

Article

Not peer-reviewed version

Secondary Analysis of Human Bulk RNA-seq Dataset Suggest Potential Mechanisms and Transcriptional Biomarkers for Letrozole Resistance in Estrogen Positive (ER+) Breast Cancer

Lincoln Sutherland , Jacob Lang , [Norberto Gonzalez-Juarbe](#) , [Brett E. Pickett](#) *

Posted Date: 14 May 2024

doi: 10.20944/preprints202405.0904.v1

Keywords: estrogen receptor; breast cancer; treatment resistance; Letrozole; biomarker



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Secondary Analysis of Human Bulk RNA-seq Dataset Suggest Potential Mechanisms and Transcriptional Biomarkers for Letrozole Resistance in Estrogen Positive (ER+) Breast Cancer

Lincoln Sutherland ¹, Jacob Lang ¹, Norberto Gonzalez-Juarbe ^{2,3} and Brett E. Pickett ^{1,*}

¹ Department of Microbiology and Molecular Biology; Brigham Young University; Provo, Utah 84602

² J. Craig Venter Institute; Rockville, Maryland 20850

³ Department of Cell Biology and Molecular Genetics; University of Maryland; College Park, Maryland 20742

* Correspondence: Author: Brett E. Pickett; brett_pickett@byu.edu

Abstract: Estrogen receptor-positive (ER+) breast cancer is common among postmenopausal women and is frequently treated with Letrozole. Letrozole inhibits aromatase from synthesizing estrogen from androgens. Decreased estrogen slows the growth of tumors and can be an effective treatment. Letrozole resistance is increasingly being observed, which poses a unique problem for patients. We reanalyzed transcriptomic data comparing individuals who responded to Letrozole therapy (responders) to those who were resistant to treatment (non-responders). We identified SOX11 and S100A9 as two genes that significantly differed between these patients; with “PLK1 signaling events” being the most significant signaling pathway. We also identified PRDX4 and E2F8 as the top transcriptional biomarkers for ER+ treatment resistance. Many of the other significant gene products play a known role in ER+ breast cancer or other types of cancer, which partially validate our results. Several of the gene products we identified do not have a characterized role in ER+ breast cancer. Our analysis identified a set of genes that warrant further research to elucidate the molecular mechanism of Letrozole resistance in this patient population, as well as novel high-performing biomarkers. We anticipate that these findings could contribute to improved therapeutic outcomes in human patients.

Keywords: estrogen receptor; breast cancer; treatment resistance; Letrozole; biomarker

1. Introduction

Breast cancer is the second leading cause of cancer death in women [1]. Approximately 13% of women will develop invasive breast cancer, with 3% of women dying from it [1]. Roughly 83% of invasive breast cancers are diagnosed in post-menopausal women (50 years or older), with 91% of the deaths occurring in the same age group [1,2]. Given that those who are 65 years of age or older are the largest growing subpopulation in the United States (16.8% of the total population in 2020) [3], the overall number of new breast cancer cases is expected to rise accordingly. The American Cancer Society predicts that 310,720 women in the United States will be diagnosed with breast cancer in 2024 [4], which is expected to rise to 440,000 by 2030 [5].

Breast cancer can be classified as a heterogenous disease. Among the various types of breast cancer, some types have an affinity for particular hormones that contribute to rapid tumor growth. Three important markers that are generally used to classify breast cancer involve the positive or negative presence of Estrogen Receptor (ER), Progesterone Receptor (PR), and/or the human epidermal growth factor receptor 2 (HER2) protein on the cell surface [6]. The Estrogen and Progesterone receptors both belong to the steroid receptor subgroup of ligand-activated transcription factors within the nuclear receptor superfamily [7]. Their presence are strong predictors for the efficacy of adjuvant and therapeutic hormonal therapy [7]. Amplification of the HER2 gene in tumor cells leads to increased presence of the protein at the cellular membrane, leading to potent

proliferative and anti-cell death signaling [8,9]. Taken together, ER, PR and HER2 play a central role in the detection, diagnosis and therapy of breast cancer [10–13].

Given that estrogen contributes to the promotion and progression of breast cancer, some therapeutic strategies specifically target estrogen synthesis via ER and its associated intracellular signaling pathways [14]. The aromatase enzyme then catalyzes the final conversion of androgens to estrogens [14,15]. Thus, aromatase was identified as a potential therapeutic target in the late 1970's with the synthesis of aminoglutethimide as the first aromatase inhibitor (AI) [16]. The low specificity and selectivity of this early therapeutic led to the synthesis of subsequent aromatase inhibitors with the most recent being Letrozole [16]. Letrozole has improved binding affinity to the catalytic site of the aromatase, which competitively inhibits the binding of the precursor reactant to the aromatase enzyme [14]. Moreover, Letrozole has greater potency than other AIs including anastrozole, exemestane, formestane, and aminoglutethimide [14]. Seventy-seven percent of all breast cancers are ER-positive (ER+), thus treatment with aromatase inhibitors is currently the standard treatment for all postmenopausal women [7]. Although Letrozole is one of the most common treatments for ER+ breast cancer it also has limitations [17,18]. A prior study showed that the overall response rate (ORR) among a large multi-national study of postmenopausal women was 55% [17]. ORR quantitatively represents the percentage of patients that achieve a favorable response, either total (complete disappearance of cancerous tissue) or partial (defined as a decrease in tumor size by at least 30%) [19]. Response rates are determined by either a mammogram or an ultrasound [17]. To our knowledge, the ability of existing prognostic methods to accurately identify treatment resistance prior to treatment are lacking.

A prior study examining the intracellular transcriptional response to Letrozole treatment used RNA-sequencing data to identify Proline-Rich Protein 11 (*PRR11*) as the only significantly overexpressed gene among 51 genes in chromosome arm 17q23. Focusing their analysis and validation work on this single locus in treatment-sensitive vs treatment-resistant ER+ breast cancer patients was useful [20], and amplification of this locus has been shown to correlate with poor clinical outcomes in breast cancer [21].

The aim of our current study is to perform a secondary analysis of the complete transcriptome from the original study RNA-sequencing study in treatment-sensitive vs treatment-resistant ER+ breast cancer patients. We expect that re-analyzing this valuable dataset will identify genes, signaling pathways, and transcriptional biomarkers across the whole human transcriptome that are significantly associated with ER+ treatment resistance.

2. Materials and Methods

2.1. Retrieving Fastq Files, Preprocessing, and Enriching RNA-sequencing Data

The sratools software package was used to download the RNA-sequencing fastq files from the desired study (Gene Expression identifier: GSE145325) from the Sequence Read Archive (SRA) database at NCBI [20,22]. The fastq files were then pre-processed and analyzed using the snakemake-based Automated Reproducible Modular Workflow for Preprocessing and Differential Analysis of RNA-seq Data (ARMOR) [23,24]. This preprocessing workflow applies the following methods using established algorithms: 1) quality control with fastqc [25], using default parameters; 2) trimming with TrimGalore! [26], using default parameters other than a minimum quality score of 20, a phred scale of 33, and a minimum length of 20 bases; 3) pseudomapping and quantification to the human transcriptome (GRCh38 release 98) using Salmon [27], with a k-mer setting of 31, read length of 63, and the following flags: seqBias --gcBias --fldMean 250 --fldSD 25; 4) read count normalization and differential gene expression using edgeR [27,28], with default parameters; and 5) functional enrichment using CAMERA [29], with default parameters enriching against the H, C2, and C5 Molecular Signature database gene sets. Although the trimming and/or differential expression algorithms may change, each of these steps is commonly applied in preprocessing RNA-sequencing data.

The differential gene expression results were then used as input to the Signaling Pathway Impact Analysis (SPIA) algorithm [30], with default parameters other than using 2,000 bootstrap replicates and a Bonferroni-adjusted p-value cutoff of 0.05. This algorithm uses bootstrapping and a null distribution to calculate significantly perturbed pathways that are enriched in statistically significant DEGs, regardless of fold-change value or direction [31–33]. The SPIA method was specifically chosen

because of its ability to take into account the topology of gene products that contribute to a pathway, rather than simply calculating whether there is significant overlap between the significant genes and the pathway, which decreases the number of false-positive results. Another benefit is that this algorithm uses bootstrapping and permutation to calculate a custom null distribution for each pathway. This null distribution is then used to determine a specific and unique p-value for the differentially expressed genes in the dataset for each pathway. This makes it outperform many other hypergeometric-based methods and enables it to predict whether any significant pathway is activated or inhibited without needing to perform separate analyses of upregulated and downregulated DEGs [30–33].

2.2. Target and Biomarker Prediction

The Pathway2Targets algorithm was then used to predict and prioritize relevant therapeutic targets from the transcriptomics signatures that could be repurposed in the case of Letrozole resistance, similar to prior work [34–37]. Briefly, this algorithm identifies known drug targets within each of the significant signaling pathways. It then uses multiple criteria from the OpenTargets database [38], as well as signaling pathways, to rank the relevant therapeutics and their targets according to a customizable weighting scheme. The justification for this approach is that a gene product need not be differentially expressed to be a relevant therapeutic target that reduces signs and/or symptoms of a given condition—particularly if the target is upstream of the significant DEG.

