)]
43. Dai, Y.-H.; Wang, Y.-F.; Shen, P.-C.; Lo, C.-H.; Yang, J.-F.; Lin, C.-S.; Chao, H.-L.; Huang, W.-Y. Gene-associated methylation status of ST14 as a predictor of survival and hormone receptor positivity in breast Cancer. *BMC Cancer* **2021**, *21*, 945. [[CrossRef](#)]
44. Furuminato, K.; Minatoya, S.; Senoo, E.; Goto, T.; Yamazaki, S.; Sakaguchi, M.; Toyota, K.; Iguchi, T.; Miyagawa, S. The role of mesenchymal estrogen receptor 1 in mouse uterus in response to estrogen. *Sci. Rep.* **2023**, *13*, 12293. [[CrossRef](#)]
45. Zhang, X.; Zhao, P.; Ma, M.; Wu, H.; Liu, R.; Liu, Z.; Cai, Z.; Liu, M.; Xie, F.; Ma, X. Missing link between tissue specific expressing pattern of ER β and the clinical manifestations in LGBLEL. *Front. Med.* **2023**, *10*, 1168977. [[CrossRef](#)] [[PubMed](#)]
46. Li, J.; Shu, X.; Xu, J.; Su, S.M.; Chan, U.I.; Mo, L.; Liu, J.; Zhang, X.; Adhav, R.; Chen, Q. S100A9-CXCL12 activation in BRCA1-mutant breast cancer promotes an immunosuppressive microenvironment associated with resistance to immunotherapy. *Nat. Commun.* **2022**, *13*, 1481. [[CrossRef](#)] [[PubMed](#)]
47. Choi, E.; Lee, J.; Lee, H.; Cho, J.; Lee, Y.-S. BRCA1 deficiency in triple-negative breast cancer: Protein stability as a basis for therapy. *Biomed. Pharmacother.* **2023**, *158*, 114090. [[CrossRef](#)] [[PubMed](#)]
48. Sachsenweger, J.; Jansche, R.; Merk, T.; Heitmeir, B.; Deniz, M.; Faust, U.; Roggia, C.; Tzschach, A.; Schroeder, C.; Riess, A. ABRAXAS1 orchestrates BRCA1 activities to counter genome destabilizing repair pathways—Lessons from breast cancer patients. *Cell Death Dis.* **2023**, *14*, 328. [[CrossRef](#)] [[PubMed](#)]
49. Tournant, F. Stromal cells drive tumorigenesis in BRCA1 mutation carriers. *Nat. Rev. Cancer* **2023**, *23*, 349. [[CrossRef](#)] [[PubMed](#)]
50. Gajda-Walczyk, A.; Potęga, A.; Kowalczyk, A.; Sek, S.; Zięba, S.; Kowalik, A.; Kudelski, A.; Nowicka, A.M. New, fast and cheap prediction tests for BRCA1 gene mutations identification in clinical samples. *Sci. Rep.* **2023**, *13*, 7316. [[CrossRef](#)] [[PubMed](#)]
51. Chakraborty, S.; Banerjee, S. Multidimensional computational study to understand non-coding RNA interactions in breast cancer metastasis. *Sci. Rep.* **2023**, *13*, 15771. [[CrossRef](#)] [[PubMed](#)]
52. Yang, W.; Li, J.; Zhang, M.; Yu, H.; Zhuang, Y.; Zhao, L.; Ren, L.; Gong, J.; Bi, H.; Zeng, L.; et al. Elevated expression of the rhythm gene NFIL3 promotes the progression of TNBC by activating NF- κ B signaling through suppression of NFKBIA transcription. *J. Exp. Clin. Cancer Res.* **2022**, *41*, 67. [[CrossRef](#)] [[PubMed](#)]
