

Article

Identification of Tumor Budding-Associated Genes in Breast Cancer through Transcriptomic Profiling and Network Diffusion Analysis

Panisa Janyasupab¹, Kodchanan Singhanat², Malee Warnnissorn³, Peti Thuwajit², Apichat Surataneer^{4,5} , Kitiporn Plaimas^{1,*}  and Chanitra Thuwajit^{2,*}

¹ Advance Virtual and Intelligent Computing (AVIC) Center, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand; panisa.janyasupab@gmail.com

² Department of Immunology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand; ksinghanat23@gmail.com (K.S.); peti.thu@mahidol.edu (P.T.)

³ Department of Pathology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand; maleewarn@yahoo.com

⁴ Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand; apichat.s@sci.kmutnb.ac.th

⁵ Intelligent and Nonlinear Dynamics Innovations Research Center, Science and Technology Research Institute, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

* Correspondence: kitiporn.p@chula.ac.th (K.P.); chanitra.thu@mahidol.ac.th (C.T.)

Abstract: Breast cancer has the highest diagnosis rate among all cancers. Tumor budding (TB) is recognized as a recent prognostic marker. Identifying genes specific to high-TB samples is crucial for hindering tumor progression and metastasis. In this study, we utilized an RNA sequencing technique, called TempO-Seq, to profile transcriptomic data from breast cancer samples, aiming to identify biomarkers for high-TB cases. Through differential expression analysis and mutual information, we identified seven genes (*NOL4*, *STAR*, *C8G*, *NEIL1*, *SLC46A3*, *FRMD6*, and *SCARF2*) that are potential biomarkers in breast cancer. To gain more relevant proteins, further investigation based on a protein–protein interaction network and the network diffusion technique revealed enrichment in the Hippo signaling and Wnt signaling pathways, promoting tumor initiation, invasion, and metastasis in several cancer types. In conclusion, these novel genes, recognized as overexpressed in high-TB samples, along with their associated pathways, offer promising therapeutic targets, thus advancing treatment and diagnosis for breast cancer.

Keywords: breast cancer; differential expression analysis; mutual information; network diffusion; tumor budding



Citation: Janyasupab, P.; Singhanat, K.; Warnnissorn, M.; Thuwajit, P.; Surataneer, A.; Plaimas, K.; Thuwajit, C. Identification of Tumor Budding-Associated Genes in Breast Cancer through Transcriptomic Profiling and Network Diffusion Analysis. *Biomolecules* **2024**, *14*, 896. <https://doi.org/10.3390/biom14080896>

Academic Editor: Da-Tian Bau

Received: 25 June 2024

Revised: 23 July 2024

Accepted: 23 July 2024

Published: 24 July 2024



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

1. Introduction

Breast cancer is the most commonly diagnosed cancer, with an estimated 2.3 million estimated new cases and 684,996 deaths in 2020 [1]. Tumor budding (TB), defined as an isolated single cell or a cluster of up to four cancer cells located at the invasive tumor front [2], has been identified as having prognostic significance in breast cancer [3]. Identifying genes differentially expressed between high and low TB levels can be proposed as theragnostic markers for better management of breast cancer.

Nowadays, gene expression datasets have proliferated greatly due to technological advancements. There are several techniques for obtaining gene expression data, such as microarray and RNA sequencing. Recently, Templated Oligo-Sequencing (TempO-Seq) has been introduced as a new transcriptomic platform that combines the advantages of microarray and RNA sequencing technology [4]. TempO-Seq is a high-throughput method for profiling gene expression directly from tissue sample lysates, including FFPE tissues,

without needing RNA extraction [5,6]. Its advantage lies in its ability to provide quantitative measurements of gene expression with high sensitivity and precision [7]. TempO-Seq provides count data for individual genes and is employed for comprehensive whole transcriptome profiling, primarily focusing on toxicogenomic research [7]. The standard pipeline for differential expression analysis, especially DESeq2, is well suited to TempO-Seq as well. The effectiveness of different normalization methods was evaluated when comparing two classes of TempO-Seq data, revealing that DESeq2 can provide relatively accurate results for absolute fold change levels of 2.0 or greater [8]. However, clinical data extracted from breast cancer are normally limited, with small numbers of cases found. A small and imbalanced set of samples can be problematic when analyzing differential gene expression because it limits the statistical power and robustness of the analysis. In the context of small sample sizes, mutual information may be more robust than traditional statistical tests because it does not rely on assumptions about the underlying distribution of the data and can handle non-linear relationships effectively. This makes it particularly useful for identifying subtle but informative features or genes that may be missed using methods based solely on mean differences or variance [9]. Despite having been a classical method for a long time, mutual information has demonstrated robustness [10] and effectiveness in handling imbalanced data [11,12], making it applicable across various domains. For instance, mutual information was employed alongside an improved Lasso method to filter unrelated genes and eliminate redundant ones, thus selecting informative genes from microarray gene expression data [13]. The large dimensionality of microarray data is reduced through combining mutual information and the Bayes theorem [14]. Additionally, biomarkers for stomach adenocarcinoma have been successfully identified through a combination of the “limma” method and joint mutual information [15].

Currently, it is evident that to enhance our comprehension of the functional pathways associated with biomarkers of interest, the utilization of network diffusion methods has proven effective in the discovery of additional relevant genes related to these biomarkers [16–19]. Consequently, this approach yields a more confident set of genes pertinent to biomarkers, thereby facilitating more robust pathway enrichment analyses that yield biologically meaningful results. Therefore, we then employed TempO-Seq technology to extract the transcriptomic data and to identify biomarkers associated with TB levels in breast cancer. Both differential expression analysis and mutual information were utilized to screen for novel biomarkers of high TB. Functional pathways relevant to the markers obtained were also retrieved. Literature support and evidence for further investigation into the diagnosis and prognosis for breast cancer were then discussed.

2. Materials and Methods

2.1. The Study Workflow

The workflow of our experiment and gene expression analysis is summarized in Figure 1. Initially, the transcriptomic data were processed using TempO-Seq, including preprocessing, correction, and normalization, until read counts were obtained for all genes. Subsequently, DESeq2 identified differentially expressed genes (DEGs) between high- and low-TB samples. Additionally, mutual information was employed to select genes with clearly distinguishable expression levels for high or low budding levels. The overlapping genes from these methods were identified as potential biomarkers for high-TB cases. Furthermore, a protein–protein interaction network and a network diffusion technique were applied to searching for more relevant proteins associated with TB progression, using these biomarkers as seeds. Finally, biological pathway enrichment was conducted to gain insights into the mechanisms underlying these biomarkers and their interactions within the protein–protein interaction network.

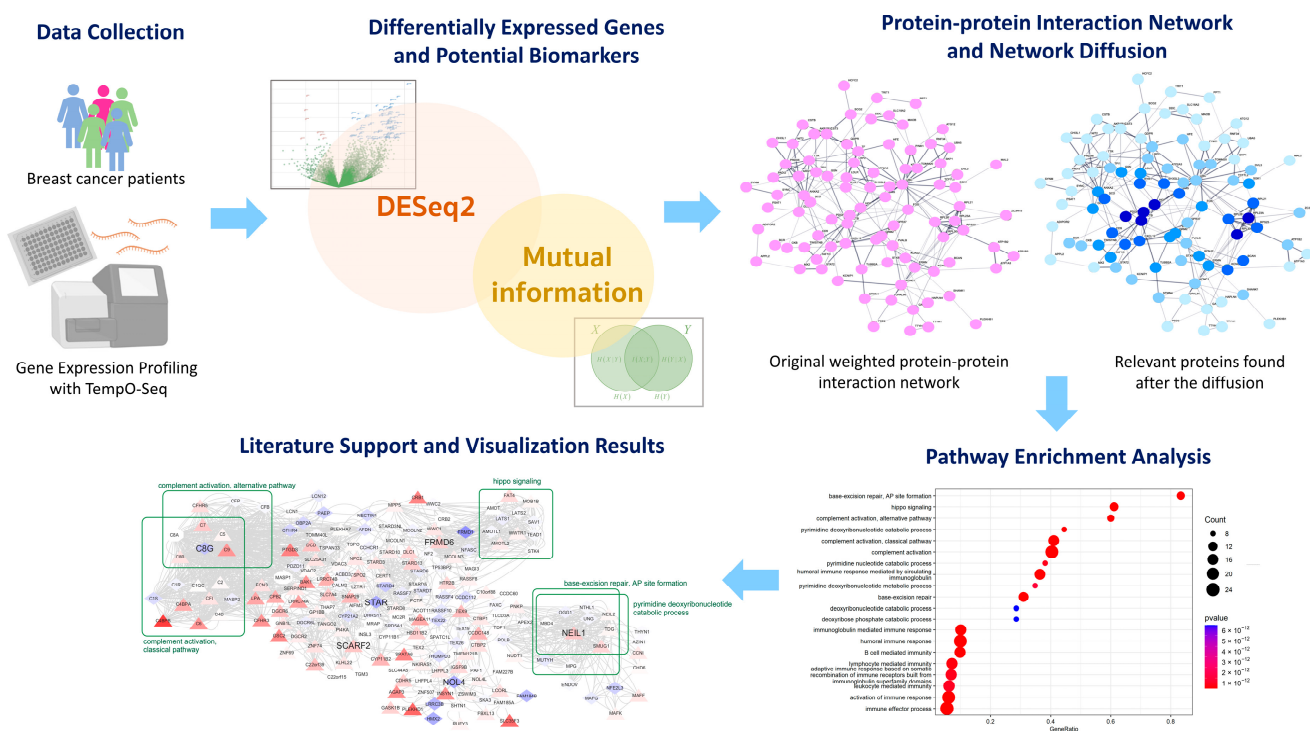


Figure 1. The workflow for identifying potential biomarkers of high-TB breast cancer and their related functional pathways.