Separately, the Salmon read counts for each sample were combined into a tabular format and labeled as “resistant” or “sensitive” to treatment with Letrozole. This table was then used as input to the XGboost algorithm [39], which uses a tree-based method to train a model from 80% of the dataset and then quantifies its performance using the remaining 20% of the data [40,41], which minimizes model overfit. For the initial analysis, the gain metric was calculated from the read counts for all detected genes across all samples. The results were sorted in descending order such that the genes/features with the highest gain value were listed first. Given the gain metrics from the whole transcriptome, the number of genes/features being evaluated was reduced to the best two biomarkers from the original analysis since this is a number that is easily accommodated by qRT-PCR (or similar) molecular methods. This approach has been successfully applied previously with acceptable performance and accuracy [40,41,123]. The XGboost algorithm was selected since prior work has shown that tree-based classifiers are faster and more accurate than other machine learning-based methods such as support vector machine, neural networks, and Bayesian approaches [126]. The XGboost parameters that were changed from default included: constructing 10,000 parallel trees to ensure appropriate coverage of the available tree space, subsample of 0.5, using a binary:logistic objective, and an area under the curve (AUC) evaluation metric.

2.3. Protein Network Analysis

The public protein-protein interaction in the STRING database were used to construct a PPI network for all significant DEGs, the top five targets, and the top five transcriptional biomarkers [130]. Cytoscape was then used to visualize the network, while the 12 algorithms incorporated into the cytoHubba app were used to compute multiple node-based metrics from the Cytoscape network [131,132]. These algorithms were specifically selected to predict key nodes from a network while minimize bias due to either high-degree or low-degree proteins [132].

3. Results

3.1. Majority of Differentially Expressed Genes in ER+ Treatment Resistance are Upregulated

We began by retrieving, preprocessing and analyzing 58 ER+ breast cancer samples from patients who had been treated with Letrozole. The RNA sequencing data from these clinical samples are publicly available in the NCBI Gene Expression Omnibus (GEO) database. Our analysis detected a total of 18,735 genes between responders (treatment-sensitive patients) versus non-responders (treatment-resistant patients), with 105 of those genes having significant differential expression (FDR p-value < 0.05; Table 1; Supplementary Table 1). This significance cutoff was used since the FDR method applies a multiple hypothesis correction to reduce false positives in the results when so many statistical tests are applied to the dataset.

We observed that among these 105 Differentially Expressed Genes (DEGs) (Figure 1), the top 20 (Table 2) included 17 protein-coding genes that consisted of two S100 Calcium binding proteins, various transcription factors and matrix metalloproteinase 7 (Figure 1, Table 2). Three of the top 20 DEGs coded for immunoglobulin variable regions, while one was a putatively transcribed unprocessed pseudogene, Ovostatin II. All but one of these 20 DEGs were upregulated in the group that failed to respond to Letrozole treatment. The sole exception to this was the downregulation of the gene that encodes the growth arrest and DNA-damage inducible gene 45 gamma expression (GADD45G), which is involved in the regulation of cell growth and apoptosis [42], and showed a log2 fold-change (log2FC) value of -1.62.

The original study identified *PRR11* as one of the few statistically significant differentially expressed genes among the 51 genes in chromosome arm 17q23 that were included in their analysis [20]. However, in our analysis, we observed *PRR11* to be the 1,211th differentially expressed gene across the whole transcriptome, which did not surpass the statistical (FDR-adjusted p-value < 0.05) threshold for significance.

Table 1. Metrics for Original 58 ER+ Breast Cancer Samples Treated with Letrozole.

Metrics	Study Information
Gene Expression Omnibus (GEO) Study Identifier	GSE145325
Title	RNA sequencing of ER+ breast tumor treated with Letrozole
Platform	Illumina HiSeq 3000
Diagnostic Criteria	Resistance to estrogen suppression was defined by a preoperative endocrine prognostic index (PEPI) ≥ 4 and/or evidence of cancer relapse after a median follow-up of 5 years
Number of Responders vs Non-Responders	36 vs 22

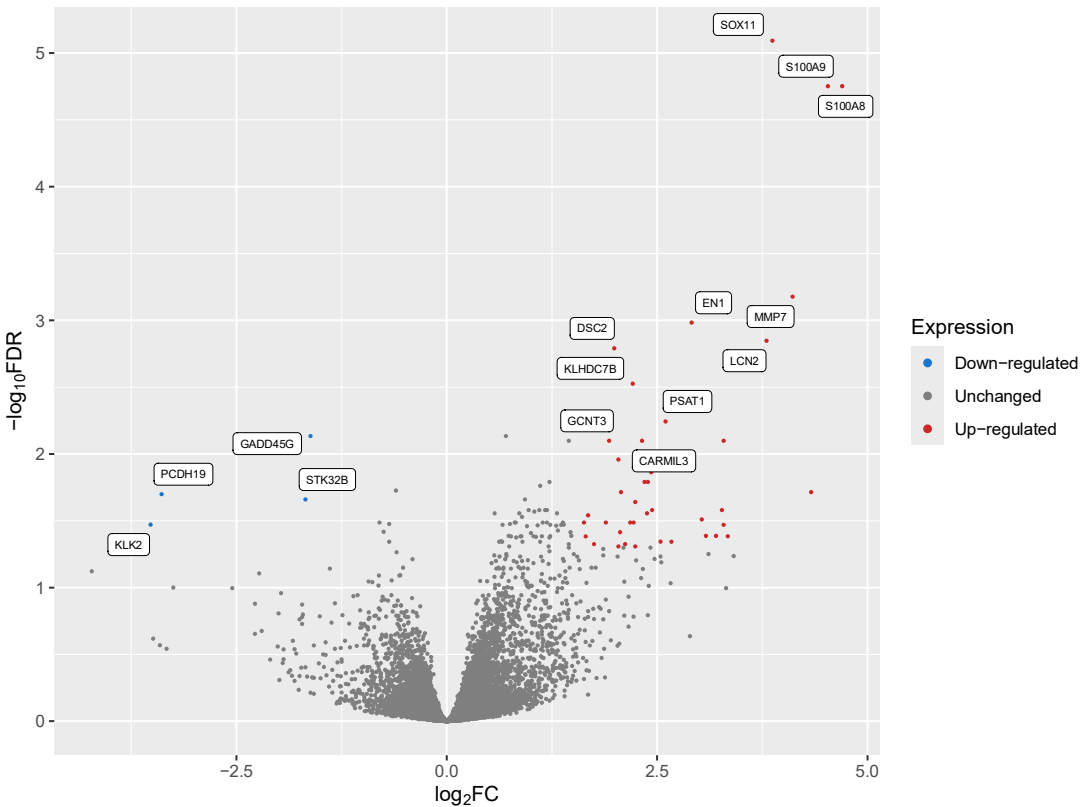


Figure 1. Volcano Plot of All Differentially Expressed Genes (DEGs). Each dot represents a gene product, with blue dots representing those that were significantly downregulated ($\log_2FC < 1$; FDR-corrected p-values < 0.05), and red representing DEGs that were significantly upregulated ($\log_2FC > 1.5$; FDR-corrected p-value < 0.05 , equivalent to $-\log_{10}FDR$ of 1.3). The 15 most statistically significant DEGs (11 up-regulated and four down-regulated) are labeled.

Table 2. Metrics and Information for Top Five Statistically Significant Differentially Expressed Genes.

Gene Symbol	Log2 Fold-Change	Log Counts Per Million (CPM)	FDR P-Value	Gene Name
SOX11	3.87	3.74	8.07E-06	SRY-box transcription factor 11
S100A9	4.53	5.06	1.77E-05	S100 calcium binding protein A9
S100A8	4.7	2.96	1.77E-05	S100 calcium binding protein A8
IGLV3-25	5.15	6.27	0.000184	immunoglobulin lambda variable 3-25
MMP7	4.11	4.84	0.000665	matrix metalloproteinase 7

3.2. Signaling Pathway Impact Analysis Identifies Four Significantly Affected Pathways

We then wanted to determine which intracellular signaling pathways were significantly enriched with the detected DEGs using the Signaling Pathway Impact Analysis (SPIA) algorithm (Bonferroni-adjusted p-value < 0.05). Briefly, this robust pathway enrichment algorithm uses a bootstrap-based approach to generate a null distribution for each pathway and to calculate a p-value. The results from our pathway enrichment analysis included four pathways that were predicted to differ between responders vs. non-responders to Letrozole Treatment (Table 3). These pathways included those associated with activated “PLK1 signaling events, anti-tumoral activity” as well as the activated “FOXM1 proliferation associated transcription factor network”.

Table 3. Metrics for All Significantly Affected Intracellular Signaling Pathways and Their Predicted Modulation.