53. Chen, S.; Shao, F.; Zeng, J.; Guo, S.; Wang, L.; Sun, H.; Lei, J.H.; Lyu, X.; Gao, S.; Chen, Q.; et al. Cullin-5 deficiency orchestrates the tumor microenvironment to promote mammary tumor development through CREB1-CCL2 signaling. *Sci. Adv.* **2023**, *9*, eabq1395. [[CrossRef](#)]
54. Huss, L.; Butt, S.T.; Borgquist, S.; Elebro, K.; Sandsveden, M.; Rosendahl, A.; Manjer, J. Vitamin D receptor expression in invasive breast tumors and breast cancer survival. *Breast Cancer Res.* **2019**, *21*, 84. [[CrossRef](#)]
55. Sannappa Gowda, N.G.; Shiragannavar, V.D.; Puttahanumantharayappa, L.D.; Shivakumar, A.T.; Dallavalasa, S.; Basavaraju, C.G.; Bhat, S.S.; Prasad, S.K.; Vamadevaiah, R.M.; Madhunapantula, S.V.; et al. Quercetin activates vitamin D receptor and ameliorates breast cancer induced hepatic inflammation and fibrosis. *Front. Nutr.* **2023**, *10*, 1158633. [[CrossRef](#)] [[PubMed](#)]
56. Ray, A.; Provenzano, P.P. Aligned forces: Origins and mechanisms of cancer dissemination guided by extracellular matrix architecture. *Curr. Opin. Cell Biol.* **2021**, *72*, 63–71. [[CrossRef](#)] [[PubMed](#)]
57. Jones, C.E.; Sharick, J.T.; Sizemore, S.T.; Cukierman, E.; Strohecker, A.M.; Leight, J.L. A miniaturized screening platform to identify novel regulators of extracellular matrix alignment. *Cancer Res. Commun.* **2022**, *2*, 1471–1486. [[CrossRef](#)] [[PubMed](#)]
58. Wiseman, B.S.; Werb, Z. Stromal effects on mammary gland development and breast cancer. *Science* **2002**, *296*, 1046–1049. [[CrossRef](#)]
59. Hannan, F.M.; Elajnaf, T.; Vandenberg, L.N.; Kennedy, S.H.; Thakker, R.V. Hormonal regulation of mammary gland development and lactation. *Nat. Rev. Endocrinol.* **2023**, *19*, 46–61. [[CrossRef](#)] [[PubMed](#)]
60. Batbayar, G.; Ishimura, A.; Lyu, H.; Wanna-Udom, S.; Meguro-Horike, M.; Terashima, M.; Horike, S.-I.; Takino, T.; Suzuki, T. ASH2L, a COMPASS core subunit, is involved in the cell invasion and migration of triple-negative breast cancer cells through the epigenetic control of histone H3 lysine 4 methylation. *Biochem. Biophys. Res. Commun.* **2023**, *669*, 19–29. [[CrossRef](#)] [[PubMed](#)]
61. Bradley, R.; Braybrooke, J.; Gray, R.; Hills, R.K.; Liu, Z.; Pan, H.; Peto, R.; Dodwell, D.; McGale, P.; Taylor, C. Aromatase inhibitors versus tamoxifen in premenopausal women with oestrogen receptor-positive early-stage breast cancer treated with ovarian suppression: A patient-level meta-analysis of 7030 women from four randomised trials. *Lancet Oncol.* **2022**, *23*, 382–392. [[CrossRef](#)] [[PubMed](#)]