2.2. Breast Cancer Sample Collection, Transcriptomic Data Preparation, and Tumor Budding Evaluation

Fifteen formalin-fixed paraffin-embedded (FFPE) breast cancer tissues from breast cancer patients who underwent surgery at Faculty of Medicine Siriraj Hospital, Mahidol University, between 2004 and 2017 were utilized for the study. Clinicopathological data, including patient age at diagnosis, histological subtype, tumor size, lymph node status, tumor staging, and survival status (measured by overall survival, OS), were collected (see Table 1). The protocol for collecting the samples and the clinicopathological data collection was approved by the Siriraj Institution Review Board according to human ethics standards (COA no. si542/2022). The cancer cell areas in the FFPE samples were labeled by a pathologist and sent for transcriptomic analysis using Templated Oligo-Sequencing (TempO-Seq) analysis (BioClavis, Glasgow, UK). Briefly, the FFPE tissues were deparaffinized and then digested. The tissue lysate was incubated with detector oligos annealing to the targeted RNA template [7]. Amplification of the ligated oligos was performed using a unique primer set for each sample, introducing a sample-specific barcode and Illumina adaptors. The samples were then sequenced on an Illumina HiSeq 2500 High Output v4 flow cell and analyzed using BCL2FASTQ software v2.20 (Illumina, San Diego, CA, USA). The obtained FASTQ files were aligned with the Human Whole Transcriptome v2.0 panel, consisting of 22,537 probes, using STAR [20]. The gene expression profile contains 15 cases, including 12 high- and 3 low-TB samples. Our gene expression dataset and its associated FASTQ files can be found in the NCBI Gene Expression Omnibus [21], with accession number GSE262825. Initially, the maximum mean gene expression was selected in case of duplication, resulting in a reduction in the total 22,537 probes to 19,703 genes.

Table 1. Demographic data, tumor budding level, and clinicopathological factors of patients.

| BCA Case | Age (y) | Tumor Size (cm) | OS (y) | DFS (y) | Tumor Staging | Pathological Differentiation | TB Level |
|----------|---------|-----------------|--------|---------|---------------|------------------------------|----------|
| B168 | 33 | 7.3 | 11.3 | 11.3 | IIIB | PD | low |
| TNBC-G06 | 67 | 4.5 | 5.3 | 5.1 | IIA | PD | low |
| B041 | 26 | 5.2 | 5.9 | 2.3 | IIIA | PD | high |
| B031 | 50 | 3 | 4 | 4.1 | IIIC | PD | high |
| B166 | 55 | 4 | 6.6 | 1.5 | IIIA | PD | high |
| TNBC-G01 | 67 | 2.4 | 5.6 | 5.3 | IIIA | MD | high |
| TNBC-G03 | 61 | 4 | 0.6 | 0.6 | IIA | MD | high |
| TNBC-G19 | 51 | 4.3 | 5.1 | 5.1 | IIIB | PD | high |
| TNBC-P04 | 54 | 3.5 | 3.2 | 3.2 | IIA | PD | high |
| TNBC-P06 | 50 | 3 | 5.2 | 4 | IIIC | MD | high |
| PC-B-099 | 50 | 6.5 | 2 | 1.3 | IIIA | PD | high |
| PC-B-130 | 44 | 6 | 0.8 | 0.6 | IIIC | PD | high |
| TNBC-P09 | 59 | 2.5 | 2.5 | 1.2 | IIA | PD | high |
| B165 | 49 | 4 | 11.5 | 1.8 | III | PD | low |
| IBC-11B | 60 | 4.5 | 15.1 | 15.1 | III | PD | high |

OS: overall survival; y: year; DFS: disease-free survival; PD: poorly differentiated; MD: moderately differentiated.

For tumor budding evaluation, the FFPE slides were sent to the Department of Pathology, Faculty of Medicine Siriraj Hospital, Mahidol University, for cytokeratin staining using routine services. Briefly, 4- μ m-thickness FFPE sections were incubated overnight at 4 °C in a humidified chamber with a 1:350 dilution of anti-pan-cytokeratin (AE1/AE3, Dako). Subsequently, the sections were processed using the K8002 detection kit, which includes an anti-mouse EnVision+ System HRP-labeled polymer (K4001), a substrate buffer, and diaminobenzidine (DAB) solution (Dako). The slides were then counterstained with hematoxylin. The stained slides were scanned at a resolution of 0.12 microns/pixel using a 3DHISTECH PANNORAMIC 1000 microscope by 3DHISTECH Ltd. and CaseViewer/QuantCenter software 2.4.0. (Sysmex, Budapest, Hungary), with a 40 \times objective lens, and saved in NRXS format. Quantification of positive membranous staining for pan-cytokeratin was performed using the image analysis software CaseViewer version 2.4 (3DHISTECH Ltd., Budapest, Hungary).

Tumor budding was defined as either one isolated tumor cell or up to four cells in clusters detached from the main tumor mass [22]. Within a low-power field (\times 50 original magnification), five areas with the highest budding number at the invasive margins were selected. Within a high-power field (200 \times original magnification), the number of budding cells was counted. According to the International Tumor Budding Conference (ITBCC) guidelines, the number of buds was categorized as low (0–4 buds), intermediate (5–9 buds), and high (\geq 10 buds). The tumor budding counts were independently reviewed by two examiners, one of whom was a medical doctor.

All the methods described in this study were performed in accordance with the relevant guidelines and regulations set forth by Siriraj Institutional Review Board (SIRB Protocol No.: 277/2565(IRB4), COA no. Si 542/2022), ensuring full compliance with the international guidelines for human research protection, such as the Declaration of Helsinki, the Belmont Report, CIOMS Guidelines, and the International Conference on Harmonization in Good Clinical Practice (ICH-GCP). Clinical samples from the patients were obtained after acquiring informed consent from the patients and were archival specimens which covered their usage in accordance with the protocol approved by Siriraj Institutional Review Board.

2.3. Differential Expression Analysis

Differential expression analysis is conducted using the “DESeq2” package in R [23]. The first step involves count normalization, which is performed to facilitate comparisons between samples. Next, the model estimates gene-wise dispersions to generate more accurate estimates of dispersion. Subsequently, the negative binomial model is fitted, and the Wald test is conducted to test the hypothesis. Finally, statistical information is prepared, including the log fold change (\log_{FC}) and p -value. \log_{FC} can be computed according to $\log_{FC} = \log_2 \left(\frac{x_{highTB}}{x_{lowTB}} \right)$, where x_{highTB} and x_{lowTB} are the normalized counts of high- and

low-TB samples, respectively. Positive and negative $\log FC$ values occur when a gene is up-regulated and down-regulated, respectively. Next, the p -value from DESeq2 tests the null hypothesis that there is no differential expression across the two groups. If the mean expression levels of the two groups are significantly different, the p -value of the gene is smaller. The adjusted p -value ($adj.p.val$) is then computed using the Benjamini–Hochberg method. Following the standard procedure, we select a subset of genes with the condition that $adj.p.val < 0.05$ for DEGs.

2.4. Mutual Information (MI)

The idea behind mutual information is to measure the relationship between two random variables or quantify the amount of information obtained from one random variable by observing another random variable. Let X and Y be random variables. MI quantifies the information shared between X and Y and is evaluated by $MI(X, Y) = H(X) + H(Y) - H(X, Y)$, where $H(X) = -\sum_{x \in X} p(x) \log(p(x))$, which is the entropy or the expected uncertainty in a random variable X , and $H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log(p(x, y))$ is the joint entropy, which measures the uncertainty when considering two random variables simultaneously. $p(x)$ and $p(y)$ stand for the marginal probability functions of X and Y , respectively, while $p(x, y)$ refers to the joint probability of both variables occurring together. In this study, the “praznik” [24] package in R is utilized, where X represents the normalized count of each gene and Y is a binary class consisting of high- and low-TB samples. In the scenario of scoring genes, $H(Y)$ is equal for all genes. Therefore, a high entropy of X , $H(X)$, indicates substantial unpredictability in the gene expression across the samples. Similarly, a high joint entropy $H(X, Y)$ signifies considerable variability in the combined distribution of gene expression and class labels. The MI score ranges from 0 to infinity. A mutual information score of 0 implies that knowing one variable provides no information about the other. Conversely, a high MI value suggests a strong relationship or dependency between variables X and Y . In the context of genes and classes, a gene with a higher MI score implies that it gains more information when the class is known. Therefore, when selecting genes based on mutual information scores, the set of genes with the highest MI scores is chosen. These genes are considered to have the strongest association with the class variable, making them potentially more informative for tasks such as classification or understanding the relationship between gene expression patterns and class labels.

2.5. The Protein–Protein Interaction Network and Network Diffusion Technique

The protein–protein interaction network was constructed from STRING database version 12.0 [25]. A subnetwork is selected based on the criterion of a combined score greater than 0.4, resulting in 19,038 protein nodes and 901,089 edges. Network diffusion is a network-based technique that requires the input of a network with initial seed nodes. Let G be the network with n nodes, and W be the weighted adjacency matrix of G . Assume that D is an $n \times n$ diagonal matrix where $D_{ii} = \sum_j W_{ij}$, and $L = D - W$ is called the Laplacian of the graph. The regularized Laplacian kernel is introduced in [26] as $K = (I + \alpha L)^{-1}$, where I is an $n \times n$ identity matrix. The network diffusion score is computed as $S_{diff} = K \cdot y$, where y is a binary vector of length n , with 1 assigned to the seed nodes and 0 otherwise. The overlapping set of genes found by the DESeq2 and mutual information techniques was identified as potential biomarkers. The proteins corresponding to these selected genes were considered seed nodes whose relevant scores were distributed along the edges of the PPI network. The diffusion scores were then sorted in descending order, and a top percentage was determined as a cutoff. This set of genes has a strong biological relationship with our identified genes and is further used for pathway enrichment analysis.