Rank	Pathway Name	Total # of Pathway Members	# DEGs in Pathway	Bonferroni P-Value	Predicted Modulation
1	PLK1 signaling events, anti-tumoral activity	44	23	5.84E-10	Activated
2	Syndecan-1-mediated signaling events, leukocyte adhesion	30	8	0.00018429	Activated
3	FOXM1 proliferation associated transcription factor network	36	14	0.0105863	Activated
4	HIF-1-alpha transcription factor network, oxygen homeostasis	60	17	0.01923722	Inhibited

3.3. Targets Prioritized for Repurposing from Identified Pathways

We then wanted to use the pathway results to predict any existing therapeutic targets that could be repurposed to treat the resistance phenotype with small molecules, monoclonal antibodies, peptides, or other modalities. We consequently applied the Pathway2Targets algorithm to perform this target prioritization analysis, which predicted 60 therapeutic targets (Table 4; Supplementary Table 2). Briefly, this algorithm calculates a weighted score for each target based on the sum of values assigned to 26 attributes. Notably, the predicted targets for the non-response (resistance) phenotype included Vascular endothelial growth factor A (VEGFA), a current target for solid tumors; as well as Estrogen Receptor 1 (ESR1); Nitric Oxide Synthase 2 (NOS2); and various Matrix Metalloproteinases (MMP9 and MMP2).

Table 4. Ranked List of Top Five Repurposed Therapeutic Targets Predicted to Aid with Resistance Phenotype.

Target Symbol	Target Name	Weighted Score (Higher is Better)
VEGFA	Vascular endothelial growth factor A	2220.5

ESR1	Estrogen Receptor 1	1771
MMP9	Matrix Metalloproteinase 9	1699
FGFR3	Fibroblast Growth Factor Receptor 9	1589
AKT1	AKT serine/threonine kinase 1	1571

3.4. Machine Learning Predicts Two Robust Transcriptional Biomarkers

We next wanted to predict a subset of transcriptional biomarkers that could potentially be used as prognostics to better identify patients who do not respond to treatment with Letrozole. To perform this analysis, we applied a decision tree-based machine learning approach to predict features (i.e. expressed genes) that most accurately classify Letrozole treatment responders vs non-responders. We used the read counts for all samples as the input for this analysis to identify and rank 278 transcriptional biomarkers by the gain metric (Table 5). Performance metrics for all 278 transcriptional biomarkers yielded an area under the receiver-operator characteristic (AUROC) curve of 0.972 (97.2%; Figure 2), indicating an exceptional ability to classify Letrozole responders vs non-responders based on global gene expression. Sensitivity for this complete set of transcriptional biomarkers was predicted to be 100%, with a predicted specificity of 94%.

Given the resources required to generate whole transcriptome data using RNA-sequencing, we repeated our biomarker analysis to predict the transcriptional biomarkers that are most associated with treatment nonresponse in this patient population. To do so, we used the top two biomarkers from our first analysis as input for a subsequent analysis. This more focused second analysis identified PRDX4 and E2F8 as potential diagnostic biomarkers, with an AUROC, for only these two gene products of 0.854 and an overall accuracy of 88.2% (Figure 3).

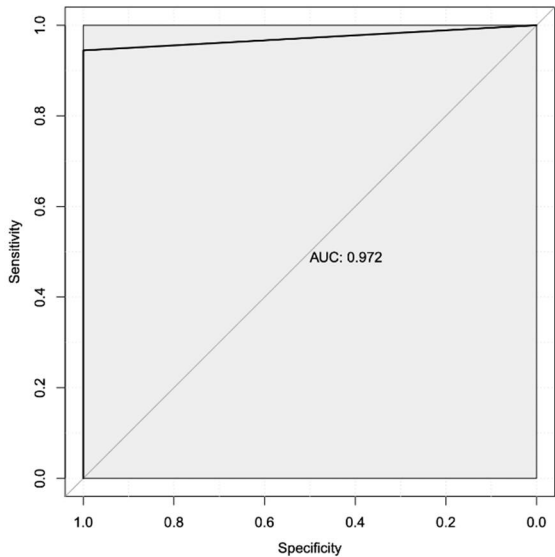


Figure 2. Area Under the Receiver-Operator Characteristic (AUROC) Curve Shows High Predicted Sensitivity vs Specificity Using All Transcripts. An initial machine learning analysis using all detected gene products as input showed high overall specificity (x-axis) and sensitivity (y-axis) for detecting Letrozole resistance in patients, with an AUC value of 97.2%.

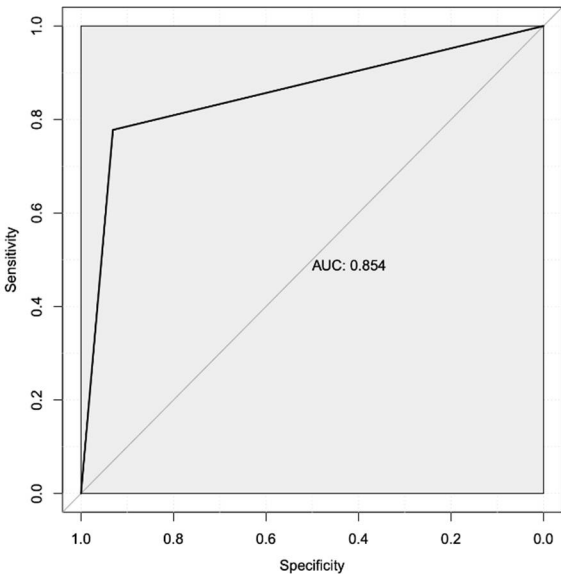


Figure 3. Area Under the Receiver-Operator Characteristic Curve Shows Acceptable Performance for the Top Two Transcripts. A secondary computational analysis using the highest ranked biomarkers from the whole transcriptome biomarker prediction (PRDX4 and E2F8). This analysis showed the combination of these two gene products have acceptable specificity (x-axis) and sensitivity (y-axis) for detecting Letrozole resistance in patients, with an AUC value of 85.4%.

Table 5. Description and Metrics for Top Five Predicted Transcriptional Biomarkers from all detected genes, Ranked in Decreasing Order by Gain.

Feature	Gain	Cover	Frequency	Antibody Available	Log2FC (FDR p-value)	Location
PRDX4	0.032	0.027	0.023	Monoclonal	0.807 (0.0326)	Secretory Granule; ER; Exosome
E2F8	0.031	0.025	0.021	Monoclonal	1.4 (0.266)	Cytosol; nucleus
IQGAP3	0.024	0.020	0.016	Monoclonal	1.2 (0.266)	Cytosol; plasma membrane
ATP6V1C2	0.021	0.018	0.015	Monoclonal	2.44 (0.0263)	Cytosol; lysosome; exosome
CDCA8	0.017	0.014	0.011	Monoclonal	1.37 (0.101)	Cytosol; nucleus

3.5. Protein-Protein Interactions Reveals Potential Treatment Resistance Network

We then constructed a protein-protein interaction (PPI) network to predict the protein(s) that potentially played a role in Letrozole resistance for these patients (Figure 4). To do so, we retrieved the known PPIs from the public STRING database for the 105 DEGs, as well as the top five drug targets and the top five transcriptional biomarkers. Interestingly, this analysis revealed a relatively well-connected local network, which indicates that the majority (87/115) of the input gene products

directly interacted with each other and at least partially supports our earlier predictions. We applied all 12 available algorithms within the CytoHubba app to this network in an effort to identify the key proteins with minimal bias to any individual approach. We then calculated the rank determined by each algorithm and then averaged them together. This analysis identified the top 10 “key” proteins in the network as CDK1, CCNB1, CCNA2, BIRC5, AKT1, ESR1, MMP9, CDCA8, CCND1, and PLK1. We found the ESR1, PLK1, MMP9, and AKT1 to be of particular interest since they were identified previously as the target of Letrozole, a significant signaling pathway, and two therapeutic targets (respectively). This profile at least partially represents the intracellular transcriptional response to Letrozole therapy.