62. Wang, J.; Cai, L.; Song, Y.; Sun, T.; Tong, Z.; Teng, Y.; Li, H.; Ouyang, Q.; Chen, Q.; Cui, S. Clinical efficacy of fulvestrant versus exemestane as first-line therapies for Chinese postmenopausal oestrogen-receptor positive/human epidermal growth factor receptor 2-advanced breast cancer (FRIEND study). *Eur. J. Cancer* **2023**, *184*, 73–82. [[CrossRef](#)] [[PubMed](#)]
63. Torrisi, R.; Vaira, V.; Giordano, L.; Destro, A.; Basilico, V.; Mazzara, S.; Salvini, P.; Gaudio, G.; Fernandes, B.; Rudini, N. Predictors of fulvestrant long-term benefit in hormone receptor-positive/HER2 negative advanced breast cancer. *Sci. Rep.* **2022**, *12*, 12789. [[CrossRef](#)]
64. Sultan, M.H.; Moni, S.S.; Madkhali, O.A.; Bakkari, M.A.; Alshahrani, S.; Alqahtani, S.S.; Alhakamy, N.A.; Mohan, S.; Ghazwani, M.; Bukhary, H.A. Characterization of cisplatin-loaded chitosan nanoparticles and rituximab-linked surfaces as target-specific injectable nano-formulations for combating cancer. *Sci. Rep.* **2022**, *12*, 468. [[CrossRef](#)]

65. Fatehi, R.; Rashedinia, M.; Akbarizadeh, A.R.; Firouzabadi, N. Metformin enhances anti-cancer properties of resveratrol in MCF-7 breast cancer cells via induction of apoptosis, autophagy and alteration in cell cycle distribution. *Biochem. Biophys. Res. Commun.* **2023**, *644*, 130–139. [[CrossRef](#)] [[PubMed](#)]
66. Bu, J.; Zhang, Y.; Niu, N.; Bi, K.; Sun, L.; Qiao, X.; Wang, Y.; Zhang, Y.; Jiang, X.; Wang, D. Dalpiciclib partially abrogates ER signaling activation induced by pyrotinib in HER²⁺ HR⁺ breast cancer. *life* **2023**, *12*, e85246. [[CrossRef](#)] [[PubMed](#)]
67. Bjørklund, S.S.; Aure, M.R.; Häkkinen, J.; Vallon-Christersson, J.; Kumar, S.; Evensen, K.B.; Fleischer, T.; Tost, J.; Bathen, T.F.; Borgen, E.; et al. Subtype and cell type specific expression of lncRNAs provide insight into breast cancer. *Commun. Biol.* **2022**, *5*, 834. [[CrossRef](#)] [[PubMed](#)]
68. Kanyomse, Q.; Le, X.; Tang, J.; Dai, F.; Mobet, Y.; Chen, C.; Cheng, Z.; Deng, C.; Ning, Y.; Yu, R.; et al. KLF15 suppresses tumor growth and metastasis in Triple-Negative Breast Cancer by downregulating CCL2 and CCL7. *Sci. Rep.* **2022**, *12*, 19026. [[CrossRef](#)] [[PubMed](#)]
69. Chen, H.; Wang, X.; Lan, X.; Yu, T.; Li, L.; Tang, S.; Liu, S.; Jiang, F.; Wang, L.; Zhang, J. A radiomics model development via the associations with genomics features in predicting axillary lymph node metastasis of breast cancer: A study based on a public database and single-centre verification. *Clin. Radiol.* **2023**, *78*, e279–e287. [[CrossRef](#)] [[PubMed](#)]
70. Liao, H.; Li, H.; Song, J.; Chen, H.; Si, H.; Dong, J.; Wang, J.; Bai, X. Expression of the prognostic marker IL-8 correlates with the immune signature and epithelial-mesenchymal transition in breast cancer. *J. Clin. Lab. Anal.* **2023**, *37*, e24797. [[CrossRef](#)] [[PubMed](#)]