2.6. Pathway Enrichment Analysis

The pathway enrichment analysis is performed using “Metascape” [27]. This web-based tool includes functional and pathway terms from Gene Ontology (GO) processes, KEGG pathways, Reactome pathways, canonical pathways, WikiPathways, and CORUM pathways. The hypergeometric test with the Benjamini–Hochberg statistical correction algorithm is utilized. GO biological processes and KEGG pathways are chosen, with a minimum pathway size of 10 and a maximum of 500. Pathways meeting the significance criterion of a p -value < 0.01 are selected for further analysis.

3. Results

3.1. Potential Biomarkers by Differentially Expressed Genes and Mutual Information

In this study, we obtained fifteen formalin-fixed paraffin-embedded (FFPE) breast cancer tissues from breast cancer patients (see Section 2.2 for more details). With a low number of samples in comparison between high and low TB, traditional analysis of differentially expressed genes using DESeq2 may provide some bias due to the low data entry when calculating either a mean value or median value for comparing between the two groups of samples. Therefore, in our analysis, we propose the use of the mutual information technique (very well known as a great tool for feature selection techniques [13,28]). Since the mutual information will capture the values of all the members in one group and separate them from the other group, the final list comprises overlapping genes found to have significantly different expressions by DESeq2 and high uncertainty by the mutual information technique.

Using DESeq2, we identified 99 significantly differently expressed genes with an adjusted p -value < 0.05 and $|\log FC| > 1.5$ (see Table S1 for the complete list). Concurrently, a higher MI score indicates a closer relationship between a gene and its class, with each gene’s MI score provided in Table S2. However, there is no standard cutoff for MI scores. The distribution of the MI scores, shown in Figure S1, indicates that only 56 genes have the highest MI score of 0.5004. Interestingly, most of these genes exhibit a high classification power, as shown in Figure S2. The results from DESeq2 revealed a significant difference in the mean of the normalized counts between the two classes, while mutual information demonstrated greater efficacy in class separation, particularly for small sample sizes. The effectiveness of mutual information in providing better classification is discussed in Sections 3.2 and 3.3. In short, we demonstrate that the inclusion of MI or mutual information yields more reliable results and related functional pathways. Therefore, we utilized the intersection of these two approaches to enhance the likelihood of identifying potential biomarkers. Table 2 lists seven potential biomarkers identified using both methods, along with their corresponding statistical information. The baseMean values represent the average of the normalized read counts, while the logFC values indicate up-regulation or down-regulation. *NEIL1* exhibited the highest baseMean, indicating high expression across most samples, while *STAR* had the lowest. *NEIL1*, *SLC46A3*, *FRMD6*, and *SCARF2* were up-regulated, while *NOL4*, *STAR*, and *C8G* were down-regulated. Figure 2 depicts the normalized gene expression counts of all the biomarkers obtained.

Table 2. Statistical information on seven potential biomarkers.

| Genes | baseMean | logFC | lfcSE | Stat | p -Value | Adjusted p -Value |
|----------------|----------|-------|--------|---------|-----------------------|---------------------|
| <i>NOL4</i> | 200.02 | −5.14 | 1.1013 | −4.6685 | 3.03×10^{-6} | 0.002490 |
| <i>STAR</i> | 36.16 | −5.00 | 1.1492 | −4.3530 | 1.34×10^{-5} | 0.005210 |
| <i>C8G</i> | 216.37 | −4.84 | 0.9440 | −5.1302 | 2.89×10^{-7} | 0.000580 |
| <i>NEIL1</i> | 588.06 | 3.16 | 0.8325 | 3.8001 | 0.000145 | 0.029210 |
| <i>SLC46A3</i> | 244.48 | 2.97 | 0.6544 | 4.5456 | 5.48×10^{-6} | 0.003100 |
| <i>FRMD6</i> | 441.97 | 2.00 | 0.4376 | 4.5627 | 5.05×10^{-6} | 0.003100 |
| <i>SCARF2</i> | 159.33 | 3.28 | 0.8192 | 4.0070 | 6.15×10^{-5} | 0.015180 |



Figure 2. Plot depicting the expression of the identified gene set, where the y -axis represents the normalized counts. Note that for each gene, points with identical or closely similar values were slightly offset to enhance the visibility of individual points.

As shown in Table 2 and Figure 2, we identified seven potential genes within breast cancer: *NOL4*, *STAR*, *C8G*, *NEIL1*, *SLC46A3*, *FRMD6*, and *SCARF2*. Their biological and tumor-related functions are summarized in Table S3. In short, *NEIL1* (Nei Like DNA Glycosylase 1) is involved in DNA repair and is up-regulated in breast and lung cancer, reducing cancer cell death and promoting cancer progression [29–31]. *SCARF2* (Scavenger Receptor Class F Member 2) is involved in endocytosis, lipid metabolism, and cell adhesion and is up-regulated in various cancers, promoting tumor progression [32,33]. *NOL4* (Nucleolar Protein 4) is involved in ribosome biogenesis and RNA processing and is up-regulated in various cancers, also promoting tumor progression [34–38]. *STAR* (Steroidogenic Acute Regulatory Protein) is crucial for steroid hormone synthesis, with its increased expression promoting certain types of breast cancer [39–42]. *C8G* (complement C8 gamma chain) facilitates membrane attack complex (MAC) formation and cell lysis and is down-regulated in breast cancer, suggesting the suppression of cell lysis [43]. *FRMD6* (FERM Domain-Containing 6, known as Willin) is an upstream regulator of the Hippo signaling pathway and is mentioned to have both tumor-suppressive and tumor-promoting effects [44–47]. *SLC46A3* (solute carrier family 46 member 3) transports folate derivatives [44]. Loss of *SLC46A3* increases the resistance of HER2+ breast cancer to trastuzumab emtansine [44,48–50]. Further details on these genes and the supporting literature are discussed in our discussion section.

3.2. Effectiveness of Potential Biomarkers in Distinguishing between Low- and High-TB Samples

We assessed the effectiveness of potential biomarkers in distinguishing between low- and high-TB samples by employing heatmap plots to illustrate sample clusters based on their expression profiles. Figure 3 contains two columns. The first column illustrates the correlation between the expression profiles of the identified biomarkers for different techniques. The color intensity reflects the strength of the correlation between samples, revealing distinct expression patterns within each sample. The second column displays the corresponding principal component analysis (PCA), which visualizes the distributions of the samples based on the first two principal components derived from the expression data of the identified genes. Each sample is represented by a point, with colors and shapes indicating different TB levels. Figure 3A presents a heatmap of the correlations between the expression profiles of the 99 DEGs obtained across the BCA samples using DESeq2. Figure 3B shows a PCA plot of the first two principal components, where each point represents a sample, with the colors and shapes indicating different levels of tumor budding (TB). Notably, the three samples of low TB could not be grouped together into one cluster for the BCA samples, highlighting the limitations of solely relying on the traditional method of comparing the mean values for small sample sizes. In contrast, key biomarkers identified by the mutual information technique (Figure 3C,D) exhibit a

clear distinction between the low- and high-TB samples. Finally, the heatmap and PCA plots using the expression profiles of the biomarkers detected by both DESeq2 and mutual information are shown in Figure 3E,F. These potential biomarkers can classify samples with different TB levels, a task which is not achievable using the genes identified by DESeq2 or mutual information alone. Moreover, the expression profiles of the genes discovered via the mutual information technique demonstrate improved clustering of the low- and high-TB samples compared to those of DESeq2. However, the overlapping genes found by both DESeq2 and mutual information constitute the best set for distinguishing between low- and high-TB samples.