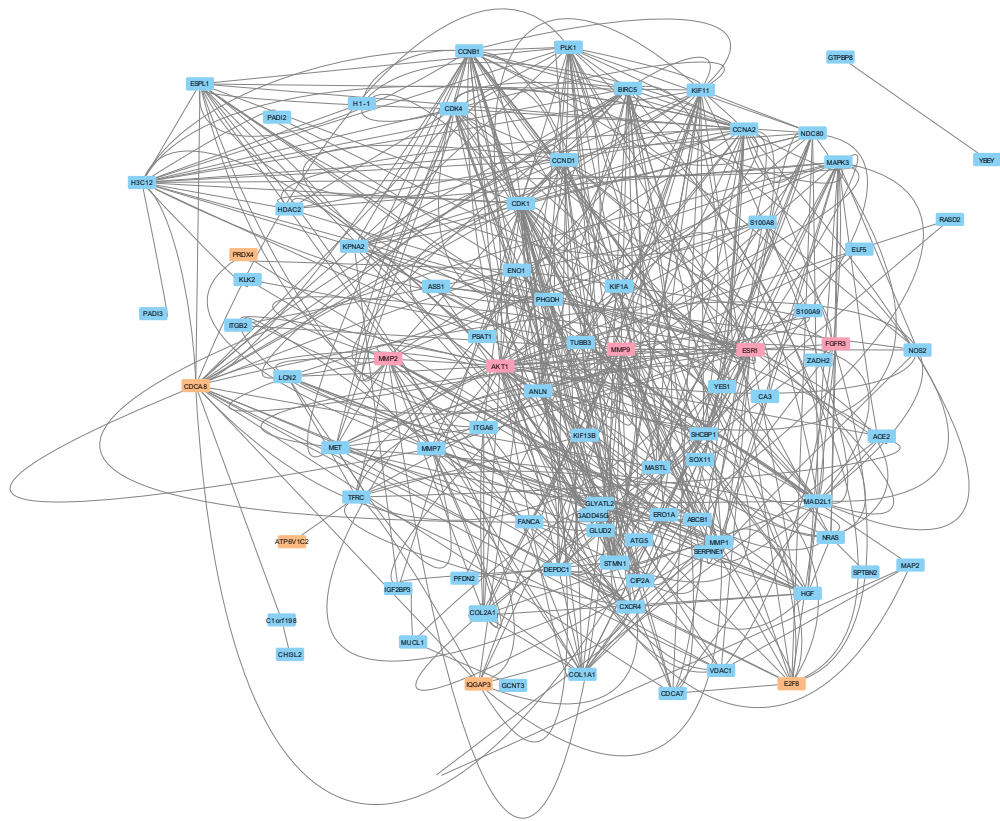


Figure 4. Protein-protein interaction network of DEGs, predicted targets, and top biomarkers. Graphical representation showing the interactions (edges) for each of the proteins (nodes) from the STRING-DB database. The 87 gene products that directly interact with each other were included in a single network. The input consisted of the 105 DEGs (blue-colored nodes), top five predicted therapeutic targets (salmon-colored nodes), and the top five predicted transcriptional biomarkers (orange-colored nodes).

4. Discussion

The primary aim of the current study was to perform a secondary analysis of a publicly available RNAseq dataset produced from 58 samples from patients with ER+ breast cancer following Letrozole treatment. Our analysis identified 105 statistically significant DEGs, with the top 20 gene products consisting of 17 protein-coding genes, including three encoding immunoglobulin variable regions, and one pseudogene. We also identified four significant signaling pathways, 20 therapeutic targets that could be repurposed to reduce Letrozole resistance and/or target resistance cells, and two potential transcriptional biomarkers that demonstrate high combined specificity and sensitivity in identifying Letrozole non-responders. Below we describe our primary findings in the context of other breast cancer-related studies.

4.1. Differentially Expressed Genes

In contrast to our findings, which identified 105 statistically significant DEGs, the original study that utilized the same data identified PRR11 as the only differentially expressed gene among the 51 genes in chromosome arm 17q23 [20]. We believe that the mostly likely contributor to this seeming discrepancy was the focus of the original study on the particular locus in 17q23 that contains 51 genes. In contrast, our analysis evaluated all of the detected mRNAs in the cells, which increased the scope and the statistical stringency needed to obtain significant results after incorporating multiple hypothesis correction. PRR11 underwent substantial differential gene expression in our analysis, although our thresholds did not categorize it as significant when looking across the whole transcriptome.

Transcription of the SRY-Box Transcription Factor 11 (SOX11) gene (with a log2FC value of 3.87) was our most statistically significant finding. Briefly, SOX11 is part of the Sox gene family comprising 20 transcription factors, which can further be divided into nine subgroups based on functional similarity [43]. SOX11 belongs to the SoxC group (composed of Sox4, Sox11, and Sox12), which are critical in the development of the nervous system and show substantial transactivation potential [44], while also being required for generation/proliferation of immature neuron progenitors [45]. SOX11 has recently been shown to behave as an oncogene and is a critical regulator of basal-like and Luminal B breast-cancers and their metastasis to the brain [46–48].

The upregulated S100 Calcium Binding Protein A8 and A9 (S100A8 and S100A9) genes encode a heterodimer of two calcium-binding proteins, referred to as calprotectin [49]. This protein is secreted by neutrophils and is an important proinflammatory mediator in both chronic and acute inflammation [49]. Calprotectin has recently been discovered to play an important role in tumorigenesis, cell proliferation, and resistance to traditional cancer therapies [50,51]. S100 proteins lack the signal peptide required for secretion via the classical endoplasmic reticulum / Golgi route, and its secretion mechanism is not well understood [50,52]. Recent discoveries implicate BRCA1, a known tumor suppressor, as essential for regulating the levels of S100A8 and S100A9 in the body [50,53]. These S100 gene products have also been shown to be secreted due to IL-22 and IL-17 signaling [54], and have been associated with poor outcomes in ductal carcinomas [55], and to be affected after AI treatment in an animal model [56]. While S100 proteins have been shown to be host markers of tumor development and progression, their role as modulators of Letrozole resistance has not been characterized previously. Thus, we believe that further experiments will help to elucidate the underlying mechanism(s) of S100A8 and S100A9 proteins in treatment-resistant ER+ breast cancer.

The Immunoglobulin Lambda Variable 3-25 (IGLV3-25) gene codes for light chain variable regions involved with antigen recognition [57,58] and has previously been implicated in various cancers [57,59,60]. Our analysis also found Matrix Metalloproteinase 7 (MMP7) to be upregulated in this dataset. This gene product encodes a proteolytic enzyme that secretes zinc and calcium endopeptidases, normally involved in wound healing and bone growth [61]. It has also been implicated in cancer progression, proliferation, differentiation, and/or apoptosis [62–64]. Basal-like breast cancer has upregulated MMP7 expression, which has been linked to DKK1 knockdown, a known tumor suppressor in breast cancer [65]. Interestingly, Zoledronate, a common drug to treat bone damage and osteoporosis [66–68], has been shown to decrease expression of MMP7 in breast cancer cell lines [69].

Engrailed Homeobox 1 (EN1) was another upregulated gene that we identified, which is a protein-coding gene that helps control development [70], and has been shown to affect bone metastases from breast cancer tumors [71]. It has been implicated in quadruple-negative breast cancer and basal-like breast cancer [72–74]. Our findings could potentially expand its scope of relevance to treatment-resistant ER+ breast cancer. We believe that future validation experiments should be performed to confirm these differential expression results.

4.2. Intracellular Signaling Pathway

Our analysis found four notable intracellular pathways that were significantly modified among non-responders. This analysis predicted the “PLK1 signaling events”, “Syndecan-1 mediated signaling events”, “FOXO1 proliferation transcription factor network” pathways to be activated; the “HIF-1-alpha transcription factor network” was predicted to be inhibited. Although the number of pathways is relatively small compared to other algorithms, we believe that applying a Bonferroni p-

value correction minimizes the number of false-results. It is possible that this approach could be complemented by other approaches in future studies.

PLK1 was our most significantly impacted intracellular signaling pathway (activated among non-responders vs. responders) and is primarily involved in cell mitosis. PLK1 contributes to many functions including controlling the G2/M checkpoint, coordinating the cell cycle and serving as a regulator of cell division in eukaryotes, as well as in both DNA replication and chromosome segregation [75,76]. In general, overexpression of PLK1 has been associated with various cancer types, including ER+ breast cancer [76,77]. PLK1 inhibition has been shown to exhibit anti-tumoral activity by blocking mitosis in cells with higher PLK1 expression, making it an attractive target for cancer therapy [78,79]. PLK1 knockdown in hormone-independent ER+ breast cancer has been associated with decreased cell viability and sensitization to radiation treatment [80]. These past studies serve to partially validate our significant pathway findings and lend credence to ongoing research into PLK1 inhibition as a potential treatment for patients with ER+ breast cancer.

The pathway for Syndecan-1 (SDC1, CD138) signaling was activated in non-responders. SDC1 is a transmembrane heparan sulfate proteoglycan responsible for maintaining typical cell morphology and is found abundantly in epithelial cells [81]. The SDC1 pathway is generally involved in cell proliferation, cell-matrix interactions, growth factor signaling, and angiogenesis [82,83]. SDC1 overexpression tends to correlate with more aggressive tumors, and it has an inverse correlation with ER expression [84]. This inverse correlation between ER expression and SDC1 signaling could help explain why Letrozole was ineffective for at least a subset of non-responders and merits further experimentation. Our analysis also showed MMP-7 as a highly significant DEG, which has been shown to anchor itself to syndecans and results in the shedding of important membrane-bound ligands such as EGF and TGF- α , leading to cell invasion [63]. Future follow-up experiments to validate these findings could include using CRISPR-i (or similar) systems and/or drugs to affect these pathways.

