Gümüşhane University Journal of Science

GUFBD / *GUJS* (2025) 15(1): 260-273 doi: 10.17714/gumusfenbil.1624557 Research Article

Actionable targets in breast cancer: a multi-omics approach to uncover tumor-specific vulnerabilities

Meme kanserinde uygulanabilir hedefler: tümör-spesifik hassasiyetleri ortaya çıkarmak için multi-omik bir yaklaşım

Abdulkadir ELMAS*[1,2]

[1]*Giresun University, Faculty of Engineering, Dept. Of Electrical & Electronics Engineering, Giresun 28200, Türkiye*
[2]*Icahn School of Medicine at Mount Sinai, Department of Genetics and Genomic Sciences, New York, NY 10029, USA*

• Received: 21.01.2025 • Accepted: 06.03.2025

Abstract

Breast cancer remains a leading cause of cancer-related deaths worldwide, necessitating innovative therapeutic strategies. This study integrates proteomic, transcriptomic, and functional dependency data to systematically identify tumor-specific markers and vulnerabilities in breast cancer. We analyzed mass-spectrometry-based proteomic data from 115 tumor samples and 18 matched normal tissues, quantifying 10,468 proteins. By combining overexpression and differential expression analyses, we identified 172 tumor-specific proteins, including well-characterized targets such as ERBB2, EGFR, and CCND1, as well as novel candidates like TRPS1, UBE2C, and FOXP4. Functional validation of these candidate targets was performed through CRISPR-based expression-driven dependency analysis using the BEACON method and DepMap data, which revealed both gene- and protein-level dependencies, uncovering novel protein-unique cancer vulnerabilities. Notably, protein-specific dependencies such as UBE2C and E2F3 highlight potential therapeutic targets overlooked in transcriptomic analyses. In particular, markers such as TRPS1 and UBE2C, which exhibit strong protein expression-driven dependencies, may serve as potential candidates for precision oncology approaches, guiding drug development and patient stratification. This study presents a systematically prioritized set of actionable targets, emphasizing the critical role of multi-omics integration in driving precision oncology advancements for breast cancer.

Keywords: BEACON method, Breast cancer, Expression-driven dependency, Multi-omics approach, Precision oncology

Öz

Meme kanseri, dünya genelinde kanser kaynaklı ölümlerin başlıca nedenlerinden biri olup, önlenimi/tedavisi için yenilikçi terapötik stratejilerine gereksinim duyulmaktadır. Bu çalışma, meme kanserinde tümör-spesifik biyobelirteçleri ve zayıf noktaları sistematik olarak belirlemek için proteomik, transkriptomik ve fonksiyonel bağımlılık verilerini birleştirmektedir. 115 tümör örneği ve 18 eşlenmiş normal dokudan elde edilen kütle spektrometrisine dayalı proteomik veriler analiz edilmiş ve 10,468 adet protein incelenmiştir. Aşırı ekspresyon ve diferansiyel ekspresyon analizlerinin birleştirilmesiyle, ERBB2, EGFR ve CCND1 gibi iyi bilinen hedeflerin yanı sıra TRPS1, UBE2C ve FOXP4 gibi yeni adayların da bulunduğu 172 tümör-spesifik protein tespit edilmiştir. Bu aday hedeflerin fonksiyonel doğrulaması, BEACON yöntemi ve DepMap verileri kullanılarak CRISPR tabanlı ekspresyon odaklı bağımlılık analizi ile gerçekleştirilmiş olup, hem gen hem de protein düzeyinde bağımlılıkları ortaya koyarak proteine özgü yeni kanser hassasiyetlerini belirlemiştir. Özellikle UBE2C ve E2F3 gibi protein-spesifik bağımlılıklar, transkriptomik analizlerde gözden kaçan potansiyel terapötik hedefleri vurgulamaktadır. Özellikle, ekspresyon odaklı bağımlılık gösteren TRPS1 ve UBE2C gibi belirteçler, ilaç geliştirme ve hasta sınıflandırmasını yönlendiren hassas onkoloji yaklaşımları için potansiyel adaylar olarak hizmet edebilir. Bu çalışma, uygulanabilir hedefleri sistematik olarak önceliklendirerek, multi-omik entegrasyonun meme kanserinde hassas onkolojiyi ilerletmedeki kritik rolünü vurgulamaktadır.

Anahtar kelimeler: *BEACON yöntemi, Meme kanseri, Ekspresyon temelli bağımlılık, Çoklu-omik yaklaşım, Hassas onkoloji*

*Abdulkadir ELMAS; abdulkadir.elmas@giresun.edu.tr

1. Introduction

Breast cancer (BRCA) remains a leading cause of cancer-related deaths worldwide, emphasizing the need for innovative therapeutic strategies (Giaquinto et al., 2024). Advances in genomic and proteomic technologies have transformed our understanding of this heterogeneous disease, enabling the discovery of novel biomarkers and therapeutic targets (Engel & Kaklamani, 2007; Mertins et al., 2016; Krug et al., 2020). Our prior work identified a subset of overexpressed protein kinases in breast cancer, as well as a number of tumor-expressed proteins, revealing potential tumor-specific vulnerabilities in breast cancer (Elmas, 2024). Building on these insights, the present study aims to broaden the scope of analysis to include both kinase and non-kinase proteins (Manning et al., 2002), focusing on tumor-specific overexpression patterns in breast cancer cohort. By integrating overexpression and differential expression data, we aim to distinguish markers uniquely upregulated in breast cancer tumors relative to their background expressions in matched normal tissue. Additionally, by comparing mRNA and protein-level overexpression, we investigate post-translational regulation to identify novel, protein-specific markers that are not evident at the transcriptomic level (Lang et al., 2020).

Previous multi-omics studies integrating transcriptomic and proteomic data have significantly advanced our understanding of breast cancer biology (Mertins el al., 2016). However, these approaches often face challenges in capturing protein-level regulatory mechanisms, as mRNA expression does not always correlate with protein abundance due to post-transcriptional modifications, protein degradation, and translational regulation (Vogel & Marcotte, 2012). Additionally, most studies lack functional validation and primarily focus on differential expression, without considering whether overexpressed genes/proteins are functionally essential for cancer cell survival. To overcome these limitations, our study integrates “expression-driven dependency” analysis, which allows for the identification of both gene- and protein-level functional dependencies, uncovering protein-specific vulnerabilities that may be overlooked in transcriptomic analyses.

A central aspect of this study is validating the functional importance of tumor-specific markers through expression-driven dependency (Elmas et al., 2024). Using cancer cell dependency and epression data sets from the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012; Nusinow et al., 2020), we assess the correlation between gene knockout sensitivity and expression levels of candidate markers. This concept highlights cancer cells' heightened vulnerability to the depletion of highly expressed genes, presenting a promising avenue for precision oncology. To advance this approach, we employ our previously developed BEACON method (Elmas et al., 2024; Elmas & Huang, n.d.), a Bayesian framework that integrates multi-omics data—including transcriptomic, proteomic, and genetic dependency information—to systematically identify critical cancer dependencies across diverse molecular levels and tissue types.

Through these expanded analyses, we aim to uncover novel tumor protein biomarkers supported by functional *in vitro* cell assays and deepen our understanding of the proteomic underpinnings of breast cancer, paving the way for new targeted therapeutic strategies.

2. Materials and methods

2.1. Proteomic, transcriptomic and functional data acquisition

Proteomic data were obtained through mass spectrometry (MS)-based analysis of a cohort comprising 116 breast cancer patients, including 115 tumor samples and 18 matched normal tissues (Krug et al., 2020). Protein quantification was performed using the PDAC workflow with stringent quality control measures, as detailed on the PDC portal (https://pdc.cancer.gov/pdc/browse). To ensure data consistency and reliability, normalization protocols were applied to address batch effects and standardize protein abundance levels, facilitating downstream analysis of 10,468 proteins, including 481 kinases. Median Absolute Deviation (MAD) normalization was employed for each sample, ensuring unit MAD across the dataset. Proteins with more than 20% missing values were excluded from the analysis to maintain data integrity. Additionally, principal component analysis (PCA) was performed to assess technical biases and potential batch effects. The PCA results indicated that sample clustering was driven by biological factors (Figure S1) rather than technical grouping (Figure S2), supporting the conclusion that batch effects were minimal.