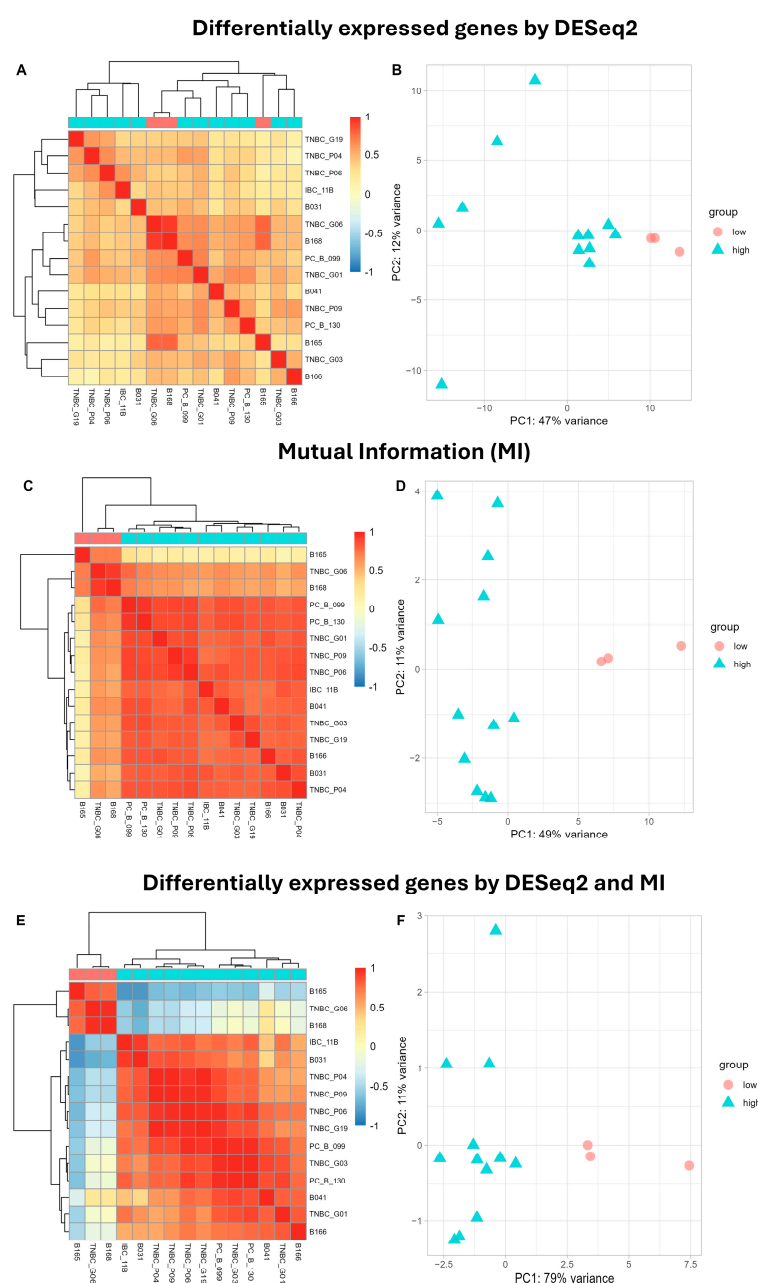


Figure 3. Heatmap and PCA plots utilizing the gene expression profiles of DEGs detected by DESeq2 (A,B), biomarkers detected by mutual information (C,D), and the seven potential biomarkers detected by DESeq2 and MI (E,F). The first column depicts the clusters and the correlations among expression profiles, while the second column displays the corresponding PCA plot of the first two principal components. Color intensities in the heatmaps reflect the strength of the correlation. Points in the PCA plots represent samples, with colors and shapes denoting different groups.

3.3. Relevant Proteins and Functional Pathway Enrichment Analysis

To enhance our understanding of the functional implications associated with the seven biomarkers (*NOL4*, *STAR*, *C8G*, *NEIL1*, *SLC46A3*, *FRMD6*, and *SCARF2*), we employed a network diffusion technique within a human protein–protein interaction network (details in Section 2.5). By using these biomarkers as seeds, we identified an additional 191 proteins related to the seeds for pathway enrichment analysis. Enrichment analysis was conducted using “Metascape” [27], focusing on Gene Ontology (GO) biological processes and KEGG pathways, resulting in 115 significant pathways (p -value < 0.01), as presented in Table S4. The top 20 enriched pathways were sorted based on the gene ratio within each pathway and visualized using the “clusterProfiler” R package [51], as depicted in Figure 4A. Using Cytoscape v3.10.1 [52], we visualized the PPI subnetwork enriched with the top five pathways, showcasing their interconnectedness in Figure 5.

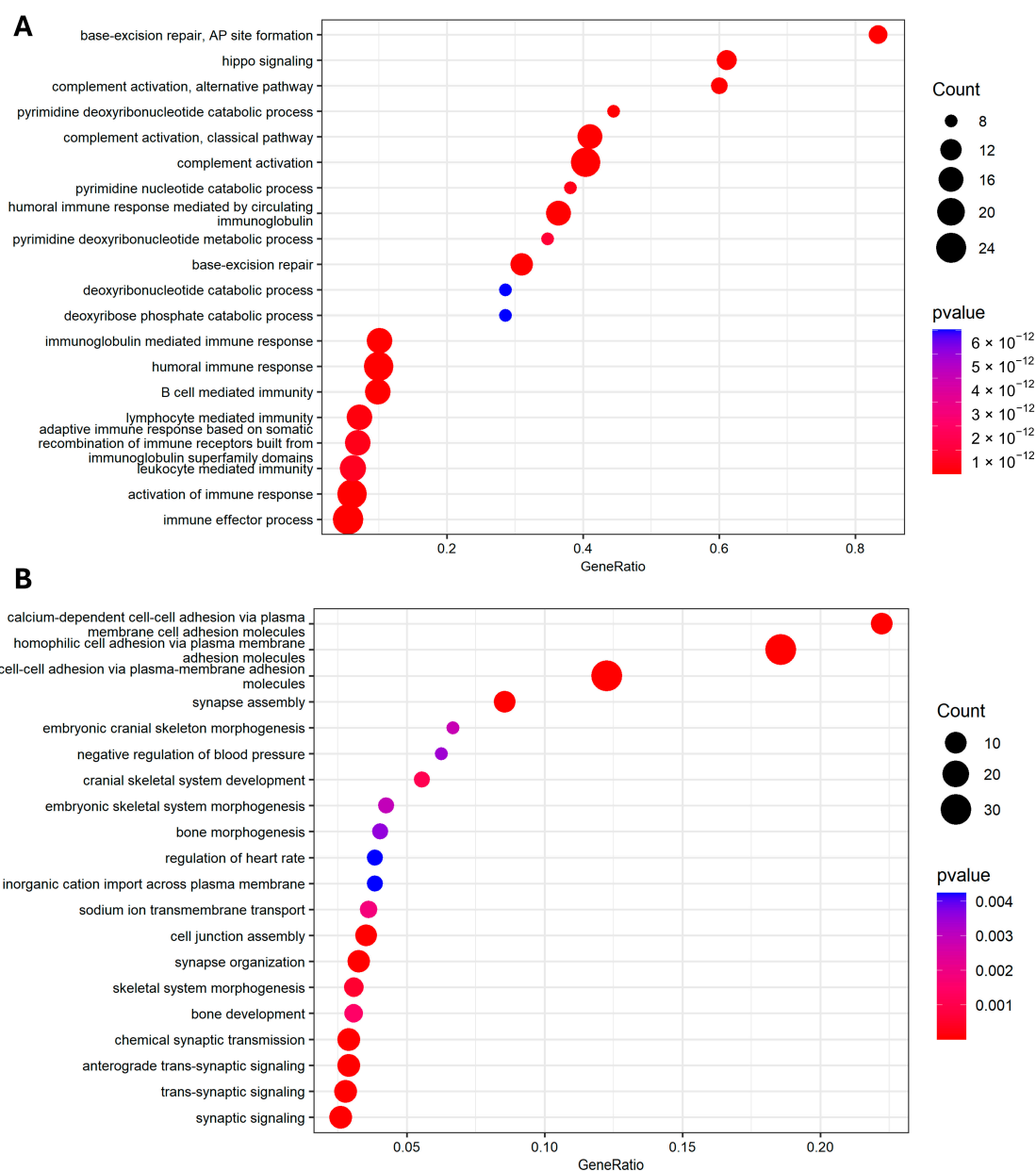


Figure 4. Functional enrichment for associated proteins using DESeq2 with MI (A) and without MI (B). Dot plots depicting the top 20 enriched GO biological processes. GeneRatio represents the ratio of the genes of interest appearing in the pathway to the total number of genes in the pathway.

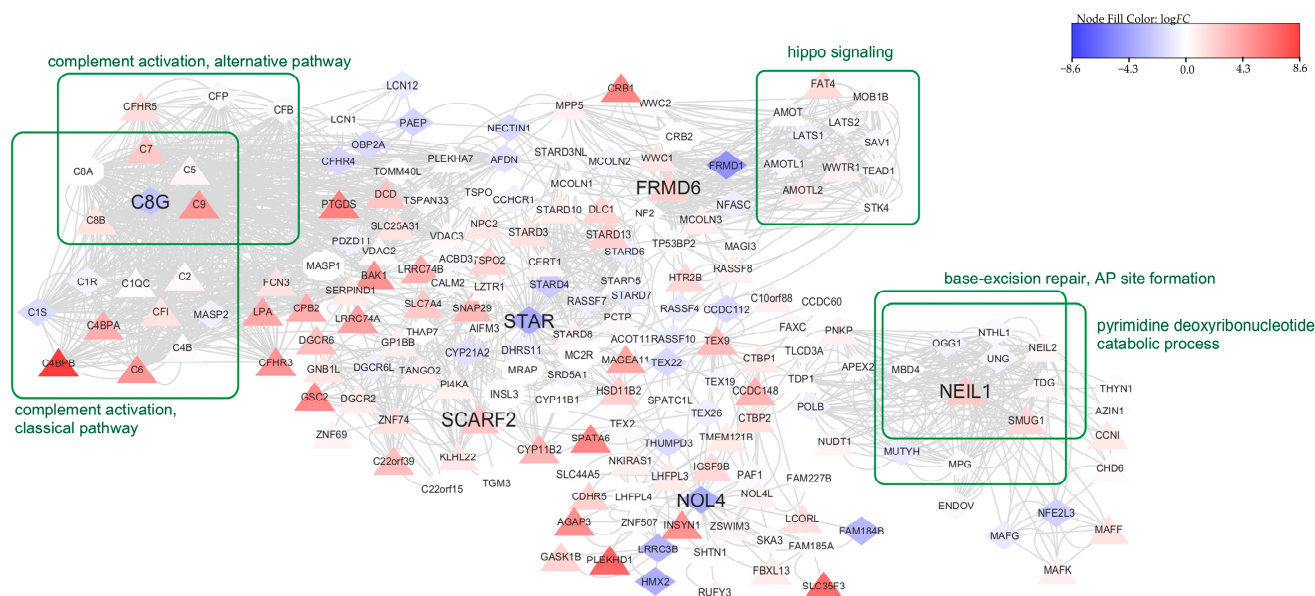


Figure 5. The protein–protein interaction subnetwork comprises proteins associated with the seven biomarkers detected by the diffusion technique as their diffusion scores rank above the 99th percentile. Node colors reflect the \log_{FC} values, with upward triangles denoting up-regulated genes, diamonds indicating down-regulated genes, and octagons representing genes with no significant change. Genes associated with a pathway of interest are grouped together.