4.3. Target Prioritization and Repurposing

Our approach to predicting potential therapeutic targets by combining public data with analysis-specific pathway information and a customizable weighted scoring method is unique. While this approach has been useful in the past [32,33], other approaches exist that could complement our findings. Specifically, since the target data has more antagonists than agonists, this could potentially be a limitation. Vascular endothelial growth factor A (VEGFA) was the top potential target identified by our prioritization algorithm, which is already approved for certain indications including breast cancer. This growth factor belongs to a ligand family composed of six related proteins and plays an important role in breast cancer and other forms of cancer [85]. The primary role of this protein is to stimulate angiogenesis, which facilitates tumor growth [85,86]. Our computational prediction of VEGFA as a relevant target warrants future experimentation and serves as at least a partial validation of our approach.

Estrogen Receptor 1 (ESR1) encodes the estrogen receptor and was identified as a therapeutic target. Of note, ESR1 has been found to encode gain-of-function mutations that promote tumor metastasis and resistance to endocrine therapy [87–89], and can be detected with a liquid biopsy [88]. The detection of ESR1 as a therapeutic target is logical given that resistance to AI treatment enables the continued activation of signaling pathways that are modulated by ESR1.

Two additional highly ranked targets identified by our algorithm were Matrix Metalloproteinase 9 (MMP9) and Matrix Metalloproteinase 2 (MMP2), both members of the extracellular matrix metalloproteinase family (MMPs). MMPs assist in extracellular matrix remodeling through the activation of substrates via enzymatic cleavage [90]. They are involved in a variety of normal processes in the body, and their dysregulation plays a vital role in processes such as aging. Their dysregulation also plays a role in several abnormal conditions, including preeclampsia among pregnant women, and breast cancer [90–92]. Higher levels of MMP9 have been correlated with higher tumor grades and resistance to endocrine therapy [90]. MMP9 was further characterized as a vital component of the metastatic cascade early in tumorigenesis that promotes colonization in the lungs [93]. Other studies have implicated MMP2 with tumor metastasis to other organs, including the brain [94,95]. Overall, both MMP9 and MMP2 have been identified as promising markers for predicting the prognosis of patients with breast cancer [96,97]. Interestingly, a phase-III clinical trial targeting MMP

proteins with Marimastat sought to reduce tumor blood flow in patients with metastatic breast cancer [98]. Although MMP7 was identified as a significant DEG, it was not predicted as a useful target in this analysis. However, MMP7 is one of the several MMPs that are inhibited by doxycycline, which is a common treatment for various forms of cancer.

Fibroblast growth receptor 3 (FGFR3) belongs to a family of four highly conserved, transmembrane receptor tyrosine kinases. Once activated, they initiate intracellular cascades that carry out a variety of functions, including promoting cellular proliferation and survival [99]. This gene has been reported in other studies to be significantly upregulated in breast cancer [100]. FGFR3 is a known target of microRNA-593-3p, and overexpression of miR-593-3p down-regulates FGFR3 expression, which slows breast cancer progression [101]. Various inhibitors of FGFR3 are at various stages in phase I and phase II clinical trials for breast cancer.

Another potential target identified was AKT Serine/Threonine Kinase 1 (AKT1). AKT1, AKT2 and AKT3 are all isoforms of protein kinase B (AKT), an essential member of the PI3K/AKT signaling pathway [102]. Both AKT1 and AKT2 have been implicated in breast cancer previously, with some mixed results regarding their specific function. However, it is generally accepted that AKT1 plays an important role in the early stages of tumor development (tumor initiation & proliferation) while AKT2 primarily assists in tumor metastasis [102,103]. The original study for this dataset identified PI3K as an important player in AI treatment [20], which is supported in other studies [104]. Regardless of its specific function, AKT1 has a well-documented correlation with breast cancer and is an appropriate target for pharmaceutical intervention. Additional work at the discovery, pre-clinical, and clinical stages is needed to determine whether such treatments could reduce the effect of Letrozole-resistance in breast cancer patients.

4.4. Biomarker Analysis

Our DEG analysis identified expressed genes that significantly differed between the two groups of samples (resistant vs sensitive). In contrast, the aim of our biomarker analysis was to determine the expressed genes that most consistently differed between the two groups of samples. Although the input for both analyses was the read counts, we employed a tree-based machine learning model for the biomarker analysis that does not incorporate traditional statistics. Consequently, we did not expect a complete overlap between the DEGs and our predicted biomarkers.

Peroxisiredoxin-4 (PRDX4), which was an up-regulated gene among non-responders, was the top transcriptional biomarker identified in our analysis. This gene product belongs to a family of six small antioxidant isozymes and is generally localized in the endoplasmic reticulum [105]. PRDX4 plays an essential role in maintaining localized redox homeostasis and is upregulated when cells are under oxidative stress [106]. PRDX4 has been described as having a tumor-promoting effect that is well-documented among various forms of cancer, including lung and renal cancers, leukemia, and glioblastoma [107,108]. Prior work has shown overexpression of PRDX4 to be associated with poor overall survival rates, shorter relapse-free survival among breast cancer patients [109,110], and metastasis. PRDX4 has also been shown to be associated with more advanced breast cancer tumors [111], which could also contribute to treatment resistance. Thus, the role of PRDX4 as a potential transcriptional biomarker in treatment-resistant breast cancer deserves future exploration.

Early Region 2 Binding Factor (E2F8), an up-regulated DEG from our study, was another potential biomarker identified by our machine learning analysis. E2F8 is a member of the E2F family, which assists in assembling the core transcriptional machinery, making it crucial for cell division [112]. E2F8 has also been shown to promote angiogenesis, which helps establish the tumor microenvironment and is positively correlated with tumor malignancy [113,114]. E2F8 has been associated with CDK4/6 inhibitor resistance [115], is overexpressed in patients with breast cancer, and has been significantly correlated with poor patient survival and cancer progression [116,117]. E2F8 has been suggested as a marker of breast relapse-free survival and distant metastasis-free survival in breast cancer patients [118]. Interestingly, a prior study in HER2- has identified a different E2F Transcription Factor (E2F4) as having a role in treatment resistance in ER+ breast cancer, which at least partially supports our results [119]. However, the role of E2F8 as a biomarker specifically in treatment-resistant breast cancer warrants further exploration.

We acknowledge inherent deficiencies in training and applying a machine learning approach to a relatively small dataset. Unfortunately, the GEO database had other bulk RNA-sequencing datasets

that characterized HER2+ samples, there were no other public ER+ Letrozole-treated clinical samples [120]. Consequently, augmenting our model with additional data was not possible. Additional wet-lab validation experiments using qRT-PCR, or similar transcriptional assays will be required in one or more distinct patient populations to better quantify the accuracy and performance of our predicted biomarkers.

4.5. Potential Treatment Resistance Mechanisms

Unfortunately, the original dataset only included samples from responsive and non-responsive tumors. As such, it was not possible to directly and accurately analyze and assess the pre-treatment state of the tumor. Similarly, we believe it would not be justified to incorporate additional comparisons from other datasets that evaluated triple-negative or ER- breast cancers given the distinct patient populations, enrollment criteria, and other experimental variables. Even so, our protein network analysis did identify a set of gene products that may contribute to an increased understanding of the intracellular transcriptional response to Letrozole treatment.

We recognize that a subset of the genes we identify in our analysis of this dataset are opposite of what has been reported in prior work. Specifically, MMP9 and ESR1 as being upregulated in Letrozole-resistant breast cancer tumors, which agrees with our findings [121]. Another study found that Estrogen is known to attract and activate neutrophils, which consequently upregulates expression of both S100A8 and S100A9 in agreement with our findings [122]. Multiple immunoglobulin lambda genes, which were upregulated in our analysis, agrees with prior work in humans and mice that showed a similar response in resistant samples [123,124]. SOX11 has been reported to be downregulated due to epigenetic effects in some breast cancer tumors after treatment w/ Letrozole [47]; however, SOX11 was found to be upregulated in our analysis. The fact that SOX11 was included in our protein network suggests that changes in its expression could be driven by Letrozole treatment, though we cannot be confident about whether this upregulation is primarily taking place within tumor cells or the tumor-associated immune cells [122]. We anticipate that additional experiments in cell culture and/or clinical cohorts will be necessary to validate these findings.

4.6. Study Limitations

We purposefully selected our cutoff criteria to primarily allow the statistical significance value to determine the threshold. Our decision to not impose a fold-change cutoff enabled us to identify genes with relatively small, but statistically significant changes in gene expression. This approach at least partially accounts for the cellular heterogeneity in bulk RNA-seq experiments [127–129]. We anticipate that adjusting our fixed threshold approach to a different set of cutoff values could alter the results.

Although our use of robust and established large-scale statistical methods should improve confidence in our results, the possibility of false-positives is inevitable. Consequently, we anticipate follow-up experiments in the wet lab to further validate our results. Additional studies that compare our results to those from triple-negative and/or ER- breast cancer patients could also be beneficial in efforts to anticipate precision medicine efforts.