For transcriptomic analyses, RNA sequencing (RNA-seq) data encompassing gene expression profiles of 106 tumor samples were sourced from the GDC portal (https://portal.gdc.cancer.gov/). Quantile normalization and log2 transformation were applied to FPKM-normalized counts, and genes with no expression (FPKM ≤ 1) in at least 20% of the samples were excluded. This filtering resulted in 21,395 protein-coding genes, of which 596 encoded kinases.

Functional validation was conducted by integrating RNA-seq data (Ghandi et al., 2019) and protein expression data (Nusinow et al., 2020) from breast cancer cell lines, available through the Cancer Cell Line Encyclopedia (CCLE) at the Cancer Dependency Map (DepMap) portal (https://depmap.org/portal/) (public release 22Q2). Genetic dependency data were also acquired from DepMap (Tsherniak et al., 2017), incorporating genome-wide CRISPR knockout screening (Meyers et al., 2017; Pacini et al., 2021; Dempster et al., 2021) performed on the same breast cancer cell lines.

The prioritized markers were evaluated based on their druggability (DGIdb), consistency across molecular layers (mRNA/protein overexpression), and functional relevance (GED/PED analyses).

2.2. Overexpression and differential expression analyses

In our previous work (Elmas, 2024), overexpression and differential expression analyses were performed to identify proteins showing abundance in the overall BRCA cohort and the proteins showing significant difference in expression (DEP) between tumor and matched normal tissues, respectively. In this work, we integrated the overexpressed and DEP markers to prioritize tumor-specific overexpressed markers for subsequent analyses and functional validation, which were further stratified by druggability (Figures 1, 2), molecular consistency (Figure 3), and cellular expression-driven dependency patterns (Figure 4).

2.3. Identification of concordant and discordant overexpressed targets

To investigate the relationship between mRNA and protein overexpression in breast cancer, we analyzed proteomic and transcriptomic data from the BRCA cohort (Figures 1, 2). A marker's (mRNA/protein) overexpression rate was assessed by OPPTI method (Elmas et al., 2021), based on comparing the marker's expression levels to the inferred (background) expressions across tumors computed by a weighted k-nearest neighbor (KNN) algorithm. This algorithm improves upon existing methods by using a k-nearest neighbor approach to infer a protein's background expression level based on co-expressed protein neighbors, rather than relying on univariate analysis or a single neighbor, reducing bias from noise. The algorithm computes an overexpression score for each protein in each tumor sample, measuring the deviation of observed expression from the inferred (background) value. Statistical analyses are performed to identify targets with significant overexpression rates via permutation approach. To generate a null distribution, dysregulation scores are randomly reassigned among proteins within each sample, and overexpression events are recalculated. This process is repeated N = 100 × 'number of samples' times to account for the cohort size. The null overexpression values from all iterations are then used to construct a permutation-based distribution, and a p-value is assigned to each marker by estimating the likelihood of its observed overexpression occurring by random chance. Additionally, false discovery rate (FDR) corrections were applied by BH method (Benjamini & Hochberg, 1995) to control for multiple testing.

Proteins were categorized as concordant if overexpression was observed at both mRNA and protein levels, while discordant proteins exhibited significant protein overexpression (PRO > 10%, FDR < 0.05) with minimal RNA-level alteration (RNA < PRO/2). We restricted our (protein/mRNA) discordant overexpression analyses to proteins with an overexpression rate of at least 10%, ensuring that a substantial relative (2-fold) difference is achieved between the protein and mRNA expression levels across the cohort. This threshold was chosen to specifically capture cases where the relative disparity between mRNA and protein overexpression is pronounced (i.e., "PRO > 10%" and "RNA < PRO/2" yields "PRO-RNA > 5%"), ensuring that at least five tumor samples within the cohort exhibit a meaningful difference (112 × 5% ≈ 5). We focused on proteins implicated in key oncogenic signaling pathways (Sanchez-Vega et al., 2018), including Cell Cycle, WNT, RTK/RAS, TP53, NOTCH, and HIPPO pathways. The identified concordant and discordant targets were visualized on scatter plots and heatmaps (Figure 3).

2.4. Integration of functional assays (expression-driven dependency)

The BEACON method (Elmas et al., 2024) was employed to integrate transcriptomic, proteomic, and genetic dependency datasets corresponding to breast cancer cell lines to identify expression-driven dependencies (ED) at both mRNA and protein levels. BEACON incorporates prior distributions to account for outliers and uneven data distributions, ensuring robust correlation measurements. Unlike traditional correlation-based methods, BEACON models gene expression levels and dependency scores using a bivariate Gaussian distribution and employs Markov Chain Monte Carlo (MCMC) sampling to estimate the correlation coefficient, rho (ρ). To determine statistical significance, BEACON compares each gene's ρ estimate to a null distribution, assuming no correlation ($\rho = 0$). The method calculates a z-score for each gene by measuring how far its MCMC-derived ρ deviates from the expected null value, standardized by the variance observed in the simulated distribution. Finally, p-values are computed to evaluate the likelihood of obtaining the observed correlation by chance, and the false discovery rate (FDR) correction is applied by Benjamini-Hochberg (BH) procedure (Elmas et al., 2024). In this work, expression vs. dependency relationships were quantified by estimating the Bayesian correlation coefficient ρ through MCMC sampling, using 3 parallel chains (n.chains) and 500 iterations (n.iter) with 100 exhausted for adaptation phase (n.adapt). The significant dependencies were defined as $\rho < -0.25$ and FDR < 0.05.

Expression-driven dependency (ED) analyses were conducted to evaluate the functional importance of our prioritized candidate targets. Gene expression levels were correlated with CRISPR knockout sensitivity scores from DepMap across 23 breast cancer cell lines with mRNA expression data and 44 with protein expression data. GED (gene expression-driven dependency) and PED (protein expression-driven dependency) were defined based on significant negative correlations between expression levels and knockout sensitivity (Figure 4). Additionally, proteins exhibiting consistent ED at both mRNA and protein levels were classified as concordant dependencies, while those with PED but not GED were considered protein-specific dependencies.

3. Results

3.1. Breast cancer proteomics cohort

We analyzed genomic and extensive mass-spectrometry (MS)-based proteomic data from a cohort of 116 breast cancer patients, which included 115 tumor samples and 18 matched normal tissues (Krug et al., 2020). To ensure high-quality results, we implemented a robust normalization protocol and rigorous quality-control procedures (detailed in Methods). This enabled us to accurately quantify a total of 10,468 proteins for downstream analysis. Furthermore, we utilized the Drug-Gene Interaction database (DGIdb) (Cotto et al., 2018) to compile a comprehensive list of genes associated with drug compounds. By cross-referencing this list with the proteins quantified in our dataset, we identified 1,694 druggable proteins that are potential targets for existing therapeutic interventions.

3.2. Inter-tumor and tumor-normal analyses

Tumor-expressed proteins (DEPs) often exhibit altered expression profiles; however, not all DEPs represent suitable candidates for therapeutic interventions due to variability in expression and functional redundancy (Elmas et al., 2022). To address this, we conducted an integrative analysis combining (inter-tumor) protein overexpression and (tumor-normal) differential expression data to identify aberrantly expressed proteins that consistently (across the cohort) show elevated levels in tumor tissue samples relative to matched normal samples. This approach enables the identification of robust therapeutic targets that are more pronounced in cancerous tissues.

In the BRCA cohort, we identified 296 kinases that were both differentially expressed and overexpressed, as defined by DGIdb as targetable. Among these, DCLK2 kinase exhibited significantly high (FDR < 0.05) alterations: (log2-fold-change [FC] = 2.6, FDR = 0.03; overexpression rate [PRO] = 19.3%, FDR = 1.6e-4) (Figure 1A) (Table S1). For non-kinase targets, 1,350 proteins demonstrated concurrent differential expression and overexpression, with 24 reaching high statistical significance (FDR < 0.05), including CD5 (FC = 4.8, PRO = 23%), SRD5A3 (FC = 2.9, PRO = 25%), FCER1G (FC = 2.6, PRO = 21%), SDS (FC = 3.4, PRO = 17%), RARA (FC = 2.9, PRO = 16%), VKORC1 (FC = 2, PRO = 21%) (Figure 1B) (Table S1).