Figure 4A illustrates the pathway enrichment results for the biomarkers detected by both DESeq2 and the MI technique. Base excision repair AP site formation and Hippo signaling were found to be significantly enriched with high gene ratios, both exceeding 0.6. Base excision repair proteins including *XRCC1*, *APE1*, *SMUG1*, and *FEN1* were significantly associated with poor breast-cancer-specific survival [53]. Additionally, Hippo transducers play crucial roles in breast cancer formation, progression, and dissemination and are therefore suggested as potential therapeutic targets [54]. Without using MI, DEGs identified by DESeq2 served as seeds for computing the network diffusion in the PPI network. Subsequently, the top 99th percentile of the results was selected for the enrichment analysis, yielding 37 GO and KEGG pathways (Table S5).

Figure 4B presents the top 20 GO pathways sorted by gene ratio. These pathways, when compared to those identified using MI, are distinct and free of duplicates. All of the enriched pathways have very low gene ratios. The pathway with the highest gene ratio, close to 0.225, is calcium-dependent cell–cell adhesion via plasma. The absence of MI results in a variety of genes in different pathways, which consequently provides a broader seed set. Notice that this also leads to higher diffusion scores across nodes in different components, resulting in multiple top-ranked genes within the subgraphs of the PPI network, as illustrated in Figure S3. These subgraphs exhibit lower connectivity compared to the single component obtained with the seeds from DESeq2 and MI, as shown in Figure 5. Nodes that are well connected in the PPI network typically share similar functions and contribute more effectively to the enrichment analysis. Thus, incorporating MI yields a higher gene ratio (up to >0.8 , as shown in Figure 4A), underscoring its significance in enhancing the analysis outcomes.

4. Discussion and Conclusions

In this study, we demonstrate using TempO-Seq to extract transcriptomic data on cancer cells in breast cancer subtypes to identify key biomarkers associated with high levels of TB, indicative of disease pathogenesis and progression. TempO-Seq has transformed gene expression analysis by directly measuring the activity in tissue samples without requiring RNA extraction. This technique offers exceptional precision and sensitivity [7].

Although FFPE samples typically yield fewer detected genes compared to fresh tissues, improvements have been made over time [6,55]. While RNA-Seq can provide broader gene detection and has become a standard method in many studies, it is often more prone to noise [5]. In breast cancer research, TempO-Seq was used to identify DEGs in tumors with low versus high cytoplasmic CAIX, highlighting the effective use of TempO-Seq [56].

The challenges posed by small sample sizes in cancer research are well recognized, particularly in the context of detecting differential gene expression. Traditional methods often rely on statistical tests that may lack robustness with limited sample sizes [57]. To address this, we employed mutual information techniques, which are adept at identifying features or genes differentiating two groups of small samples, such as low and high TB, as in our case. With our small sample sizes, mutual information proves more robust than traditional statistical tests, as it does not rely on assumptions about the underlying data distribution. Additionally, our study demonstrates that using mutual information is crucial for screening significant genes in network diffusion and functional enrichment analyses. While single methods like DESeq2 may detect numerous genes, including non-relevant ones, this can result in network diffusion valuing various pathways less important to high-tumor-budding breast cancer. Moreover, a larger number of seed genes in the diffusion technique can lead to diverse clusters in the PPI network, which may correspond to different functions and may be less suitable for enrichment analysis.

Combining significant differentially expressed genes (DEGs) and important genes identified by mutual information provides better potential biomarkers than either approach alone. Subsequently, we identified seven potential genes related to tumor budding in breast cancer (*NOL4*, *STAR*, *C8G*, *NEIL1*, *SLC46A3*, *FRMD6*, and *SCARF2*).

Each gene found in all breast cancer subtypes with high TB plays a different role in cellular processes. Firstly, *NEIL1* (Nei Like DNA Glycosylase 1) is involved in the DNA repair process, specifically in the base excision repair pathway, which repairs damaged DNA bases [29]. Dysregulation of the DNA repair process leads to genomic instability. Consistent with this study, previous studies showed that the expression of *NEIL1* was up-regulated in breast and lung cancer cell lines, and *NEIL1* silencing resulted in cancer cell apoptosis by increasing the pro-apoptotic protein (Bax) and decreasing the anti-apoptotic protein (Bcl-2) [30,31]. These results suggest that *NEIL1* helps maintain genomic stability, leading to a reduction in cancer cell death and eventually promoting the progression of breast and lung cancer.

SCARF2 (Scavenger Receptor Class F Member 2) is a member of the scavenger receptor family involved in various cellular processes, including endocytosis, lipid metabolism, and cell adhesion [32]. It functions as a receptor for various ligands, including modified lipoproteins and extracellular matrix proteins [32]. Meanwhile, *NOL4* (Nucleolar Protein 4) is a nucleolar protein involved in ribosome biogenesis and RNA processing [34]. It plays a role in the assembly and maturation of ribosomes, the cellular machinery responsible for protein synthesis [34]. It was reported that *SCARF2* and *NOL4* were up-regulated in various types of cancers [33,35–38], and they promoted tumor progression and metastasis [33,35–38]. However, direct evidence linking *SCARF2* or *NOL4* to breast cancer is limited. Further research is needed to elucidate the specific mechanisms by which *SCARF2* and *NOL4* contribute to breast cancer pathogenesis.

Moreover, in this study, we found that Steroidogenic Acute Regulatory Protein (*STAR*) and complement C8 gamma chain (*C8G*) were down-regulated. *STAR* plays a crucial role in steroid hormone synthesis, particularly in transporting cholesterol into the mitochondria for steroidogenesis [39], and the increased expression of *STAR* promotes the development and progression of certain types of breast cancer, particularly hormone receptor-positive breast cancer [40–42]. Interestingly, *C8G* is responsible for stabilizing the interaction between the complement components C8 α and C8 β and facilitating their assembly into the complement membrane attack complex (MAC) [43]. Once the MAC is formed, it inserts into the membrane of the target cell, disrupting its integrity and ultimately causing cell lysis [43]. We found that in all the breast cancer subtypes with high TB, the expression

of the *C8G* gene was down-regulated, suggesting that cell lysis is suppressed and may contribute to breast cancer progression. However, a previous study reported that *C8G* expression was increased in the exosome of rectal cancer patients who responded well to neoadjuvant therapy [58]. Therefore, *C8G* might have specific function in each type of cancer.

FERM Domain-Containing 6 (*FRMD6*), also known as Willin, is an upstream regulator of the Hippo signaling pathway that has recently been shown to modulate actin cytoskeleton dynamics and the mechanical phenotype of neuronal cells [44]. *FRMD6* is reported to exert both tumor-suppressive [45–47] and tumor-promoting [59,60] effects in various types of cancers, but still no evidence has been reported on breast cancer. However, in this study, we found that *FRMD6* gene expression was up-regulated, suggesting *FRMD6* might exhibit tumor-promoting effects in this scenario.

SLC46A3 (solute carrier family 46 member 3) belongs to the SLC46 family of solute carriers, which are responsible for transporting folate derivatives into cells [44]. Several studies have reported the association of *SLC46A3* with breast cancer [44,48–50]. Loss of *SLC46A3* was reported to increase the resistance of HER2+ breast cancer to trastuzumab emtansine [44,48–50], and the efficacy of trastuzumab emtansine was reduced when *SLC46A3* was inhibited [44,48–50]. Apart from breast cancer, a low level of this gene was related to the aggressive type of hepatocellular carcinoma [61], indicating that *SLC46A3* is an important molecule that exhibits anti-tumoral effects.

Further investigation based on protein–protein interaction networks and network diffusion was performed to elucidate the specific mechanisms by which these genes contribute to breast cancer pathogenesis. This revealed enrichment in the Hippo signaling and Wnt signaling pathways, which are known to promote tumor initiation, invasion, and metastasis in several cancer types. Additionally, the use of network diffusion techniques allowed us to uncover additional proteins closely related to our identified biomarkers, thereby expanding our understanding of the functional pathways implicated in breast cancer development and progression. This integrative approach enables the identification of interconnected signaling pathways and regulatory networks that drive oncogenesis, offering new avenues for therapeutic intervention. The crucial issue with this approach is the selection of the initial seed genes or proteins in the PPI network, which iteratively distribute their scores to related partners. It is important to consider both the number of seeds and their potential significance as representatives of the specific interest, which in our case is high tumor budding in breast cancer. We demonstrated that the enrichment results, with small seed sets highly related to breast cancer, identified by DESeq2 with mutual information, provided reasonable functional relevance to cancer and more robust clusters in the PPI network.

Taken together, our study demonstrates the utility of Tempo-Seq in uncovering the dynamic gene expression changes associated with breast cancer and highlights the potential of integrated analyses to elucidate the complex regulatory networks underlying cancer pathogenesis and identify novel therapeutic targets. Our findings provide insights into cancer biology and personalized treatment approaches, emphasizing the need for further investigation into the mechanisms underlying the involvement of *NOLA*, *STAR*, *C8G*, *NEIL1*, *SLC46A3*, *FRMD6*, and *SCARF2* in breast cancer pathogenesis. This presents opportunities for the development of novel therapeutic strategies and personalized treatment approaches tailored to individual patients.