5. Conclusions

Our analysis identified genes, pathways, and transcriptional biomarkers that could contribute to improved understanding, prognostics, and characterization of mechanisms for Letrozole resistance in ER+ breast cancer patients. Future work is needed to better characterize the genes we identified that do not display a well-understood function to better understand how the tumor microenvironment interacts with Letrozole. Further analysis of our identified biomarkers could also yield clinical applications that enable treatment regimens specified to particular subsets of patients.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Table S1: Excel download for differentially expressed gene information. Table S2: Excel download for therapeutic targets.

Author Contributions: Conceptualization, B.E.P.; methodology, B.E.P.; software, B.E.P.; validation, L.S. and J. L.; formal analysis, L.S. and J. L.; investigation, L.S., J. L. and B.E.P.; data curation, L.S.; writing—original draft

preparation, L.S., J. L. and B.E.P; writing—review and editing, L.S., J. L., N.J.G. and B.E.P; visualization, L.S. and B.E.P.; supervision, B.E.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable.

Acknowledgments: We acknowledge the technical expertise and the high-performance computing environment provided by the BYU Office of Research Computing.

Conflicts of Interest: BEP owns equity in Pythia Biosciences.

Appendix A

The appendix is an optional section that can contain details and data supplemental to the main text—for example, explanations of experimental details that would disrupt the flow of the main text but nonetheless remain crucial to understanding and reproducing the research shown; figures of replicates for experiments of which representative data is shown in the main text can be added here if brief, or as Supplementary data. Mathematical proofs of results not central to the paper can be added as an appendix.

Appendix B

All appendix sections must be cited in the main text. In the appendices, Figures, Tables, etc. should be labeled starting with “A”—e.g., Figure A1, Figure A2, etc.

References

1. Giaquinto, A. N. *et al.* Breast Cancer Statistics, 2022. *CA Cancer J. Clin.* **72**, 524–541 (2022).
2. Postmenopause. *Cleveland Clinic* <https://my.clevelandclinic.org/health/diseases/21837-postmenopause>.
3. US Census Bureau. The Older Population: 2020. (2023).
4. Breast Cancer Statistics. <https://www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html>.
5. Rosenberg, P. S., Barker, K. A. & Anderson, W. F. Estrogen Receptor Status and the Future Burden of Invasive and In Situ Breast Cancers in the United States. *J. Natl. Cancer Inst.* **107**, (2015).
6. Breast cancer types: What your type means. *Mayo Clinic* <https://www.mayoclinic.org/diseases-conditions/breast-cancer/in-depth/breast-cancer/art-20045654> (2022).
7. Breast Biomarker Immunocytochemistry. in *The Breast* 197–206.e6 (Elsevier, 2018).
8. Siddiq, A., Long, L. M., Li, L., Marciniak, R. A. & Kazhdan, I. Expression of HER-2 in MCF-7 breast cancer cells modulates anti-apoptotic proteins Survivin and Bcl-2 via the extracellular signal-related kinase (ERK) and phosphoinositide-3 kinase (PI3K) signalling pathways. *BMC Cancer* **8**, 129 (2008).
9. Zambrano, J. & Yeh, E. S. Autophagy and Apoptotic Crosstalk: Mechanism of Therapeutic Resistance in HER2-Positive Breast Cancer. *Breast Cancer* **10**, 13–23 (2016).
10. Onitilo, A. A., Engel, J. M., Greenlee, R. T. & Mukesh, B. N. Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. *Clin. Med. Res.* **7**, 4–13 (2009).
11. Calhoun, B. C. & Collins, L. C. Predictive markers in breast cancer: An update on ER and HER2 testing and reporting. *Semin. Diagn. Pathol.* **32**, 362–369 (2015).
12. Caldarella, A. *et al.* Female breast cancer status according to ER, PR and HER2 expression: a population based analysis. *Pathol. Oncol. Res.* **17**, 753–758 (2011).
13. Bagaria, S. P. *et al.* Personalizing breast cancer staging by the inclusion of ER, PR, and HER2. *JAMA Surg.* **149**, 125–129 (2014).
14. Bhatnagar, A. S. The discovery and mechanism of action of letrozole. *Breast Cancer Res. Treat.* **105 Suppl 1**, 7–17 (2007).
15. Sohl, C. D. & Guengerich, F. P. Kinetic analysis of the three-step steroid aromatase reaction of human cytochrome P450 19A1. *J. Biol. Chem.* **285**, 17734–17743 (2010).
16. Letrozole: Pharmacology, toxicity and potential therapeutic effects. *Life Sci.* **310**, 121074 (2022).
17. Ellis, M. J. & Ma, C. Letrozole in the neoadjuvant setting the P024 trial. *Breast Cancer Res. Treat.* **112**, 371–371 (2008).

18. Guarneri, V. *et al.* Double-blind, placebo-controlled, multicenter, randomized, phase IIb neoadjuvant study of letrozole-lapatinib in postmenopausal hormone receptor-positive, human epidermal growth factor receptor 2-negative, operable breast cancer. *J. Clin. Oncol.* **32**, 1050–1057 (2014).
19. Objective response rate of placebo in randomized controlled trials of anticancer medicines. *eClinicalMedicine* **55**, 101753 (2023).
20. Lee, K.-M. *et al.* Proline rich 11 (PRR11) overexpression amplifies PI3K signaling and promotes antiestrogen resistance in breast cancer. *Nat. Commun.* **11**, 1–15 (2020).
21. Liu, Y. *et al.* Targeting 17q23 amplicon to overcome the resistance to anti-HER2 therapy in HER2+ breast cancer. *Nat. Commun.* **9**, 1–16 (2018).
22. Kodama, Y., Shumway, M., Leinonen, R. & International Nucleotide Sequence Database Collaboration. The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res.* **40**, D54–6 (2012).
23. Orjuela, S., Huang, R., Hembach, K. M., Robinson, M. D. & Soneson, C. ARMOR: An automated reproducible workflow for Preprocessing and Differential Analysis of NA-seq Data. *G3* **9**, 2089–2096 (2019).
24. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **34**, 3600 (2018).
25. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. www.bioinformatics.babraham.ac.uk/projects/fastqc/.
26. Babraham Bioinformatics - Trim Galore! https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
27. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
28. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
29. Wu, D. & Smyth, G. K. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res.* **40**, e133 (2012).
30. Tarca, A. L. *et al.* A novel signaling pathway impact analysis. *Bioinformatics* **25**, 75–82 (2009).
31. Gifford, K. T. L. & Pickett, B. E. Comparative meta-analysis of host transcriptional response during *Streptococcus pneumoniae* carriage or infection. *Microb. Pathog.* **173**, 105816 (2022).
32. Ferrarini, M. G. *et al.* Genome-wide bioinformatic analyses predict key host and viral factors in SARS-CoV-2 pathogenesis. *Commun Biol* **4**, 590 (2021).
33. Martinez Viedma, M. D. P. & Pickett, B. E. Characterizing the Different Effects of Zika Virus Infection in Placenta and Microglia Cells. *Viruses* **10**, (2018).
34. Dobbs Spendlove, M. *et al.* Pathway2Targets: an open-source pathway-based approach to repurpose therapeutic drugs and prioritize human targets. *PeerJ* **11**, e16088 (2023).
35. Scott, T. M. *et al.* Comparison of Intracellular Transcriptional Response of NHBE Cells to Infection with SARS-CoV-2 Washington and New York Strains. *Front. Cell. Infect. Microbiol.* **12**, 1009328 (2022).
36. Gray, M. *et al.* Chikungunya virus time course infection of human macrophages reveals intracellular signaling pathways relevant to repurposed therapeutics. *PeerJ* **10**, e13090 (2022).
37. Scott, T. M., Jensen, S. & Pickett, B. E. A signaling pathway-driven bioinformatics pipeline for predicting therapeutics against emerging infectious diseases. *F1000Res.* **10**, 330 (2021).
38. Ochoa, D. *et al.* Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).
39. GitHub - dmlc/xgboost: Scalable, Portable and Distributed Gradient Boosting (GBDT, GBRT or GBM) Library, for Python, R, Java, Scala, C++ and more. Runs on single machine, Hadoop, Spark, Dask, Flink and DataFlow. [GitHub https://github.com/dmlc/xgboost](https://github.com/dmlc/xgboost).
40. Rapier-Sharman, N., Clancy, J. & Pickett, B. E. Joint Secondary Transcriptomic Analysis of Non-Hodgkin's B-Cell Lymphomas Predicts Reliance on Pathways Associated with the Extracellular Matrix and Robust Diagnostic Biomarkers. *J Bioinform Syst Biol* **5**, 119–135 (2022).
41. Moreno, C., Bybee, E., Tellez Freitas, C. M., Pickett, B. E. & Weber, K. S. Meta-Analysis of Two Human RNA-seq Datasets to Determine Periodontitis Diagnostic Biomarkers and Drug Target Candidates. *Int. J. Mol. Sci.* **23**, (2022).
42. Ying, J. *et al.* The Stress-Responsive Gene GADD45G Is a Functional Tumor Suppressor, with Its Response to Environmental Stresses Frequently Disrupted Epigenetically in Multiple Tumors. *Clin. Cancer Res.* **11**, 6442–6449 (2005).
43. Understanding the function and regulation of Sox2 for its therapeutic potential in breast cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1877**, 188692 (2022).
44. She, Z.-Y. & Yang, W.-X. SOX family transcription factors involved in diverse cellular events during development. *Eur. J. Cell Biol.* **94**, 547–563 (2015).
45. Potzner, M. R. *et al.* Sequential requirement of Sox4 and Sox11 during development of the sympathetic nervous system. *Development* **137**, 775–784 (2010).
46. Shepherd, J. H. *et al.* The SOX11 transcription factor is a critical regulator of basal-like breast cancer growth, invasion, and basal-like gene expression. *Oncotarget* **7**, 13106–13121 (2016).