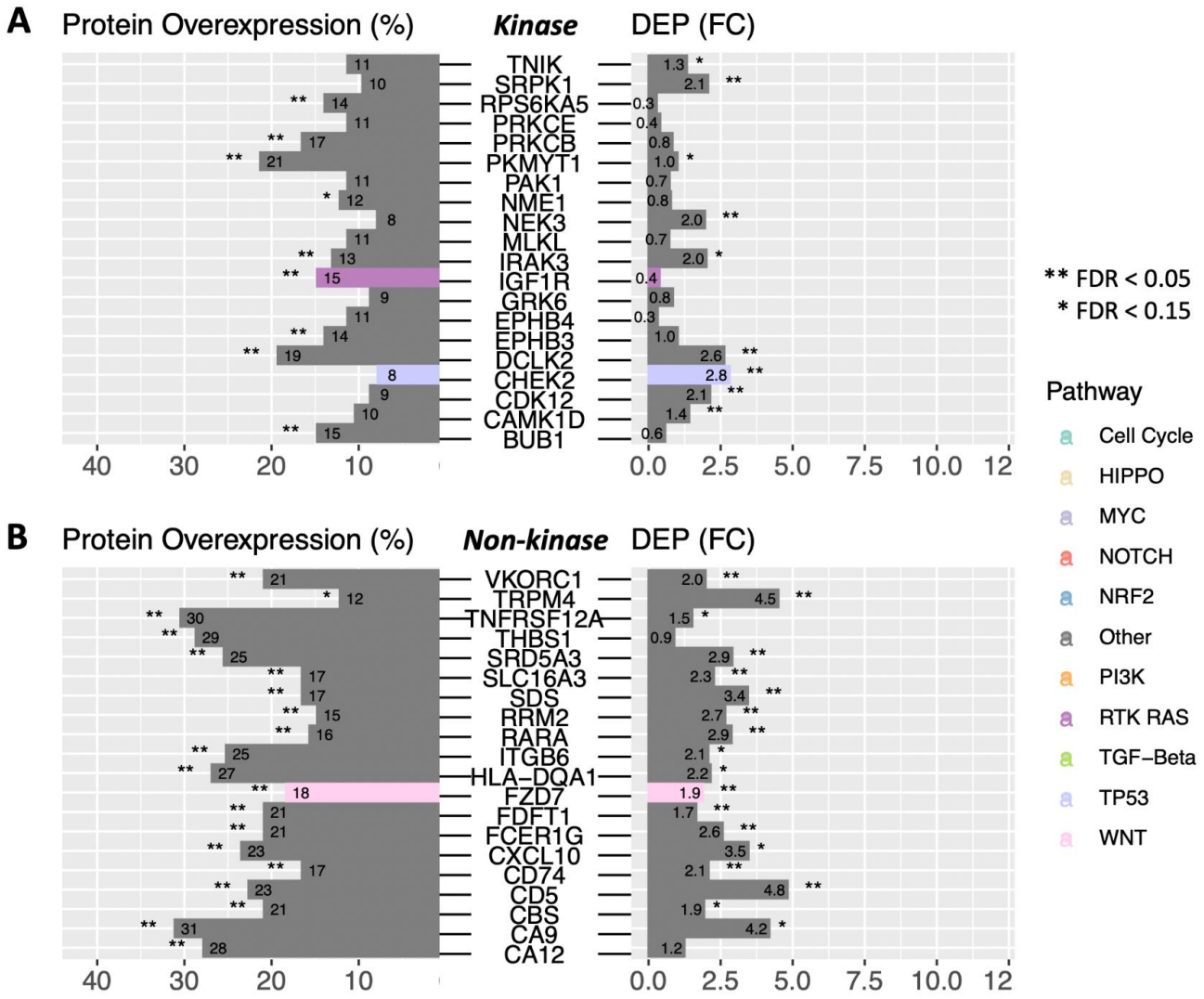

Figure 1. Top-20 tumor-specific overexpressed protein markers in BRCA identified in DGIdb. (A) Kinase proteins and (B) Non-kinase proteins are shown with their percentages of overexpression in tumors (left) and fold changes (FC) in differential expression between tumor and matched normal tissues (right). Significant fold changes are indicated with * (FDR < 0.15) and ** (FDR < 0.05). Pathways associated with each protein are color-coded.

We further analyzed proteins not targeted by existing drugs curated in DGIdb to identify novel opportunities for therapeutic intervention. Among kinases, 171 were both differentially expressed and overexpressed, with three displaying significant (FDR < 0.05) alterations: TK1 (FC = 3.6, PRO = 17%), ETNK2 (FC = 3.5, PRO = 17%), PTK7 (FC = 1.4, PRO = 13%) (Figure 2A) (Table S2). For non-kinase proteins, 144 were identified as significantly differentially expressed and overexpressed (FDR < 0.05). Notable examples include CNTN4 (FC = 3.6, PRO = 38%), ITGB8 (FC = 5.5, PRO = 28%), PEG3 (FC = 3.7, PRO = 29%), SHISA2 (FC = 3.1, PRO = 30%), DACH1 (FC = 2, PRO = 30%), HNRNPA1L2 (FC = 10.6, PRO = 23%), and UBE2C (FC = 4.7, PRO = 20%). (Figure 2B) (Table S2).

These findings highlight a subset of proteins with both significantly elevated (**) differential expression and overexpression levels, providing a prioritized list of candidate therapeutic targets in breast cancer (n=172; Tables S1, S2). By integrating protein-level data and druggability, we thus established a foundation for subsequent analyses, i.e., exploring mRNA-/protein-level abundance and functional dependencies to uncover potential vulnerabilities in breast cancer cells.