In clinical data and gene expression data analysis, small sample sizes pose significant challenges that can impact the validity and reliability of the results. In our case, we have only 15 breast cancer samples (12 samples in the high-TB group vs. 3 samples in the low-TB group). To enhance the reliability of our analysis, DESeq2 and mutual information were applied, and network diffusion analysis of the PPI network was performed to extract potential biomarkers and their functional pathways. While our results provide initial insights into the gene expression changes associated with high-TB cases, the small sample size necessitates caution in the interpretation. Further experiments and analyses with

larger, more comprehensive studies are required to validate these findings and explore their clinical implications.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/biom14080896/s1>. Table S1: Statistical information on significant DEGs; Table S2: Mutual information score for each gene [62–86]; Table S3: The biological and tumor-related functions of the potential biomarkers; Table S4: GO and KEGG pathway enrichment results; Table S5: GO and KEGG pathway enrichment results without using MI; Figure S1: Histogram of MI score; Figure S2: Normalized count of genes with the highest MI score; Figure S3: The PPI subnetwork containing nodes with diffusion scores rank above the 99th percentile without using MI.

Author Contributions: Conceptualization, P.J., A.S., K.P., P.T. and C.T. Data curation, P.J., K.S., M.W., P.T. and C.T. Formal analysis, P.J., A.S. and K.P. Funding acquisition, A.S. and C.T. Methodology, P.J., A.S. and K.P. Supervision, A.S., K.P. and C.T. Visualization, P.J., A.S. and K.P. Writing—original draft preparation, P.J. Writing—review and editing, P.J., K.S., M.W., P.T., A.S., K.P. and C.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a Mid-Career Research Grant (N42A650343) funded by the National Research Council of Thailand and Mahidol University to C.T. Apichat Surataneer was funded by the National Science, Research and Innovation Fund (NSRF) and King Mongkut's University of Technology North Bangkok with contract no. KMUTNB-FF-66-08.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki, the Belmont Report, CIOMS Guidelines, and the International Conference on Harmonization in Good Clinical Practice (ICH-GCP) and approved by Siriraj Institutional Review Board (SIRB Protocol No.: 277/2565(IRB4); COA no. Si 542/2022).

Informed Consent Statement: Informed consent was obtained from all the subjects involved in the study.

Data Availability Statement: This article, along with its Supplementary Materials files, contains all the data generated or analyzed during this investigation, which are made available to the readers. Our gene expression dataset and its associated FASTQ files are publicly available in the NCBI Gene Expression Omnibus at <https://www.ncbi.nlm.nih.gov/geo/> (accessed on 30 March 2024), under accession number GSE262825.

Acknowledgments: The authors thank Surat Phumphueng for the breast cancer clinical specimen and data collection. P.J. is supported by the Science Achievement Scholarship of Thailand (SAST).

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

References

1. Hong, R.; Xu, B. Breast cancer: An up-to-date review and future perspectives. *Cancer Commun.* **2022**, *42*, 913–936. [[CrossRef](#)] [[PubMed](#)]
2. Huang, T.; Bao, H.; Meng, Y.-H.; Zhu, J.-L.; Chu, X.-D.; Chu, X.-L.; Pan, J.-H. Tumour budding is a novel marker in breast cancer: The clinical application and future prospects. *Ann. Med.* **2022**, *54*, 1303–1312. [[CrossRef](#)] [[PubMed](#)]
3. Voutsadakis, I.A. Prognostic role of tumor budding in breast cancer. *World J. Exp. Med.* **2018**, *8*, 12. [[CrossRef](#)] [[PubMed](#)]
4. Everett, L.J.; Mav, D.; Phadke, D.P.; Balik-Meisner, M.R.; Shah, R.R. Impact of Aligner, Normalization Method, and Sequencing Depth on TempO-seq Accuracy. *Bioinform. Biol. Insights* **2022**, *16*, 11779322221095216. [[CrossRef](#)] [[PubMed](#)]
5. Trejo, C.L.; Babić, M.; Imler, E.; Gonzalez, M.; Bibikov, S.I.; Shepard, P.J.; VanSteenhouse, H.C.; Yeakley, J.M.; Seligmann, B.E. Extraction-free whole transcriptome gene expression analysis of FFPE sections and histology-directed subareas of tissue. *PLoS ONE* **2019**, *14*, e0212031. [[CrossRef](#)] [[PubMed](#)]
6. Turnbull, A.K.; Selli, C.; Martinez-Perez, C.; Fernando, A.; Renshaw, L.; Keys, J.; Figueroa, J.D.; He, X.; Tanioka, M.; Munro, A.F. Unlocking the transcriptomic potential of formalin-fixed paraffin embedded clinical tissues: Comparison of gene expression profiling approaches. *BMC Bioinform.* **2020**, *21*, 30. [[CrossRef](#)] [[PubMed](#)]
7. Yeakley, J.M.; Shepard, P.J.; Goyena, D.E.; VanSteenhouse, H.C.; McComb, J.D.; Seligmann, B.E. A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling. *PLoS ONE* **2017**, *12*, e0178302. [[CrossRef](#)] [[PubMed](#)]
8. Bushel, P.R.; Ferguson, S.S.; Ramaiahgari, S.C.; Paules, R.S.; Auerbach, S.S. Comparison of normalization methods for analysis of TempO-Seq Targeted RNA sequencing data. *Front. Genet.* **2020**, *11*, 594. [[CrossRef](#)] [[PubMed](#)]
9. Vergara, J.R.; Estévez, P.A. A review of feature selection methods based on mutual information. *Neural Comput. Appl.* **2014**, *24*, 175–186. [[CrossRef](#)]

10. Song, L.; Langfelder, P.; Horvath, S. Comparison of co-expression measures: Mutual information, correlation, and model based indices. *BMC Bioinform.* **2012**, *13*, 328. [[CrossRef](#)]
11. Li, K.; Yu, M.; Liu, L.; Li, T.; Zhai, J. Feature selection method based on weighted mutual information for imbalanced data. *Int. J. Softw. Eng. Knowl. Eng.* **2018**, *28*, 1177–1194. [[CrossRef](#)]
12. Shukla, A.K.; Singh, P.; Vardhan, M. A new hybrid feature subset selection framework based on binary genetic algorithm and information theory. *Int. J. Comput. Intell. Appl.* **2019**, *18*, 1950020. [[CrossRef](#)]
13. Zhongxin, W.; Gang, S.; Jing, Z.; Jia, Z. Feature selection algorithm based on mutual information and lasso for microarray data. *Open Biotechnol. J.* **2016**, *10*, 278–286. [[CrossRef](#)]
14. Purbolaksono, M.D.; Widiastuti, K.C.; Mubarak, M.S.; Ma'ruf, F.A. Implementation of mutual information and bayes theorem for classification microarray data. *J. Phys. Conf. Ser.* **2018**, *971*, 012011. [[CrossRef](#)]
15. Wang, Y.; Wang, J.; Hu, Y.; Shangguan, J.; Song, Q.; Xu, J.; Wang, H.; Xue, M.; Wang, L.; Zhang, Y. Identification of key biomarkers for STAD using filter feature selection approaches. *Sci. Rep.* **2022**, *12*, 19854. [[CrossRef](#)]
16. Bersanelli, M.; Mosca, E.; Remondini, D.; Castellani, G.; Milanese, L. Network diffusion-based analysis of high-throughput data for the detection of differentially enriched modules. *Sci. Rep.* **2016**, *6*, 34841. [[CrossRef](#)]
17. Janyasupab, P.; Suratane, A.; Plaimas, K. Network diffusion with centrality measures to identify disease-related genes. *Math. Biosci. Eng.* **2021**, *18*, 2909–2929. [[CrossRef](#)]
18. Sagulkoo, P.; Suratane, A.; Plaimas, K. Immune-Related Protein Interaction Network in Severe COVID-19 Patients toward the Identification of Key Proteins and Drug Repurposing. *Biomolecules* **2022**, *12*, 690. [[CrossRef](#)]
19. Suratane, A.; Plaimas, K. DDA: A Novel Network-Based Scoring Method to Identify Disease-Disease Associations. *Bioinform. Biol. Insights* **2015**, *9*, 175–186. [[CrossRef](#)]
20. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)]
21. Sayers, E.W.; Beck, J.; Bolton, E.E.; Brister, J.R.; Chan, J.; Comeau, D.C.; Connor, R.; DiCuccio, M.; Farrell, C.M.; Feldgarden, M. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2024**, *52*, D33. [[CrossRef](#)] [[PubMed](#)]
22. Koelzer, V.H.; Zlobec, I.; Lugli, A. Tumor budding in colorectal cancer—ready for diagnostic practice? *Hum. Pathol.* **2016**, *47*, 4–19. [[CrossRef](#)] [[PubMed](#)]
23. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)] [[PubMed](#)]
24. Kursu, M.B. Praznik: High performance information-based feature selection. *SoftwareX* **2021**, *16*, 100819. [[CrossRef](#)]
25. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A.L.; Fang, T.; Doncheva, N.T.; Pyysalo, S. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **2023**, *51*, D638–D646. [[CrossRef](#)] [[PubMed](#)]
26. Smola, A.J.; Kondor, R. Kernels and regularization on graphs. In Proceedings of the Learning Theory and Kernel Machines: 16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, 24–27 August 2003; pp. 144–158.
27. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A.H.; Tanaseichuk, O.; Benner, C.; Chanda, S.K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **2019**, *10*, 1523. [[CrossRef](#)] [[PubMed](#)]
28. Beraha, M.; Metelli, A.M.; Papini, M.; Tirinzoni, A.; Restelli, M. Feature selection via mutual information: New theoretical insights. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–9.
29. Singh, P.K.; Mistry, K. Human NEIL1 DNA glycosylase: Structure, function and polymorphisms. *Meta Gene* **2017**, *11*, 49–57. [[CrossRef](#)]
30. Post, A.E.M.; Bussink, J.; Sweep, F.; Span, P.N. Changes in DNA Damage Repair Gene Expression and Cell Cycle Gene Expression Do Not Explain Radioresistance in Tamoxifen-Resistant Breast Cancer. *Oncol. Res.* **2020**, *28*, 33–40. [[CrossRef](#)] [[PubMed](#)]
31. Xue, W.; Liu, Y.; Xin, N.; Miao, J.; Du, J.; Wang, Y.; Shi, H.; Wei, Y.; Zhang, H.; Chen, Y.; et al. Nei Endonuclease VIII-Like1 (NEIL1) Inhibits Apoptosis of Human Colorectal Cancer Cells. *Biomed. Res. Int.* **2020**, *2020*, 5053975. [[CrossRef](#)]
32. Vo, T.T.; Kong, G.; Kim, C.; Juang, U.; Gwon, S.; Jung, W.; Nguyen, H.; Kim, S.H.; Park, J. Exploring scavenger receptor class F member 2 and the importance of scavenger receptor family in pre-diagnostic diseases. *Toxicol. Res.* **2023**, *39*, 341–353. [[CrossRef](#)]
33. Kim, C.; Kong, G.; Lee, H.; Tran, Q.; Vo, T.T.; Kwon, S.H.; Park, J.; Kim, S.H.; Park, J. Scavenger receptor class F member 2 (SCARF2) as a novel therapeutic target in glioblastoma. *Toxicol. Res.* **2022**, *38*, 249–256. [[CrossRef](#)] [[PubMed](#)]
34. Lin, F.; Zhou, J.; Li, X.; Wang, X. NOL4L, a novel nuclear protein, promotes cell proliferation and metastasis by enhancing the PI3K/AKT pathway in ovarian cancer. *Biochem. Biophys. Res. Commun.* **2021**, *559*, 121–128. [[CrossRef](#)] [[PubMed](#)]
35. Kim, Y.R.; Kim, K.U.; Lee, J.H.; Kim, D.W.; Chung, J.H.; Kim, Y.D.; Shin, D.H.; Lee, M.K.; Shin, Y.I.; Lee, S.Y. Cancer Testis Antigen, NOL4, Is an Immunogenic Antigen Specifically Expressed in Small-Cell Lung Cancer. *Curr. Oncol.* **2021**, *28*, 1927–1937. [[CrossRef](#)] [[PubMed](#)]
36. Lee, J.H.; Shin, D.H.; Lee, S.Y.; Park, J.Y.; Kim, S.Y.; Hwang, C.S.; Lee, H.J.; Na, J.Y.; Kim, J.Y. NOL4 is a novel nuclear marker of small cell carcinoma and other neuroendocrine neoplasms. *Histol. Histopathol.* **2022**, *37*, 1091–1098. [[CrossRef](#)] [[PubMed](#)]
37. Pavithran, G.; Rathi, B.; Santoshi, S. Translational and structural vaccinomics approach to design a multi-epitope vaccine against NOL4 autologous antigen of small cell lung cancer. *Immunol. Res.* **2023**, *71*, 909–928. [[CrossRef](#)]