71. Bi, J.; Wu, Z.; Zhang, X.; Zeng, T.; Dai, W.; Qiu, N.; Xu, M.; Qiao, Y.; Ke, L.; Zhao, J. TMEM25 inhibits monomeric EGFR-mediated STAT3 activation in basal state to suppress triple-negative breast cancer progression. *Nat. Commun.* **2023**, *14*, 2342. [[CrossRef](#)] [[PubMed](#)]
72. Huo, Y.; Li, X.; Xu, P.; Bao, Z.; Liu, W. Analysis of breast cancer based on the dysregulated network. *Front. Genet.* **2022**, *13*, 856075. [[CrossRef](#)] [[PubMed](#)]
73. Ring, A.; Kaur, P.; Lang, J.E. EP300 knockdown reduces cancer stem cell phenotype, tumor growth and metastasis in triple negative breast cancer. *BMC Cancer* **2020**, *20*, 1076. [[CrossRef](#)]
74. Li, C.H.; Fang, C.Y.; Chan, M.H.; Lu, P.J.; Ger, L.P.; Chu, J.S.; Chang, Y.C.; Chen, C.L.; Hsiao, M. The activation of EP300 by F11R leads to EMT and acts as a prognostic factor in triple-negative breast cancers. *J. Pathol. Clin. Res.* **2023**, *9*, 165–181. [[CrossRef](#)]
75. Ma, S.; Tang, T.; Probst, G.; Konradi, A.; Jin, C.; Li, F.; Gutkind, J.S.; Fu, X.-D.; Guan, K.-L. Transcriptional repression of estrogen receptor alpha by YAP reveals the Hippo pathway as therapeutic target for ER⁺ breast cancer. *Nat. Commun.* **2022**, *13*, 1061. [[CrossRef](#)] [[PubMed](#)]
76. Mo, D.; Jiang, P.; Yang, Y.; Mao, X.; Tan, X.; Tang, X.; Wei, D.; Li, B.; Wang, X.; Tang, L. A tRNA fragment, 5'-tiRNAVal, suppresses the Wnt/ β -catenin signaling pathway by targeting FZD3 in breast cancer. *Cancer Lett.* **2019**, *457*, 60–73. [[CrossRef](#)] [[PubMed](#)]
77. McBean, B.N.; Michmerhuizen, A.R.; Wilder-Romans, K.; Chandler, B.C.; Lerner, L.M.; Ward, C.; Liu, M.; Boyle, A.P.; Speers, C.W. Molecular mechanisms of intrinsic radioresistance in breast cancer. *Cancer Res.* **2023**, *83*, 2401. [[CrossRef](#)]
78. Zuo, Z.; Zhou, Z.; Chang, Y.; Liu, Y.; Shen, Y.; Li, Q.; Zhang, L. Ribonucleotide reductase M2 (RRM2): Regulation, function and targeting strategy in human cancer. *Genes Dis.* **2024**, *11*, 218–233. [[CrossRef](#)]
79. Gordon, D.; Croushore, E.; Koppenhafer, S.; Goss, K.; Geary, E. Activator protein-1 (AP-1) signaling inhibits the growth of Ewing sarcoma cells in response to DNA replication stress. *Cancer Res.* **2023**, *83*, 3532. [[CrossRef](#)]
80. Rudd, S.G.; Tsesmetzis, N.; Sanjiv, K.; Paulin, C.B.; Sandhow, L.; Kutzner, J.; Hed Myrberg, I.; Bunten, S.S.; Axelsson, H.; Zhang, S.M. Ribonucleotide reductase inhibitors suppress SAMHD1 ara-CTPase activity enhancing cytarabine efficacy. *EMBO Mol. Med.* **2020**, *12*, e10419. [[CrossRef](#)]
81. Li, X.; Zhao, L.; Chen, C.; Nie, J.; Jiao, B. Can EGFR be a therapeutic target in breast cancer? *Biochim. Biophys. Acta (BBA)-Rev. Cancer*, 2022; 1877, 188789.