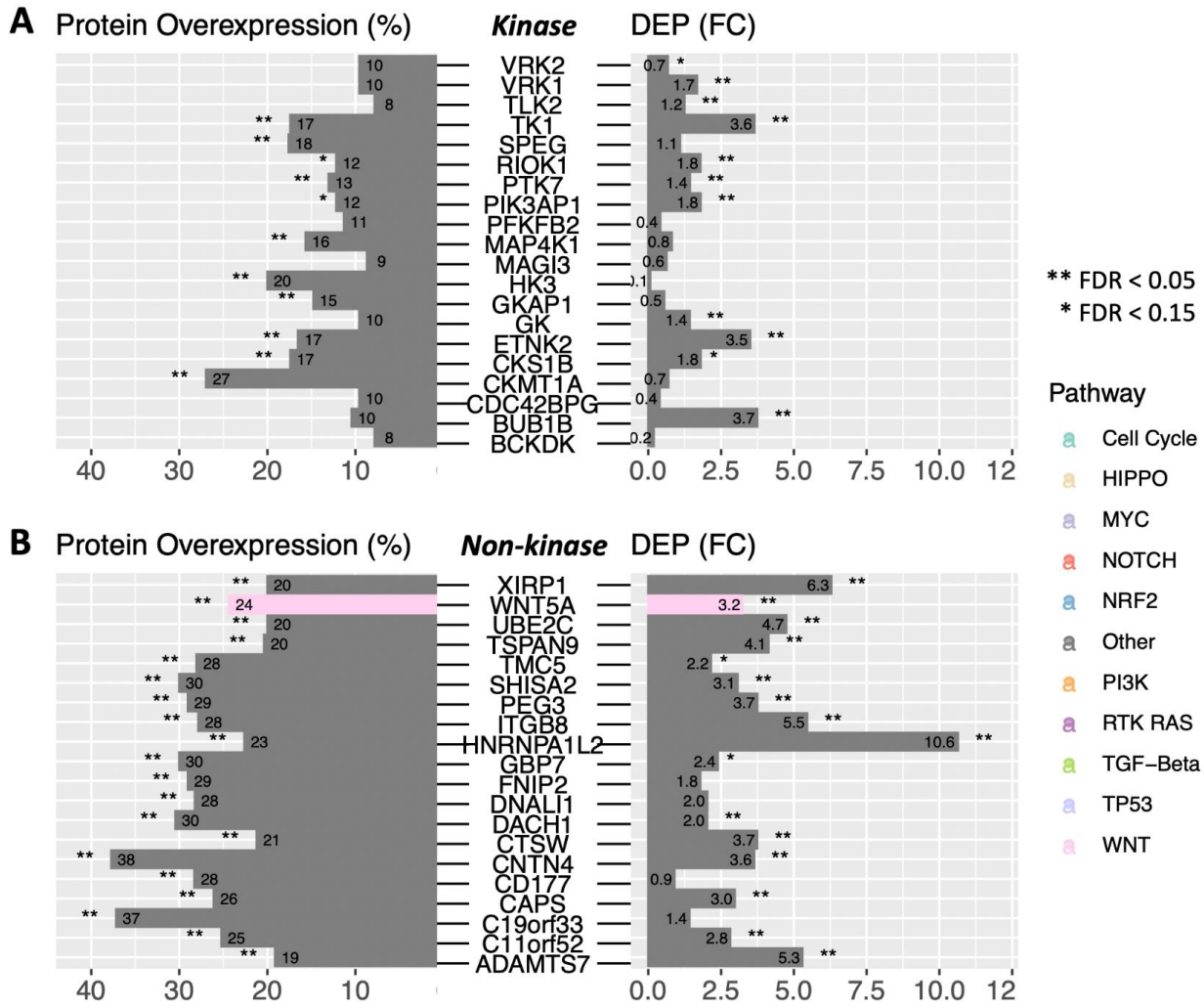

Figure 2. Top-20 tumor-specific overexpressed protein markers in BRCA cohort that are currently not targeted in DGIdb. (A) Kinase and (B) Non-kinase proteins are presented with tumor overexpression percentages (left) and fold changes (FC) in differential expression between tumor and matched normal tissues (right). Significant FC values are marked with * (FDR < 0.15) and ** (FDR < 0.05). Pathways are color-coded for each protein.

3.3. Comparing RNA and protein level alterations

Protein overexpression can stem from genomic alterations, such as copy-number amplifications, or post-transcriptional mechanisms not detectable at the DNA or RNA levels (Elmas et al., 2021). To disentangle these factors, we systematically compared the prevalence of protein overexpression with transcriptomic aberrations in the BRCA cohort, focusing on proteins involved in oncogenic signaling pathways (Sanchez-Vega et al., 2018). We aim to reveal markers of novel therapeutic potential by distinguishing those with concordent or discordant overexpression patterns in different molecular layers.

We identified 17 concordant targets with significant overexpression (≥10, FDR < 0.05) at both mRNA and protein levels, including CCND1 (mRNA overexpression rate [RNA] = 24%, PRO = 18%) and CDK6 (RNA = 15%, PRO = 13%) from Cell Cycle pathway, FZD7 (RNA = 23%, PRO = 18%), WNT5A (RNA = 23%, PRO = 24%), and SFRP4 (RNA = 15%, PRO = 23%) from WNT pathway, ERBB4 (RNA = 17%, PRO = 28%), IRS2 (RNA = 19%, PRO = 21%), ERBB2 (RNA = 16%, PRO = 18%), EGFR (RNA = 15%, PRO = 16%), and IGF1R (RNA = 14%, PRO = 15%) from RTK/RAS pathway, and TP53 (RNA = 17%, PRO = 19%) from TP53 pathway (Figure 3) (Table S3). Conversely, five discordant targets exhibited significant protein overexpression (PRO ≥ 10%, FDR < 0.05) with minimal RNA alterations (RNA < PRO/2), such as FHL1 (RNA = 8%, PRO = 18%) from NOTCH pathway, RASGRP2 (RNA = 6%, PRO = 17%) form RTK/RAS pathway, TEAD1 (RNA = 5%, PRO = 16%), FAT4 (RNA = 3%, PRO = 17%), and WWTR1 (RNA = 6%, PRO = 14%) from HIPPO pathway (Figure 3) (Table S3). These findings highlight potential therapeutic targets that may be overlooked by transcriptomic analyses alone.

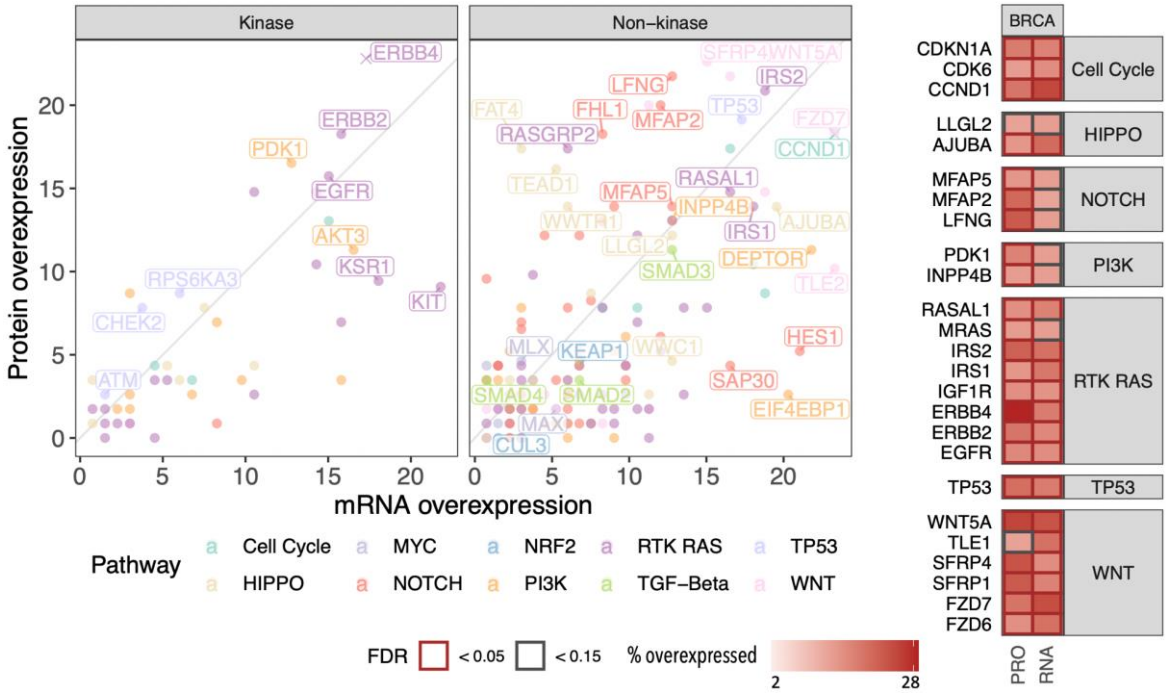

Figure 3. Concordant and discordant overexpressed targets in BRCA. Scatter plots (left) depict the relationship between mRNA and protein overexpression rates in BRCA cohort for kinase and non-kinase proteins. Overexpression rates were calculated using the OPPTI method, and significant targets were determined via permutation analysis with FDR correction (FDR < 0.15; FDR < 0.05). Concordant proteins (high mRNA and protein overexpression) align near the diagonal, while discordant proteins (high protein overexpression with minimal mRNA change) deviate significantly. Highly deviating values (e.g., ERBB4, PRO = 28%) are truncated for better visualization. Heatmaps (right) show pathway enrichment for prioritized targets across key oncogenic pathways, highlighting their potential role in BRCA tumorigenesis.

3.4. Expression-driven dependency

To explore functional gene and protein dependencies in breast cancer, we integrated transcriptomic, proteomic, and genetic dependency datasets from the Cancer Cell Line Encyclopedia (CCLE) and the Cancer Dependency Map (DepMap) (Tsherniak et al., 2017). Our goal was to evaluate the relationship between the expression levels of candidate genes and their knockout sensitivity across breast cancer cell lines, identifying potential vulnerabilities for therapeutic intervention (Methods). Given the limitations of available sensitivity data, only 23 breast cancer cell lines have corresponding mRNA-level expression and 44 with protein-level expression. To ensure a robust correlation measure despite these constraints, we employed BEACON method (Elmas et al., 2024)—a Bayesian correlation approach that incorporates prior distributions to model both expression and dependency data, effectively addressing outliers in the analysis.