38. Wei, Y.; Zheng, L.; Yang, X.; Luo, Y.; Yi, C.; Gou, H. Identification of Immune Subtypes and Candidate mRNA Vaccine Antigens in Small Cell Lung Cancer. *Oncologist* **2023**, *28*, e1052–e1064. [[CrossRef](#)] [[PubMed](#)]
39. Tugaeva, K.V.; Sluchanko, N.N. Steroidogenic Acute Regulatory Protein: Structure, Functioning, and Regulation. *Biochemistry* **2019**, *84*, S233–S253. [[CrossRef](#)]
40. Manna, P.R.; Ahmed, A.U.; Molehin, D.; Narasimhan, M.; Pruitt, K.; Reddy, P.H. Hormonal and Genetic Regulatory Events in Breast Cancer and Its Therapeutics: Importance of the Steroidogenic Acute Regulatory Protein. *Biomedicines* **2022**, *10*, 1313. [[CrossRef](#)]
41. Manna, P.R.; Ahmed, A.U.; Vartak, D.; Molehin, D.; Pruitt, K. Overexpression of the steroidogenic acute regulatory protein in breast cancer: Regulation by histone deacetylase inhibition. *Biochem. Biophys. Res. Commun.* **2019**, *509*, 476–482. [[CrossRef](#)]
42. Manna, P.R.; Ahmed, A.U.; Yang, S.; Narasimhan, M.; Cohen-Tannoudji, J.; Slominski, A.T.; Pruitt, K. Genomic Profiling of the Steroidogenic Acute Regulatory Protein in Breast Cancer: In Silico Assessments and a Mechanistic Perspective. *Cancers* **2019**, *11*, 623. [[CrossRef](#)]
43. Bayly-Jones, C.; Bubeck, D.; Dunstone, M.A. The mystery behind membrane insertion: A review of the complement membrane attack complex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2017**, *372*, 20160221. [[CrossRef](#)] [[PubMed](#)]
44. Pegram, M.D.; Miles, D.; Tsui, C.K.; Zong, Y. HER2-Overexpressing/Amplified Breast Cancer as a Testing Ground for Antibody-Drug Conjugate Drug Development in Solid Tumors. *Clin. Cancer Res.* **2020**, *26*, 775–786. [[CrossRef](#)] [[PubMed](#)]
45. Xu, Y.; Wang, K.; Yu, Q. FRMD6 inhibits human glioblastoma growth and progression by negatively regulating activity of receptor tyrosine kinases. *Oncotarget* **2016**, *7*, 70080–70091. [[CrossRef](#)] [[PubMed](#)]
46. Wang, W.; Zhao, C.; Quan, F.; Zhang, P.; Shao, Y.; Liu, L. FERM domain-containing protein 6 exerts a tumor-inhibiting role in thyroid cancer by antagonizing oncogenic YAP1. *Biofactors* **2022**, *48*, 428–441. [[CrossRef](#)] [[PubMed](#)]
47. von Koskull, A.; Hagström, J.; Haglund, C.; Kaprio, T.; Böckelman, C. High-tissue FRMD6 expression predicts better outcomes among colorectal cancer patients. *Biomarkers* **2024**, *29*, 127–133. [[CrossRef](#)] [[PubMed](#)]
48. Li, G.; Guo, J.; Shen, B.Q.; Yadav, D.B.; Sliwkowski, M.X.; Crocker, L.M.; Lacap, J.A.; Phillips, G.D.L. Mechanisms of Acquired Resistance to Trastuzumab Emtansine in Breast Cancer Cells. *Mol. Cancer Ther.* **2018**, *17*, 1441–1453. [[CrossRef](#)] [[PubMed](#)]
49. Tomabechi, R.; Kishimoto, H.; Sato, T.; Saito, N.; Kiyomiya, K.; Takada, T.; Higuchi, K.; Shirasaka, Y.; Inoue, K. SLC46A3 is a lysosomal proton-coupled steroid conjugate and bile acid transporter involved in transport of active catabolites of T-DM1. *Proc. Natl. Acad. Sci. Nexus* **2022**, *1*, pgac063. [[CrossRef](#)] [[PubMed](#)]
50. Kiyomiya, K.; Tomabechi, R.; Saito, N.; Watai, K.; Takada, T.; Shirasaka, Y.; Kishimoto, H.; Higuchi, K.; Inoue, K. Macrolide and Ketolide Antibiotics Inhibit the Cytotoxic Effect of Trastuzumab Emtansine in HER2-Positive Breast Cancer Cells: Implication of a Potential Drug-ADC Interaction in Cancer Chemotherapy. *Mol. Pharm.* **2023**, *20*, 6130–6139. [[CrossRef](#)]
51. Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2021**, *2*, 100141. [[CrossRef](#)]
52. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
53. Abdel-Fatah, T.M.; Perry, C.; Arora, A.; Thompson, N.; Doherty, R.; Moseley, P.M.; Green, A.R.; Chan, S.Y.; Ellis, I.O.; Madhusudan, S. Is there a role for base excision repair in estrogen/estrogen receptor-driven breast cancers? *Antioxid. Redox Signal.* **2014**, *21*, 2262–2268. [[CrossRef](#)] [[PubMed](#)]
54. Parambil, S.T.; Antony, G.R.; Littleflower, A.B.; Subhadradevi, L. The molecular crosstalk of the hippo cascade in breast cancer: A potential central susceptibility. *Biochimie* **2024**, *222*, 132–150. [[CrossRef](#)] [[PubMed](#)]
55. Linton, K.M.; Hey, Y.; Dibben, S.; Miller, C.J.; Freemont, A.J.; Radford, J.A.; Pepper, S.D. Methods comparison for high-resolution transcriptional analysis of archival material on Affymetrix Plus 2.0 and Exon 1.0 microarrays. *Biotechniques* **2009**, *47*, 587–596. [[CrossRef](#)] [[PubMed](#)]
56. Shamis, S.A. The Relationship between Hypoxia, Hypoxia Gene Signature and Survival in Patients with Breast Cancer. Ph.D. Thesis, University of Glasgow, Glasgow, UK, 2023.
57. Shahjaman, M.; Manir Hossain Mollah, M.; Rezanur Rahman, M.; Islam, S.M.S.; Nurul Haque Mollah, M. Robust identification of differentially expressed genes from RNA-seq data. *Genomics* **2020**, *112*, 2000–2010. [[CrossRef](#)] [[PubMed](#)]
58. Strybel, U.; Marczak, L.; Zeman, M.; Polanski, K.; Mielańczyk, Ł.; Klymenko, O.; Samelak-Czajka, A.; Jackowiak, P.; Smolarz, M.; Chekan, M.; et al. Molecular Composition of Serum Exosomes Could Discriminate Rectal Cancer Patients with Different Responses to Neoadjuvant Radiotherapy. *Cancers* **2022**, *14*, 993. [[CrossRef](#)] [[PubMed](#)]
59. Haldrup, J.; Strand, S.H.; Cieza-Borrella, C.; Jakobsson, M.E.; Riedel, M.; Norgaard, M.; Hedensted, S.; Dagnaes-Hansen, F.; Ulhøi, B.P.; Eeles, R.; et al. FRMD6 has tumor suppressor functions in prostate cancer. *Oncogene* **2021**, *40*, 763–776. [[CrossRef](#)] [[PubMed](#)]
60. Wang, T.; Guo, H.; Zhang, L.; Yu, M.; Li, Q.; Zhang, J.; Tang, Y.; Zhang, H.; Zhan, J. FERM domain-containing protein FRMD6 activates the mTOR signaling pathway and promotes lung cancer progression. *Front. Med.* **2023**, *17*, 714–728. [[CrossRef](#)] [[PubMed](#)]
61. Zhao, Q.; Zheng, B.; Meng, S.; Xu, Y.; Guo, J.; Chen, L.J.; Xiao, J.; Zhang, W.; Tan, Z.R.; Tang, J.; et al. Increased expression of SLC46A3 to oppose the progression of hepatocellular carcinoma and its effect on sorafenib therapy. *Biomed. Pharmacother.* **2019**, *114*, 108864. [[CrossRef](#)] [[PubMed](#)]
62. Sheikholeslami, S.; Azizi, F.; Ghasemi, A.; Alibakhshi, A.; Parsa, H.; Tavangar, S.M.; Shivaee, S.; Yeganeh, M.Z.; Hedayati, M.; Teimoori-Toolabi, L. NOL4 is Downregulated and Hyper-Methylated in Papillary Thyroid Carcinoma Suggesting Its Role as a Tumor Suppressor Gene. *Int. J. Endocrinol. Metab.* **2020**, *18*, 108510. [[CrossRef](#)]