82. Wang, Y.; Du, L.; Jing, J.; Zhao, X.; Wang, X.; Hou, S. Leptin and leptin receptor expression as biomarkers for breast cancer: A retrospective study. *BMC Cancer* **2023**, *23*, 260. [[CrossRef](#)]
83. Hu, Y.; Liu, L.; Chen, Y.; Zhang, X.; Zhou, H.; Hu, S.; Li, X.; Li, M.; Li, J.; Cheng, S.; et al. Cancer-cell-secreted miR-204-5p induces leptin signalling pathway in white adipose tissue to promote cancer-associated cachexia. *Nat. Commun.* **2023**, *14*, 5179. [[CrossRef](#)] [[PubMed](#)]
84. Katzenellenbogen, B.S.; Guillen, V.S.; Katzenellenbogen, J.A. Targeting the oncogenic transcription factor FOXM1 to improve outcomes in all subtypes of breast cancer. *Breast Cancer Res.* **2023**, *25*, 76. [[CrossRef](#)]
85. Ziegler, Y.; Laws, M.J.; Sanabria Guillen, V.; Kim, S.H.; Dey, P.; Smith, B.P.; Gong, P.; Bindman, N.; Zhao, Y.; Carlson, K.; et al. Suppression of FOXM1 activities and breast cancer growth in vitro and in vivo by a new class of compounds. *NPJ Breast Cancer* **2019**, *5*, 45. [[CrossRef](#)]
86. Shahbandi, A.; Nguyen, H.D.; Jackson, J.G. TP53 mutations and outcomes in breast cancer: Reading beyond the headlines. *Trends Cancer* **2020**, *6*, 98–110. [[CrossRef](#)] [[PubMed](#)]
87. Mitri, Z.I.; Abuhadra, N.; Goodyear, S.M.; Hobbs, E.A.; Kaempf, A.; Thompson, A.M.; Moulder, S.L. Impact of TP53 mutations in Triple Negative Breast Cancer. *npj Precis. Oncol.* **2022**, *6*, 64. [[CrossRef](#)] [[PubMed](#)]
88. Liu, F.; Xie, B.; Ye, R.; Xie, Y.; Zhong, B.; Zhu, J.; Tang, Y.; Lin, Z.; Tang, H.; Wu, Z. Overexpression of tripartite motif-containing 47 (TRIM47) confers sensitivity to PARP inhibition via ubiquitylation of BRCA1 in triple negative breast cancer cells. *Oncogenesis* **2023**, *12*, 13. [[CrossRef](#)] [[PubMed](#)]

89. Mateos, M.R.-C.; Santiago-Freijanes, P.; Röder, J.; Oberoi, P.; Vigo, N.; Almenar, E.; Chucla, T.C.; Mosquera, J.; Acea-Nebriil, B.; Wels, W. 17P New therapeutic target in triple-negative breast cancer for enhancing PARP inhibitor efficacy and stimulating the anti-tumour immune response. *ESMO Open* **2023**, *8*, 100983. [[CrossRef](#)]
90. Hu, Z.; Wei, F.; Su, Y.; Wang, Y.; Shen, Y.; Fang, Y.; Ding, J.; Chen, Y. Histone deacetylase inhibitors promote breast cancer metastasis by elevating NEDD9 expression. *Signal Transduct. Target. Ther.* **2023**, *8*, 11. [[CrossRef](#)]
91. Khanal, P.; Patil, V.S.; Bhandare, V.V.; Patil, P.P.; Patil, B.; Dwivedi, P.S.; Bhattacharya, K.; Harish, D.R.; Roy, S. Systems and in vitro pharmacology profiling of diosgenin against breast cancer. *Front. Pharmacol.* **2023**, *13*, 1052849. [[CrossRef](#)] [[PubMed](#)]
92. Deng, Z.; Chen, G.; Shi, Y.; Lin, Y.; Ou, J.; Zhu, H.; Wu, J.; Li, G.; Lv, L. Curcumin and its nano-formulations: Defining triple-negative breast cancer targets through network pharmacology, molecular docking, and experimental verification. *Front. Pharmacol.* **2022**, *13*, 920514. [[CrossRef](#)] [[PubMed](#)]