Using BEACON, we identified 187 genes showing significant mRNA-level expression-driven dependency (GED) (correlation coefficient $\rho < -0.25$, FDR < 0.05) in breast cancer cell lines, notably ESR1 ($\rho = -0.89$), PROX1 ($\rho = -0.83$), GATA3 ($\rho = -0.81$), SPDEF ($\rho = -0.79$), and FOXA1 ($\rho = -0.75$) (Figure 4A) (Table S4). Among them, 12 were acting in oncogenic signaling pathways (predominantly from RTK/RAS), namely ERBB2 ($\rho = -0.67$), PDGFRB ($\rho = -0.55$), FGFR2 ($\rho = -0.53$), IRS2 ($\rho = -0.48$), and FGFR1 ($\rho = -0.46$). At the protein-level, we identified 64 significant ($\rho < -0.25$, FDR < 0.05) protein expression-driven dependencies (PED), including FOXP4 ($\rho = -0.8$), WDR47 ($\rho = -0.74$), E2F3 ($\rho = -0.76$), CCND1 ($\rho = -0.74$), and NPEPL1 ($\rho = -0.73$), of which four PEDs involved in oncosignaling pathways, i.e., E2F3 ($\rho = -0.76$) and CCND1 ($\rho = -0.74$) from Cell Cycle, and FGFR1 ($\rho = -0.58$) and ARHGAP35 ($\rho = -0.57$) from RTK/RAS pathway (Figure 4A) (Table S4). Furthermore, 11 genes showed consistent significant ED (FDR < 0.05) at both molecular layers, including, FOXA1 ($\rho < -0.66$), TRPS1 ($\rho < -0.63$), CCND1 ($\rho < -0.48$), SOX17 ($\rho < -0.62$), and CDH11 ($\rho < -0.41$). Notably, 53 proteins showed significant protein-specific ED (PED without GED; $\rho_{PRO} < -0.25$, $FDR_{PRO} < 0.05$, $FDR_{RNA} > 0.05$), including FOXP4 ($\rho_{PRO} = -0.8$, $\rho_{RNA} = -0.34$, $FDR_{RNA} = 0.15$), WDR47 ($\rho_{PRO} = -0.74$, $\rho_{RNA} = -0.26$, $FDR_{RNA} = 0.3$), E2F3 ($\rho_{PRO} = -0.76$, $\rho_{RNA} = -0.3$, $FDR_{RNA} = 0.23$), and NPEPL1 ($\rho_{PRO} = -0.73$, $\rho_{RNA} = -0.21$, $FDR_{RNA} = 0.37$).

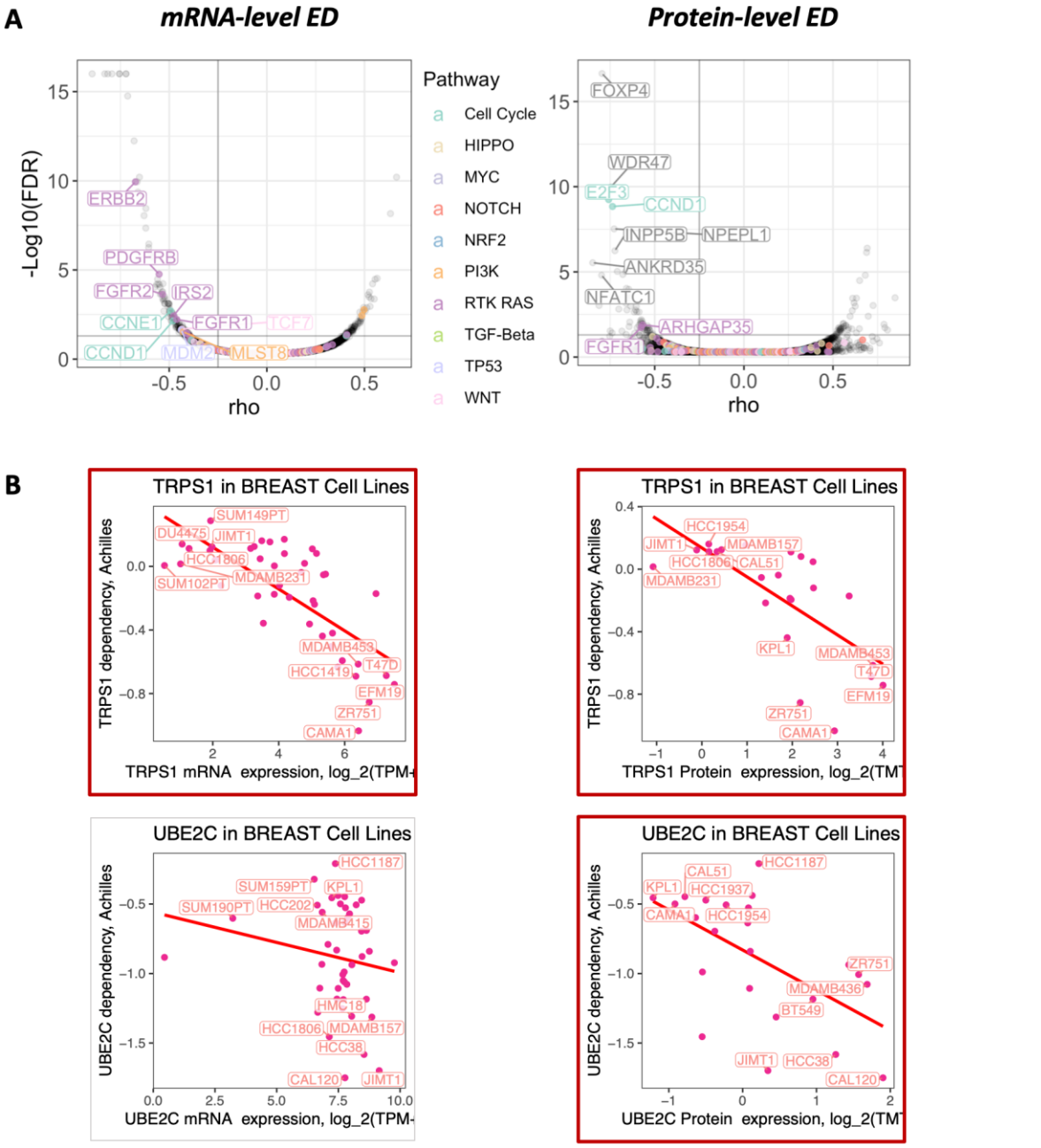

Figure 4. Identification of expression-driven dependencies (ED) in breast cancer cell lines using the BEACON method. (A) Volcano plots showing significant gene dependencies at the mRNA (left) and protein (right) levels, quantified by Bayesian correlation coefficients (ρ). Genes involved in oncogenic signaling pathways are highlighted and annotated, with significant expression-driven dependencies (GED or PED) defined as ρ < -0.25 and FDR < 0.05 (indicated by vertical and horizontal lines, respectively). (B) Correlation between gene/protein expression levels and knockout sensitivity for TRPS1 (top row) and UBE2C (bottom row) in breast cancer cell lines. Left panels depict mRNA expression-driven dependency (GED), while right panels show protein expression-driven dependency (PED). TRPS1 exhibits concordant GED and PED, while UBE2C shows PED without GED, emphasizing the value of protein-level analysis in identifying unique dependencies.

Among the 172 candidate targets we prioritized through significant results in both inter-tumor and tumor-normal analyses (Tables S1, S2; indicated by "**"), BEACON identified four candidates with significant GED at the mRNA level across breast cancer cell lines: TRPS1 (ρ = -0.71, FDR = 2e-15), DPM3 (ρ = -0.58, FDR = 2e-6), AGTRAP (ρ = -0.44, FDR = 0.02), and GIMAP6 (ρ = -0.42, FDR = 0.04) (Figure 4B) (Table S4). The negative correlation indicates that cell lines with higher TRPS1 expression exhibit greater sensitivity to TRPS1 knockout, suggesting its critical role in breast cancer cell survival. At the protein level, two proteins demonstrated significant PED: TRPS1 (ρ = -0.63, FDR = 0.001) and UBE2C (ρ = -0.53, FDR = 0.018) (Figure 4B) (Table S4). Notably, TRPS1 exhibited both GED and PED, indicating mRNA-protein concordance,

whereas UBE2C showed PED without corresponding GED, highlighting the importance of protein-level analyses in uncovering unique cancer cell vulnerabilities not readily observed by the transcriptomic data alone.

4. Discussion and conclusions

In this study, we leveraged a robust multi-omics approach to systematically identify tumor-specific protein markers and dependencies in breast cancer, revealing insights into potential therapeutic vulnerabilities. By integrating proteomic, transcriptomic, and functional dependency data, our analyses identified several previously unrecognized protein candidates, demonstrating the importance of considering both mRNA and protein-level overexpression in target prioritization, offering an additional perspective beyond conventional mutation-centered methods (Ellis et al., 2013; Krug et al., 2020).