47. Wang, H. *et al.* Integrative analysis identifies two molecular and clinical subsets in Luminal B breast cancer. *iScience* **26**, 107466 (2023).
48. Zvelebil, M. *et al.* Embryonic mammary signature subsets are activated in Brca1-/- and basal-like breast cancers. *Breast Cancer Res.* **15**, R25 (2013).
49. S100A8 and S100A9 in inflammation and cancer. *Biochem. Pharmacol.* **72**, 1622–1631 (2006).
50. Li, J. *et al.* S100A9-CXCL12 activation in BRCA1-mutant breast cancer promotes an immunosuppressive microenvironment associated with resistance to immunotherapy. *Nat. Commun.* **13**, 1–19 (2022).
51. S100A8 and S100A9 in Cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1878**, 188891 (2023).
52. Myeloid-related Protein (MRP) 8 and MRP14, Calcium-binding Proteins of the S100 Family, Are Secreted by Activated Monocytes via a Novel, Tubulin-dependent Pathway. *J. Biol. Chem.* **272**, 9496–9502 (1997).
53. Wang, L. & Di, L.-J. BRCA1 and estrogen/estrogen receptor in breast cancer: where they interact? *Int. J. Biol. Sci.* **10**, 566–575 (2014).
54. Karpishev, V. *et al.* The role of Th17 cells in the pathogenesis and treatment of breast cancer. *Cancer Cell Int.* **22**, 108 (2022).
55. Arai, K. *et al.* S100A8 and S100A9 overexpression is associated with poor pathological parameters in invasive ductal carcinoma of the breast. *Curr. Cancer Drug Targets* **8**, 243–252 (2008).
56. Boutas, I. *et al.* Effect of Aromatase Inhibitors on Serum Calprotectin Levels in an Animal Experimental Model: Trial. *Cancer Diagn Progn* **2**, 720–730 (2022).
57. Ralli, S., Jones, S. J., Leach, S., Lynch, H. T. & Brooks-Wilson, A. R. Gene and pathway based burden analyses in familial lymphoid cancer cases: Rare variants in immune pathway genes. *PLoS One* **18**, e0287602 (2023).
58. Development and validation of a 6-gene signature for the prognosis of loco-regional control in patients with HPV-negative locally advanced HNSCC treated by postoperative radio(chemo)therapy. *Radiother. Oncol.* **171**, 91–100 (2022).
59. Dai, D., Xie, L., Shui, Y., Li, J. & Wei, Q. Identification of Tumor Microenvironment-Related Prognostic Genes in Sarcoma. *Front. Genet.* **12**, 620705 (2021).
60. Li, R. *et al.* Identification and validation of the prognostic value of immune-related genes in non-small cell lung cancer. *Am. J. Transl. Res.* **12**, 5844–5865 (2020).
61. Roles of matrix metalloproteinase-7 (MMP-7) in cancer. *Clin. Biochem.* **92**, 9–18 (2021).
62. Köhrmann, A., Kammerer, U., Kapp, M., Dietl, J. & Anacker, J. Expression of matrix metalloproteinases (MMPs) in primary human breast cancer and breast cancer cell lines: New findings and review of the literature. *BMC Cancer* **9**, 188 (2009).
63. Bouris, P. *et al.* Estrogen receptor alpha mediates epithelial to mesenchymal transition, expression of specific matrix effectors and functional properties of breast cancer cells. *Matrix Biol.* **43**, 42–60 (2015).
64. Gialeli, C., Theocharis, A. D. & Karamanos, N. K. Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting. *FEBS J.* **278**, 16–27 (2011).
65. The Forkhead Box Transcription Factor FOXC1 Promotes Breast Cancer Invasion by Inducing Matrix Metalloprotease 7 (MMP7) Expression. *J. Biol. Chem.* **287**, 24631–24640 (2012).
66. Wang, B., Zhan, Y., Yan, L. & Hao, D. How zoledronic acid improves osteoporosis by acting on osteoclasts. *Front. Pharmacol.* **13**, 961941 (2022).
67. Reid, I. R. *et al.* Zoledronate. *Bone* **137**, 115390 (2020).
68. Black, D. M. *et al.* Once-yearly zoledronic acid for treatment of postmenopausal osteoporosis. *N. Engl. J. Med.* **356**, 1809–1822 (2007).
69. Dedes, P. G. *et al.* Expression of matrix macromolecules and functional properties of breast cancer cells are modulated by the bisphosphonate zoledronic acid. *Biochim. Biophys. Acta* **1820**, 1926–1939 (2012).
70. Chang, J. *et al.* EN1 Regulates Cell Growth and Proliferation in Human Glioma Cells via Hedgehog Signaling. *Int. J. Mol. Sci.* **23**, 1123 (2022).
71. Majumder, A., Singh, M. & Tyagi, S. C. Post-menopausal breast cancer: from estrogen to androgen receptor. *Oncotarget* **8**, 102739–102758 (2017).
72. Kim, Y. J. *et al.* Engrailed 1 overexpression as a potential prognostic marker in quintuple-negative breast cancer. *Cancer Biol. Ther.* (2018) doi:10.1080/15384047.2018.1423913.
73. Beltran, A. S., Graves, L. M. & Blancafort, P. Novel role of Engrailed 1 as a prosurvival transcription factor in basal-like breast cancer and engineering of interference peptides block its oncogenic function. *Oncogene* **33**, 4767–4777 (2013).
74. Yang, L. *et al.* Weighted gene co-expression network analysis of the association between upregulated AMD1, EN1 and VGLL1 and the progression and poor prognosis of breast cancer. *Exp. Ther. Med.* **22**, 1–13 (2021).
75. PLK1, A Potential Target for Cancer Therapy. *Transl. Oncol.* **10**, 22–32 (2017).
76. Chiappa, M. *et al.* Present and Future Perspective on PLK1 Inhibition in Cancer Treatment. *Front. Oncol.* **12**, 903016 (2022).
77. PLK1 Signaling in Breast Cancer Cells Cooperates with Estrogen Receptor-Dependent Gene Transcription. *Cell Rep.* **3**, 2021–2032 (2013).