93. Kong, X.; Yan, W.; Sun, W.; Zhang, Y.; Yang, H.J.; Chen, M.; Chen, H.; de Vere White, R.W.; Zhang, J.; Chen, X. Isoform-specific disruption of the TP73 gene reveals a critical role for TAp73 γ in tumorigenesis via leptin. *eLife* **2023**, *12*, e82115. [[CrossRef](#)] [[PubMed](#)]
94. Lin, X.; Chen, D.; Chu, X.; Luo, L.; Liu, Z.; Chen, J. Oxypalmitate regulates proliferation and apoptosis of breast cancer cells by inhibiting PI3K/AKT signaling and its efficacy against breast cancer organoids. *Phytomedicine* **2023**, *114*, 154752. [[CrossRef](#)]
95. Yuan, L.; Cai, Y.; Zhang, L.; Liu, S.; Li, P.; Li, X. Promoting apoptosis, a promising way to treat breast cancer with natural products: A comprehensive review. *Front. Pharmacol.* **2022**, *12*, 801662. [[CrossRef](#)]
96. Dong, S.; Yousefi, H.; Savage, I.V.; Okpechi, S.C.; Wright, M.K.; Matossian, M.D.; Collins-Burow, B.M.; Burow, M.E.; Alahari, S.K. Ceritinib is a novel triple negative breast cancer therapeutic agent. *Mol. Cancer* **2022**, *21*, 138. [[CrossRef](#)] [[PubMed](#)]
97. Jiang, W.; Zhang, M.; Gao, C.; Yan, C.; Gao, R.; He, Z.; Wei, X.; Xiong, J.; Ruan, Z.; Yang, Q. A mitochondrial EglN1-AMPK α axis drives breast cancer progression by enhancing metabolic adaptation to hypoxic stress. *EMBO J.* **2023**, *42*, e113743. [[CrossRef](#)] [[PubMed](#)]
98. Chen, J.; Ma, D.; Zeng, C.; White, L.V.; Zhang, H.; Teng, Y.; Lan, P. Solasodine suppress MCF7 breast cancer stem-like cells via targeting Hedgehog/Gli1. *Phytomedicine* **2022**, *107*, 154448. [[CrossRef](#)] [[PubMed](#)]
99. Balogh, B.; Vecsernyés, M.; Veres-Székely, A.; Berta, G.; Stayer-Harci, A.; Tarjányi, O.; Sétáló Jr, G. Urocortin stimulates ERK1/2 phosphorylation and proliferation but reduces ATP production of MCF7 breast cancer cells. *Mol. Cell. Endocrinol.* **2022**, *547*, 111610. [[CrossRef](#)] [[PubMed](#)]
100. Janacova, L.; Stenckova, M.; Lapcik, P.; Hrachovinova, S.; Bouchalova, P.; Potesil, D.; Hrstka, R.; Müller, P.; Bouchal, P. Catechol-O-methyl transferase suppresses cell invasion and interplays with MET signaling in estrogen dependent breast cancer. *Sci. Rep.* **2023**, *13*, 1285. [[CrossRef](#)] [[PubMed](#)]
101. Park, J.D.; Jang, H.J.; Choi, S.H.; Jo, G.H.; Choi, J.-H.; Hwang, S.; Park, W.; Park, K.-S. The ELK3-DRP1 axis determines the chemosensitivity of triple-negative breast cancer cells to CDDP by regulating mitochondrial dynamics. *Cell Death Discov.* **2023**, *9*, 237. [[CrossRef](#)]
102. Lei, M.; Zhang, Y.-L.; Huang, F.-Y.; Chen, H.-Y.; Chen, M.-H.; Wu, R.-H.; Dai, S.-Z.; He, G.-S.; Tan, G.-H.; Zheng, W.-P. Gankyrin inhibits ferroptosis through the p53/SLC7A11/GPX4 axis in triple-negative breast cancer cells. *Sci. Rep.* **2023**, *13*, 21916. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.