Our integrative analysis identified hundreds of differentially expressed and overexpressed proteins in breast cancer tumors. By systematically analyzing both kinase and non-kinase proteins, we identified 172 prioritized candidates based on their tumor-specific overexpression (Figures 1, 2). Many of these proteins are implicated in key oncogenic signaling pathways, such as Cell Cycle, RTK/RAS, and WNT. Notably, we identified several concordant targets, such as CCND1, ERBB2, and EGFR, which exhibited significant overexpression at both mRNA and protein levels (Figure 3). These markers represent well-characterized drivers of breast cancer, reaffirming the validity of our approach. Importantly, our study also identified discordant targets, including FHL1, TEAD1, and FAT4, which exhibited significant protein overexpression with minimal RNA-level alterations (Figure 3). Such discrepancies can arise due to post-transcriptional regulation, protein stabilization, or translational control mechanisms that are not captured at the mRNA level (Lapek et al., 2017; Mertins et al., 2016; Elmas et al., 2021). For example, microRNA-mediated repression, altered protein degradation rates, or differential subcellular localization may drive protein accumulation independently of transcript abundance. Consequently, relying solely on transcriptomic analyses may miss biologically and clinically relevant protein-level changes that impact tumor behavior and therapeutic response. By focusing on protein-specific alterations, we were able to uncover novel potential therapeutic targets, such as UBE2C and FOXP4, that might otherwise be overlooked.

To assess the functional significance of the prioritized targets, we analyzed their (CRISPR) knockout sensitivity in breast cancer cell lines using genetic dependency data from the Cancer Dependency Map (DepMap), by focusing on genes that play a critical role in cancer cell survival due to their elevated expression levels. This analysis by BEACON approach identified both mRNA-level (GED) and protein-level (PED) expression-driven dependencies, emphasizing the importance of expression in determining cellular vulnerabilities (Figure 4). Notably, TRPS1 and UBE2C demonstrated high sensitivity to gene depletion, with breast cancer cell lines showing a strong negative correlation between their expression levels and knockout viability. Furthermore, concordant dependencies, such as TRPS1 and CCND1, highlight targets with consistent functional importance across molecular layers, cross-validating their vulnerability potential. Meanwhile, protein-specific dependencies, such as, UBE2C, E2F3, NPEPL1 and WDR47, illustrate the unique insights provided by proteomic analyses and can uncover novel vulnerabilities.

Notably, TRPS1 emerged as a high-priority target, exhibiting significant GED and PED correlations in breast cancer cells, as well as tumor-specific overexpression in the BRCA cohort. TRPS1 is a transcriptional cofactor recruited by the progesterone receptor and plays a critical role in modulating RANKL expression through epigenetic regulation in endometrial and breast cancers (Yang et al., 2023). TRPS1's context-dependent behavior, acting as a repressor in endometrial cancer and an activator in breast cancer, highlights its potential as a therapeutic target to enhance the anti-tumor effects of *Medroxyprogesterone* in endometrial cancer while mitigating its carcinogenic effects in breast cancer (Yang et al., 2023). Furhermore, TRPS1 has emerged as a promising diagnostic marker for breast cancer, as it demonstrates exceptional sensitivity and specificity across all breast cancer subtypes, especially for triple-negative breast cancer, outperforming the commonly used GATA3 marker (D. Ai et al., 2021).

We identified UBE2C as a therapeutic target through our overexpression and expression-driven dependency analysis, which revealed a unique protein-level vulnerability (PED) not observed at the mRNA level (GED). As a key regulator of the cell cycle within the ubiquitin-proteasome system, UBE2C plays a critical role in breast cancer progression. While elevated mRNA expression of UBE2C has been linked to poor disease-free and overall survival (Psyrri et al., 2012), our findings emphasize the importance of proteomic analyses in

uncovering functional dependencies that are undetectable via transcriptomic data alone. Given its strong association with aggressive tumor behavior and adverse prognosis, UBE2C holds promise as a therapeutic target. Its potential applications include patient stratification for aggressive chemotherapy regimens and targeted therapies, such as proteasome inhibitors like *bortezomib*, particularly when combined with HER2-targeting agents like *trastuzumab* (Cusack, 2003). These findings call for further investigation into the mechanistic and therapeutic implications of UBE2C in breast cancer.

Our findings offer significant implications for drug development. Several of the identified targets, such as ERBB2, EGFR, and CCND1 are already implicated in breast cancer therapies (Ali & Wendt, 2017; Yoon & Oh, 2024; B. Ai et al., 2019), demonstrating the relevance of our approach. Meanwhile, the novel targets identified in this study, including TRPS1, UBE2C, FOXP4, and E2F3, warrant further investigation to evaluate their therapeutic potential. For instance, E2F3 presents a promising therapeutic target in breast cancer, as siRNA-mediated silencing of its overexpression effectively suppresses tumor cell growth, highlighting its potential for both diagnostic and treatment strategies (Vimala et al., 2012). Comprehensive functional validation in preclinical models and drug-screening studies will be critical for translating these findings into clinical practice.

Despite the strengths of this study, several limitations must be acknowledged. First, the original dataset used in this study (Krug et al., 2020) consists of breast cancer patients distributed relatively evenly across the four major molecular subtypes: Basal, HER2, Luminal A, and Luminal B. Our analyses were performed across these subtypes, supporting the generalizability of our findings. Given this balanced representation, no additional external validation dataset was included in this study. However, the relatively small cohort size, particularly the limited availability of matched normal tissues, may affect the generalizability of our findings. Validation in larger and more diverse cohorts will be necessary to confirm the relevance of the identified targets across different breast cancer subtypes. Second, bulk proteomic analyses were used to identify tumor-specific overexpressed proteins. However, bulk proteomic data may not fully capture intratumoral heterogeneity and could mask subclonal variations. As single-cell proteomic technologies continue to advance, future studies will enable a more detailed investigation of intratumoral heterogeneity. Nevertheless, because our primary objective is to identify biomarkers that are consistently present across tumor subtypes and broadly generalizable, bulk analysis remains a powerful and appropriate approach. Additionally, although the BEACON method effectively handles multi-omics data with small sample sizes, the limited availability of proteomic dependency data may restrict the scope of PED analyses. Expanding proteomic datasets in future studies will enable a more comprehensive evaluation of protein-specific vulnerabilities. Finally, while this study establishes a robust framework for target discovery, comprehensive functional validation of the identified markers is a critical next step. Further experimental studies, including gene knockdown and drug-sensitivity assays in cell lines, such as (Elmas et al., 2022), and patient-derived models, e.g., (Huang et al., 2017), will be essential to assess the therapeutic potential of these targets.

In conclusion, this study demonstrates the utility of integrating proteomic, transcriptomic, and functional dependency data to identify tumor-specific markers and vulnerabilities in breast cancer. By uncovering both concordant and protein-specific targets, we present a prioritized list of novel actionable candidates for therapeutic development (Table S4). These findings advance our understanding of breast cancer biology and open new avenues for precision oncology strategies aimed at improving patient outcomes. The integration of these targets into clinical research could lead to the development of more personalized therapeutic interventions, enabling tailored treatments based on individual tumor profiles. Such strategies have the potential to improve early detection, enhance treatment efficacy, and ultimately contribute to better patient prognoses, moving us closer to more precise and effective cancer care.

Data and Software access

The supplementary tables referenced in this study (Tables S1-S4) are available at: https://www.columbia.edu/~ae2321/workspace/BRCA/SupplementaryTables.xlsx. The BRCA cohort data (Krug et al., 2020) is accessible through the Clinical Proteomic Tumor Analysis Consortium (CPTAC) resources provided by The National Cancer Institute. The data can be found at the following links: https://proteomic.datacommons.cancer.gov/pdc/study/PDC000116 and https://pdc.cancer.gov/pdc/. The BEACON and OPPTI softwares utilized in this work are available on GitHub: https://github.com/Huang-

lab/BEACON and https://github.com/Huang-lab/Oppti, respectively. All analyses were conducted using R programming language (version 3.6.2) through custom-written scripts.