78. Kahl, I. *et al.* The cell cycle-related genes RHAMM, AURKA, TPX2, PLK1, and PLK4 are associated with the poor prognosis of breast cancer patients. *J. Cell. Biochem.* **123**, 581–600 (2022).
79. Gutteridge, R. E. A., Ndiaye, M. A., Liu, X. & Ahmad, N. Plk1 Inhibitors in Cancer Therapy: From Laboratory to Clinics. *Mol. Cancer Ther.* **15**, 1427–1435 (2016).
80. Bhola, N. E. *et al.* Kinome-wide Functional Screen Identifies Role of PLK1 in Hormone-Independent, ER-Positive Breast Cancer. *Cancer Res.* **75**, 405–414 (2015).
81. Couchman, J. R. Syndecan-1 (CD138), Carcinomas and EMT. *Int. J. Mol. Sci.* **22**, 4227 (2021).
82. Yang, Z., Chen, S., Ying, H. & Yao, W. Targeting syndecan-1: new opportunities in cancer therapy. *American Journal of Physiology-Cell Physiology* (2022) doi:10.1152/ajpcell.00024.2022.
83. Fleurot, E., Goudin, C., Hanoux, V., Bonnamy, P.-J. & Levallet, J. Estrogen receptor α regulates the expression of syndecan-1 in human breast carcinoma cells. *Endocr. Relat. Cancer* **26**, 615–628 (2019).
84. Barbareschi, M. *et al.* High syndecan-1 expression in breast carcinoma is related to an aggressive phenotype and to poorer prognosis. *Cancer* **98**, 474–483 (2003).
85. Al Kawas, H. *et al.* How VEGF-A and its splice variants affect breast cancer development – clinical implications. *Cell. Oncol.* **45**, 227–239 (2022).
86. Claesson-Welsh, L. & Welsh, M. VEGFA and tumour angiogenesis. *J. Intern. Med.* **273**, 114–127 (2013).
87. Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M. & Schiff, R. ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. *Nat. Rev. Clin. Oncol.* **12**, 573–583 (2015).
88. Dustin, D., Gu, G. & Fuqua, S. A. W. ESR1 mutations in breast cancer. *Cancer* **125**, 3714–3728 (2019).
89. Brett, J. O., Spring, L. M., Bardia, A. & Wander, S. A. ESR1 mutation as an emerging clinical biomarker in metastatic hormone receptor-positive breast cancer. *Breast Cancer Res.* **23**, 1–15 (2021).
90. Joseph, C. *et al.* Elevated MMP9 expression in breast cancer is a predictor of shorter patient survival. *Breast Cancer Res. Treat.* **182**, 267–282 (2020).
91. Cancemi, P. *et al.* The Role of Matrix Metalloproteinases (MMP-2 and MMP-9) in Ageing and Longevity: Focus on Sicilian Long-Living Individuals (LLIs). *Mediators Inflamm.* **2020**, 8635158 (2020).
92. Nikolov, A. & Popovski, N. Role of Gelatinases MMP-2 and MMP-9 in Healthy and Complicated Pregnancy and Their Future Potential as Preeclampsia Biomarkers. *Diagnostics (Basel)* **11**, (2021).
93. Owyong, M. *et al.* MMP9 modulates the metastatic cascade and immune landscape for breast cancer anti-metastatic therapy. *Life Sci Alliance* **2**, (2019).
94. Mendes, O., Kim, H.-T., Lungu, G. & Stoica, G. MMP2 role in breast cancer brain metastasis development and its regulation by TIMP2 and ERK1/2. *Clin. Exp. Metastasis* **24**, 341–351 (2007).
95. Mendes, O., Kim, H.-T. & Stoica, G. Expression of MMP2, MMP9 and MMP3 in Breast Cancer Brain Metastasis in a Rat Model. *Clin. Exp. Metastasis* **22**, 237–246 (2005).
96. Jiang, H. & Li, H. Prognostic values of tumoral MMP2 and MMP9 overexpression in breast cancer: a systematic review and meta-analysis. *BMC Cancer* **21**, 1–13 (2021).
97. Somiari, S. B. *et al.* Circulating MMP2 and MMP9 in breast cancer—Potential role in classification of patients into low risk, high risk, benign disease and breast cancer categories. *International Journal of Cancer* **119**, 1403–1411 (2006).
98. ClinicalTrials.gov. <https://www.clinicaltrials.gov/study/NCT00003010>.
99. Chew, N. J. *et al.* FGFR3 signaling and function in triple negative breast cancer. *Cell Commun. Signal.* **18**, 1–17 (2020).
100. Long, X. *et al.* MicroRNA-99a Suppresses Breast Cancer Progression by Targeting FGFR3. *Front. Oncol.* **9**, 499482 (2020).
101. Xie, J., Wan, Y., Zhang, M., Jin, Z. & Yao, Y. Circ_0061825 Acts as a miR-593-3p Sponge to Promote Breast Cancer Progression by Regulating FGFR3 Expression. *CMAR* **12**, 11243–11255 (2020).
102. Hinz, N. & Jücker, M. Distinct functions of AKT isoforms in breast cancer: a comprehensive review. *Cell Commun. Signal.* **17**, 1–29 (2019).
103. Riggio, M. *et al.* AKT1 and AKT2 isoforms play distinct roles during breast cancer progression through the regulation of specific downstream proteins. *Sci. Rep.* **7**, 1–12 (2017).
104. Sanchez, C. G. *et al.* Preclinical modeling of combined phosphatidylinositol-3-kinase inhibition with endocrine therapy for estrogen receptor-positive breast cancer. *Breast Cancer Res.* **13**, R21 (2011).
105. Jia, W., Chen, P. & Cheng, Y. PRDX4 and Its Roles in Various Cancers. *Technol. Cancer Res. Treat.* (2019) doi:10.1177/1533033819864313.
106. Elko, E. A. *et al.* Oxidation of peroxiredoxin-4 induces oligomerization and promotes interaction with proteins governing protein folding and endoplasmic reticulum stress. *J. Biol. Chem.* **296**, (2021).
107. Roles of peroxiredoxins in cancer, neurodegenerative diseases and inflammatory diseases. *Pharmacol. Ther.* **163**, 1–23 (2016).
108. Kocatürk, B. In silico analysis reveals PRDX4 as a prognostic and oncogenic marker in renal papillary cell carcinoma. *Gene* **859**, 147201 (2023).
109. Mei, J. *et al.* Comprehensive analysis of peroxiredoxins expression profiles and prognostic values in breast cancer. *Biomarker Research* **7**, 1–11 (2019).

110. Peroxiredoxin 4: A novel secreted mediator of cancer induced osteoclastogenesis. *Cancer Lett.* **361**, 262–270 (2015).
111. Wang, G. *et al.* The Prognosis Of Peroxiredoxin Family In Breast Cancer. *Cancer Manag. Res.* **11**, 9685–9699 (2019).
112. Emerging role of E2F8 in human cancer. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1869**, 166745 (2023).
113. Jiang, X. *et al.* The role of microenvironment in tumor angiogenesis. *J. Exp. Clin. Cancer Res.* **39**, 1–19 (2020).
114. Website. <https://doi.org/10.1038/emboj.2012.231> doi:10.1038/emboj.2012.231.
115. Whittle, J. R. *et al.* Dual Targeting of CDK4/6 and BCL2 Pathways Augments Tumor Response in Estrogen Receptor–Positive Breast Cancer. *Clin. Cancer Res.* **26**, 4120–4134 (2020).
116. Ye, L. *et al.* Upregulation of E2F8 promotes cell proliferation and tumorigenicity in breast cancer by modulating G1/S phase transition. *Oncotarget* **7**, 23757–23771 (2016).
117. Li, Y. *et al.* Expression patterns of E2F transcription factors and their potential prognostic roles in breast cancer. *Oncol. Lett.* **15**, 9216–9230 (2018).
118. Oshi, M. *et al.* The E2F Pathway Score as a Predictive Biomarker of Response to Neoadjuvant Therapy in ER+/HER2- Breast Cancer. *Cells* **9**, (2020).
119. Hoogstraat, M. *et al.* Comprehensive characterization of pre- and post-treatment samples of breast cancer reveal potential mechanisms of chemotherapy resistance. *NPJ Breast Cancer* **8**, 60 (2022).
120. Guerrero-Zotano, A. L. *et al.* ER Breast Cancers Resistant to Prolonged Neoadjuvant Letrozole Exhibit an E2F4 Transcriptional Program Sensitive to CDK4/6 Inhibitors. *Clin. Cancer Res.* **24**, 2517–2529 (2018).
121. Selli, C. *et al.* Molecular changes during extended neoadjuvant letrozole treatment of breast cancer: distinguishing acquired resistance from dormant tumours. *Breast Cancer Res.* **21**, 2 (2019).
122. Huang, H. *et al.* The immunomodulatory effects of endocrine therapy in breast cancer. *J. Exp. Clin. Cancer Res.* **40**, 19 (2021).
123. Pires, B. R. B. *et al.* Label-Free Proteomics Revealed Oxidative Stress and Inflammation as Factors That Enhance Chemoresistance in Luminal Breast Cancer. *Oxid. Med. Cell. Longev.* **2019**, 5357649 (2019).
124. Dabydeen, S. A. *et al.* Comparison of tamoxifen and letrozole response in mammary preneoplasia of ER and aromatase overexpressing mice defines an immune-associated gene signature linked to tamoxifen resistance. *Carcinogenesis* **36**, 122–132 (2015).
125. Clancy, J., Hoffmann, C. S. & Pickett, B. E. Transcriptomics secondary analysis of severe human infection with SARS-CoV-2 identifies gene expression changes and predicts three transcriptional biomarkers in leukocytes. *Comput. Struct. Biotechnol. J.* **21**, 1403–1413 (2023).
126. Piccolo, S. R., Mecham, A., Golightly, N. P., Johnson, J. L. & Miller, D. B. The ability to classify patients based on gene-expression data varies by algorithm and performance metric. *PLoS Comput. Biol.* **18**, e1009926 (2022).
127. Vinkel, J., Rib, L., Buil, A., Hedetoft, M. & Hyldegaard, O. Key pathways and genes that are altered during treatment with hyperbaric oxygen in patients with sepsis due to necrotizing soft tissue infection (HBOmic study). *Eur. J. Med. Res.* **28**, 507 (2023).
128. Li, F. *et al.* Weakly activated core neuroinflammation pathways were identified as a central signaling mechanism contributing to the chronic neurodegeneration in Alzheimer's disease. *Front. Aging Neurosci.* **14**, 935279 (2022).
129. Roche, K. E. & Mukherjee, S. The accuracy of absolute differential abundance analysis from relative count data. *PLoS Comput. Biol.* **18**, e1010284 (2022).
130. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
131. Su, G., Morris, J. H., Demchak, B. & Bader, G. D. Biological network exploration with Cytoscape 3. *Curr. Protoc. Bioinformatics* **47**, 8.13.1–24 (2014).
132. Chin, C.-H. *et al.* cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8** Suppl 4, S11 (2014).

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.