63. Bayat, A.; Iqbal, S.; Borredy, K.; Amiel, J.; Zweier, C.; Barcia, G.; Kraus, C.; Weyhreter, H.; Bassuk, A.G.; Chopra, M.; et al. PRICKLE2 revisited—Further evidence implicating PRICKLE2 in neurodevelopmental disorders. *Eur. J. Hum. Genet.* **2021**, *29*, 1235–1244. [[CrossRef](#)]
64. Zhu, L.; Liu, Y.; Tang, H.; Wang, P. Circular RNA Circ_0001777 Suppresses Lung Adenocarcinoma Progression In Vitro and In Vivo. *Biochem. Genet.* **2023**, *61*, 704–724. [[CrossRef](#)]
65. Qian, H.; Cui, N.; Zhou, Q.; Zhang, S. Identification of miRNA biomarkers for stomach adenocarcinoma. *BMC Bioinform.* **2022**, *23*, 1–12. [[CrossRef](#)] [[PubMed](#)]
66. Kim, H.; Ryu, J.; Lee, C. Genome-wide identification of expression quantitative trait loci for human telomerase. *Medicine* **2016**, *95*, e5209. [[CrossRef](#)] [[PubMed](#)]
67. Aloe, L.; Rocco, M.L.; Balzamino, B.O.; Micera, A. Nerve growth factor: Role in growth, differentiation and controlling cancer cell development. *J. Exp. Clin. Cancer Res.* **2016**, *35*, 1–7. [[CrossRef](#)] [[PubMed](#)]
68. Fernández-Nogueira, P.; Bragado, P.; Almendro, V.; Ametller, E.; Rios, J.; Choudhury, S.; Mancino, M.; Gascón, P. Differential expression of neurogenes among breast cancer subtypes identifies high risk patients. *Oncotarget* **2015**, *7*, 5313–5326. [[CrossRef](#)] [[PubMed](#)]
69. Chakravarthy, R.; Mnich, K.; Gorman, A.M. Nerve growth factor (NGF)-mediated regulation of p75NTR expression contributes to chemotherapeutic resistance in triple negative breast cancer cells. *Biochem. Biophys. Res. Commun.* **2016**, *478*, 1541–1547. [[CrossRef](#)] [[PubMed](#)]
70. Wu, R.; Li, K.; Yuan, M.; Luo, K.Q. Nerve growth factor receptor increases the tumor growth and metastatic potential of triple-negative breast cancer cells. *Oncogene* **2021**, *40*, 2165–2181. [[CrossRef](#)] [[PubMed](#)]
71. Bashir, N.; Ishfaq, M.; Mazhar, K.; Khan, J.S.; Shahid, R. Upregulation of CD271 transcriptome in breast cancer promotes cell survival via NFκB pathway. *Mol. Biol. Rep.* **2021**, *49*, 487–495. [[CrossRef](#)] [[PubMed](#)]
72. Hasan, N.; Pangen, G.; Ray, T.A.; Fransen, K.M.; Noel, J.; Borghuis, B.G.; McCall, M.A.; Gregg, R.G. LRIT3 is Required for Nyctalopin Expression and Normal ON and OFF Pathway Signaling in the Retina. *eNeuro* **2020**, *7*. [[CrossRef](#)]
73. Tadini-Buoninsegni, F.; Smeazzetto, S. Mechanisms of charge transfer in human copper ATPases ATP7A and ATP7B. *IUBMB Life* **2017**, *69*, 218–225. [[CrossRef](#)]
74. Shi, B.; Zhang, W.; Wang, T.; Cui, Z. The therapeutic and prognostic role of cuproptosis-related genes in triple negative breast cancer. *BMC Bioinform.* **2023**, *24*, 1–15. [[CrossRef](#)] [[PubMed](#)]
75. Zhang, D.; Lu, W.; Zhuo, Z.; Wang, Y.; Zhang, W.; Zhang, M. Comprehensive analysis of a cuproptosis-related ceRNA network implicates a potential endocrine therapy resistance mechanism in ER-positive breast cancer. *BMC Med. Genom.* **2023**, *16*, 1–15. [[CrossRef](#)] [[PubMed](#)]
76. Yu, Z.; Cao, W.; Ren, Y.; Zhang, Q.; Liu, J. ATPase copper transporter A, negatively regulated by miR-148a-3p, contributes to cisplatin resistance in breast cancer cells. *Clin. Transl. Med.* **2020**, *10*, 57–73. [[CrossRef](#)] [[PubMed](#)]
77. Shanbhag, V.; Jasmer-McDonald, K.; Zhu, S.; Martin, A.L.; Gudekar, N.; Khan, A.; Ladomersky, E.; Singh, K.; Weisman, G.A.; Petris, M.J. ATP7A delivers copper to the lysyl oxidase family of enzymes and promotes tumorigenesis and metastasis. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 6836–6841. [[CrossRef](#)] [[PubMed](#)]
78. Hauge, H.; Fjelland, K.E.; Sioud, M.; Aasheim, H.-C. Evidence for the involvement of FAM110C protein in cell spreading and migration. *Cell. Signal.* **2009**, *21*, 1866–1873. [[CrossRef](#)] [[PubMed](#)]
79. Li, M.; Zhong, D.; Li, G. Regulatory role of local tissue signal Del-1 in cancer and inflammation: A review. *Cell. Mol. Biol. Lett.* **2021**, *26*, 1–12. [[CrossRef](#)] [[PubMed](#)]
80. Lee, S.J.; Lee, J.; Kim, W.W.; Jung, J.H.; Park, H.Y.; Park, J.-Y.; Chae, Y.S. Del-1 Expression as a Potential Biomarker in Triple-Negative Early Breast Cancer. *Oncology* **2018**, *94*, 243–256. [[CrossRef](#)] [[PubMed](#)]
81. Moon, P.-G.; Lee, J.-E.; Cho, Y.-E.; Lee, S.J.; Jung, J.H.; Chae, Y.S.; Bae, H.-I.; Kim, Y.-B.; Kim, I.-S.; Park, H.Y.; et al. Identification of Developmental Endothelial Locus-1 on Circulating Extracellular Vesicles as a Novel Biomarker for Early Breast Cancer Detection. *Clin. Cancer Res.* **2016**, *22*, 1757–1766. [[CrossRef](#)] [[PubMed](#)]
82. Lee, S.J.; Lee, J.; Jung, J.H.; Park, H.Y.; Moon, P.-G.; Chae, Y.S.; Baek, M.-C. Exosomal Del-1 as a Potent Diagnostic Marker for Breast Cancer: Prospective Cohort Study. *Clin. Breast Cancer* **2021**, *21*, E748–E756. [[CrossRef](#)]
83. Lee, J.; Jeong, J.-H.; Jung, J.H.; Kim, W.W.; Lee, S.J.; Park, J.-Y.; Park, J.Y.; Kang, S.H.; Kim, E.A.; Park, J.H.; et al. Overcoming Tamoxifen Resistance by Regulation of Del-1 in Breast Cancer. *Oncology* **2019**, *97*, 180–188. [[CrossRef](#)]
84. Nommick, A.; Boutin, C.; Rosnet, O.; Schirmer, C.; Bazellieres, E.; Thomé, V.; Loiseau, E.; Viallat, A.; Kodjabachian, L. Lrrcc1 and Ccdc61 are conserved effectors of multiciliated cell function. *J. Cell Sci.* **2022**, *135*, 258960. [[CrossRef](#)] [[PubMed](#)]
85. Chen, D.; Yu, W.; Aitken, L.; Gunn-Moore, F. Willin/FRMD6: A Multi-Functional Neuronal Protein Associated with Alzheimer's Disease. *Cells* **2021**, *10*, 3024. [[CrossRef](#)] [[PubMed](#)]
86. Sun, M.; Shen, Y.; Jia, G.; Deng, Z.; Shi, F.; Jing, Y.; Xia, S. Activation of the HNRNPA2B1/miR-93-5p/FRMD6 axis facilitates prostate cancer progression in an m6A-dependent manner. *J. Cancer* **2023**, *14*, 1242–1256. [[CrossRef](#)] [[PubMed](#)]

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.