Acknowledgement

The author extends gratitude to the BRCA cohort, along with the patients and families involved, whose generous contribution of data was essential to this work. The author also expresses gratitude to the members of the Huang lab (https://labs.icahn.mssm.edu/kuanhuanglab/members/) for their insightful discussions that significantly contributed to the development of this research. The author would like to thank the editor and the referees for their valuable contributions during the review and evaluation of this manuscript.

Declaration of ethical code

This research utilized publicly available data from previously published sources. The dataset is open access and can be freely obtained from the CPTAC Data Portal (https://pdc.cancer.gov/pdc/). As such, no separate ethics committee approval and/or legal-special permission were required for this study.

Conflicts of interest

The author declares that there is no conflict of interest to disclose.

References

Ai, B., Kong, X., Wang, X., Zhang, K., Yang, X., Zhai, J., Gao, R., Qi, Y., Wang, J., Wang, Z., & Fang, Y. (2019). LINC01355 suppresses breast cancer growth through FOXO3-mediated transcriptional repression of CCND1. *Cell Death & Disease*, *10*(7), 502. https://doi.org/10.1038/s41419-019-1741-8

Ai, D., Yao, J., Yang, F., Huo, L., Chen, H., Lu, W., Soto, L. M. S., Jiang, M., Raso, M. G., Wang, S., Bell, D., Liu, J., Wang, H., Tan, D., Torres-Cabala, C., Gan, Q., Wu, Y., Albarracin, C., Hung, M.-C., … Ding, Q. (2021). TRPS1: a highly sensitive and specific marker for breast carcinoma, especially for triple-negative breast cancer. *Modern Pathology*, *34*(4), 710–719. https://doi.org/10.1038/s41379-020-00692-8

Ali, R., & Wendt, M. K. (2017). The paradoxical functions of EGFR during breast cancer progression. *Signal Transduction and Targeted Therapy*, *2*(1), 16042. https://doi.org/10.1038/sigtrans.2016.42

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V, Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M. F., Monahan, J. E., Morais, P., Meltzer, J., Korejwa, A., Jané-Valbuena, J., … Garraway, L. A. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, *483*(7391), 603–607. https://doi.org/10.1038/nature11003

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Cotto, K. C., Wagner, A. H., Feng, Y.-Y., Kiwala, S., Coffman, A. C., Spies, G., Wollam, A., Spies, N. C., Griffith, O. L., & Griffith, M. (2018). DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Research*, *46*(D1), D1068–D1073. https://doi.org/10.1093/nar/gkx1143

Cusack, J. (2003). Rationale for the treatment of solid tumors with the proteasome inhibitor bortezomib. *Cancer Treatment Reviews*, *29*, 21–31. https://doi.org/10.1016/S0305-7372(03)00079-3

Dempster, J. M., Boyle, I., Vazquez, F., Root, D. E., Boehm, J. S., Hahn, W. C., Tsherniak, A., & McFarland, J. M. (2021). Chronos: a cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. *Genome Biology*, *22*(1), 343. https://doi.org/10.1186/s13059-021-02540-7

Ellis, M. J., Gillette, M., Carr, S. A., Paulovich, A. G., Smith, R. D., Rodland, K. K., Townsend, R. R., Kinsinger, C., Mesri, M., Rodriguez, H., & Liebler, D. C. (2013). Connecting genomic alterations to cancer biology with proteomics: The NCI clinical proteomic tumor analysis consortium. *Cancer Discovery*, *3*(10), 1108–1112. https://doi.org/10.1158/2159-8290.CD-13-0219

Elmas, A. (2024). Proteomic landscape of breast cancer tumors identifies novel therapeutic targets. In Altınok Bahar (Ed.), *3. Bilsel International Aspendos Scientific Researches Congress* (pp. 82–91). Astana.

Elmas, A., & Huang, K. (n.d.). *https://github.com/Huang-lab/BEACON*.

Elmas, A., Layden, H. M., Ellis, J. D., Bartlett, L. N., Zhao, X., Kawabata-Iwakawa, R., Obinata, H., Hiebert, S. W., & Huang, K. (2024). Expression-driven genetic dependency reveals targets for precision medicine. In *bioRxiv*. https://doi.org/10.1101/2024.10.17.618926

Elmas, A., Lujambio, A., & Huang, K.-L. (2022). Proteomic analyses identify therapeutic targets in hepatocellular carcinoma. *Frontiers in Oncology*, *12*, 814120. https://doi.org/10.3389/fonc.2022.814120

Elmas, A., Tharakan, S., Jaladanki, S., Galsky, M. D., Liu, T., & Huang, K.-L. (2021). Pan-cancer proteogenomic investigations identify post-transcriptional kinase targets. *Communications Biology*, *4*(1), 1112. https://doi.org/10.1038/s42003-021-02636-7

Engel, R. H., & Kaklamani, V. G. (2007). HER2-positive breast cancer: current and future treatment strategies. *Drugs*, *67*(9), 1329–1341. https://doi.org/10.2165/00003495-200767090-00006

Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., Barretina, J., Gelfand, E. T., Bielski, C. M., Li, H., Hu, K., Andreev-Drakhlin, A. Y., Kim, J., Hess, J. M., Haas, B. J., Aguet, F., Weir, B. A., Rothberg, M. V., Paolella, B. R., … Sellers, W. R. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*, *569*(7757), 503–508. https://doi.org/10.1038/s41586-019-1186-3

Giaquinto, A. N., Sung, H., Newman, L. A., Freedman, R. A., Smith, R. A., Star, J., Jemal, A., & Siegel, R. L. (2024). Breast cancer statistics 2024. *CA: A Cancer Journal for Clinicians*. https://doi.org/10.3322/caac.21863

Huang, K.-L., Li, S., Mertins, P., Cao, S., Gunawardena, H. P., Ruggles, K. V, Mani, D. R., Clauser, K. R., Tanioka, M., Usary, J., Kavuri, S. M., Xie, L., Yoon, C., Qiao, J. W., Wrobel, J., Wyczalkowski, M. A., Erdmann-Gilmore, P., Snider, J. E., Hoog, J., … Ding, L. (2017). Proteogenomic integration reveals therapeutic targets in breast cancer xenografts. *Nature Communications*, *8*, 14864. https://doi.org/10.1038/ncomms14864

Krug, K., Jaehnig, E. J., Satpathy, S., Blumenberg, L., Karpova, A., Anurag, M., Miles, G., Mertins, P., Geffen, Y., Tang, L. C., Heiman, D. I., Cao, S., Maruvka, Y. E., Lei, J. T., Huang, C., Kothadia, R. B., Colaprico, A., Birger, C., Wang, J., … Clinical Proteomic Tumor Analysis Consortium. (2020). Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell*, *183*(5), 1436-1456.e31. https://doi.org/10.1016/j.cell.2020.10.036

Lang, G.-T., Jiang, Y.-Z., Shi, J.-X., Yang, F., Li, X.-G., Pei, Y.-C., Zhang, C.-H., Ma, D., Xiao, Y., Hu, P.-C., Wang, H., Yang, Y.-S., Guo, L.-W., Lu, X.-X., Xue, M.-Z., Wang, P., Cao, A.-Y., Ling, H., Wang, Z.-H., … Shao, Z.-M. (2020). Characterization of the genomic landscape and actionable mutations in Chinese breast cancers by clinical sequencing. *Nature Communications*, *11*(1), 5679. https://doi.org/10.1038/s41467-020-19342-3

Lapek, J. D., Greninger, P., Morris, R., Amzallag, A., Pruteanu-Malinici, I., Benes, C. H., & Haas, W. (2017). Detection of dysregulated protein-association networks by high-throughput proteomics predicts cancer vulnerabilities. *Nature Biotechnology*. https://doi.org/10.1038/nbt.3955

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science (New York, N.Y.)*, *298*(5600), 1912–1934. https://doi.org/10.1126/science.1075762

Mertins, P., Mani, D. R., Ruggles, K. V, Gillette, M. A., Clauser, K. R., Wang, P., Wang, X., Qiao, J. W., Cao, S., Petralia, F., Kawaler, E., Mundt, F., Krug, K., Tu, Z., Lei, J. T., Gatza, M. L., Wilkerson, M., Perou, C. M., Yellapantula, V., … NCI CPTAC. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature*, *534*(7605), 55–62. https://doi.org/10.1038/nature18003

Meyers, R. M., Bryan, J. G., McFarland, J. M., Weir, B. A., Sizemore, A. E., Xu, H., Dharia, N. V, Montgomery, P. G., Cowley, G. S., Pantel, S., Goodale, A., Lee, Y., Ali, L. D., Jiang, G., Lubonja, R., Harrington, W. F., Strickland, M., Wu, T., Hawes, D. C., … Tsherniak, A. (2017). Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nature Genetics*, *49*(12), 1779–1784. https://doi.org/10.1038/ng.3984

Nusinow, D. P., Szpyt, J., Ghandi, M., Rose, C. M., McDonald, E. R., Kalocsay, M., Jané-Valbuena, J., Gelfand, E., Schweppe, D. K., Jedrychowski, M., Golji, J., Porter, D. A., Rejtar, T., Wang, Y. K., Kryukov, G. V., Stegmeier,

F., Erickson, B. K., Garraway, L. A., Sellers, W. R., & Gygi, S. P. (2020). Quantitative proteomics of the Cancer Cell Line Encyclopedia. *Cell*, *180*(2), 387-402.e16. https://doi.org/10.1016/j.cell.2019.12.023

Pacini, C., Dempster, J. M., Boyle, I., Gonçalves, E., Najgebauer, H., Karakoc, E., van der Meer, D., Barthorpe, A., Lightfoot, H., Jaaks, P., McFarland, J. M., Garnett, M. J., Tsherniak, A., & Iorio, F. (2021). Integrated cross-study datasets of genetic dependencies in cancer. *Nature Communications*, *12*(1), 1661. https://doi.org/10.1038/s41467-021-21898-7

Psyrri, A., Kalogeras, K. T., Kronenwett, R., Wirtz, R. M., Batistatou, A., Bournakis, E., Timotheadou, E., Gogas, H., Aravantinos, G., Christodoulou, C., Makatsoris, T., Linardou, H., Pectasides, D., Pavlidis, N., Economopoulos, T., & Fountzilas, G. (2012). Prognostic significance of UBE2C mRNA expression in high-risk early breast cancer. A hellenic cooperative oncology group (HECOG) study. *Annals of Oncology*, *23*(6), 1422–1427. https://doi.org/10.1093/annonc/mdr527

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., Dimitriadoy, S., Liu, D. L., Kantheti, H. S., Saghafinia, S., Chakravarty, D., Daian, F., Gao, Q., Bailey, M. H., Liang, W.-W., Foltz, S. M., Shmulevich, I., Ding, L., Heins, Z., … Schultz, N. (2018). Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell*, *173*(2), 321-337.e10. https://doi.org/10.1016/j.cell.2018.03.035

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., Gill, S., Harrington, W. F., Pantel, S., Krill-Burger, J. M., Meyers, R. M., Ali, L., Goodale, A., Lee, Y., Jiang, G., Hsiao, J., Gerath, W. F. J., Howell, S., Merkel, E., … Hahn, W. C. (2017). Defining a cancer dependency map. *Cell*, *170*(3), 564-576.e16. https://doi.org/10.1016/j.cell.2017.06.010

Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nature Reviews Genetics, 13(4), 227-232. https://doi.org/10.1038/nrg3185

Vimala, K., Sundarraj, S., Sujitha, M. V., & Kannan, S. (2012). Curtailing overexpression of E2F3 in breast cancer using siRNA (E2F3)-based gene silencing. *Archives of Medical Research*, *43*(6), 415–422. https://doi.org/10.1016/j.arcmed.2012.08.009

Yang, L., Fan, Q., Wang, J., Yang, X., Yuan, J., Li, Y., Sun, X., & Wang, Y. (2023). TRPS1 regulates the opposite effect of progesterone via RANKL in endometrial carcinoma and breast carcinoma. *Cell Death Discovery*, *9*(1), 185. https://doi.org/10.1038/s41420-023-01484-0

Yoon, J., & Oh, D.-Y. (2024). HER2-targeted therapies beyond breast cancer — an update. *Nature Reviews Clinical Oncology*, *21*(9), 675–700. https://doi.org/10.1038/s41571-024-00924-9

Supplementary Figures

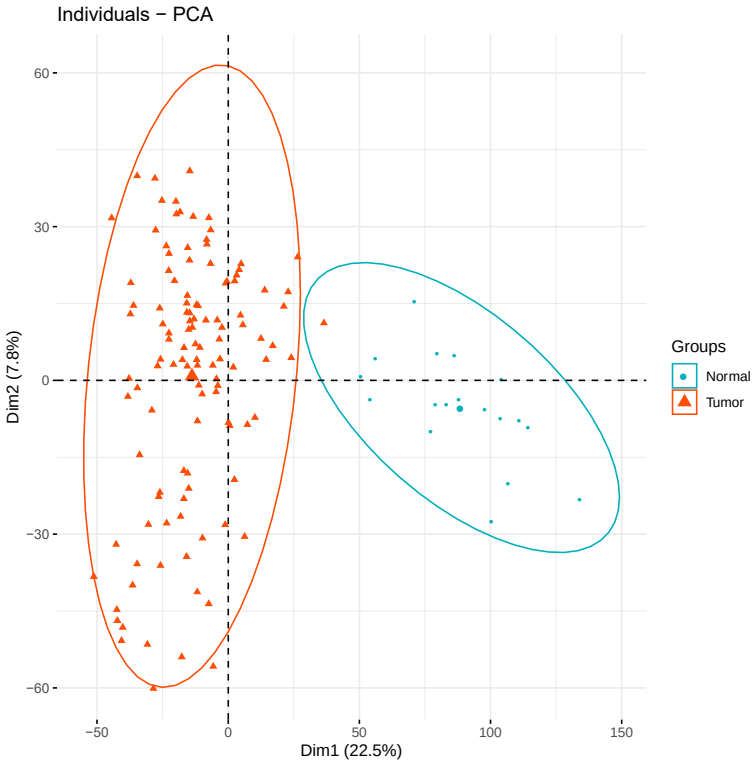

Figure S1. Principal Component Analysis (PCA) of normal and tumor samples. PCA was conducted to explore the primary sources of variation in the dataset. The analysis reveals clear separation between normal (blue) and tumor (red) samples, indicating that clustering is primarily driven by biological differences rather than technical artifacts.

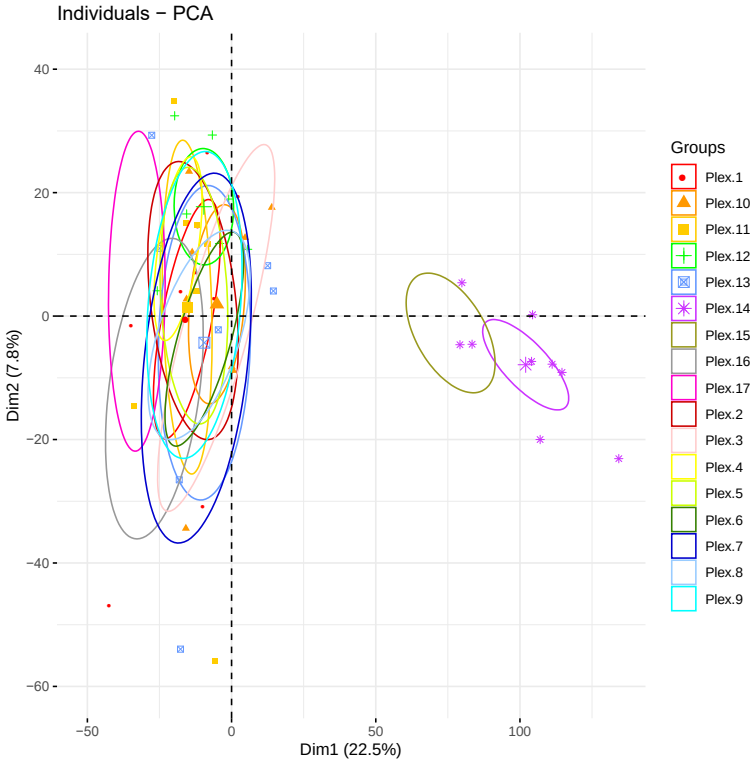

Figure S2. Principal Component Analysis (PCA) colored by sample batch. PCA was performed to evaluate potential batch effects. The absence of distinct clustering based on batch grouping suggests that batch effects are minimal, further supporting the robustness of the dataset.