



Article

# Identification of Prognostic Candidate Genes in Breast Cancer by Integrated Bioinformatic Analysis

Charles C.N. Wang <sup>1,†</sup>, Chia Ying Li <sup>2,3,†</sup>, Jia-Hua Cai <sup>1</sup>, Phillip C.-Y. Sheu <sup>4</sup>, Jeffrey J.P. Tsai <sup>1</sup>, Meng-Yu Wu <sup>5,6</sup>, Chia-Jung Li <sup>7,8,\*</sup> and Ming-Feng Hou <sup>9,10,11,12,\*</sup>

- <sup>1</sup> Department of Bioinformatics and Medical Engineering, Asia University, Taichung 413, Taiwan
  - <sup>2</sup> Department of Surgery, Show Chwan Memorial Hospital, Changhua 500, Taiwan
  - <sup>3</sup> Graduate Institute of Biomedical Engineering, National Chung Hsing University, Taichung 402, Taiwan
  - <sup>4</sup> Department of EECS and BME, University of California, Irvine, CA 92697, USA
  - <sup>5</sup> Department of Emergency Medicine, Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, New Taipei 231, Taiwan
  - <sup>6</sup> Department of Emergency Medicine, School of Medicine, Tzu Chi University, Hualien 970, Taiwan
  - <sup>7</sup> Department of Obstetrics and Gynecology, Kaohsiung Veterans General Hospital, Kaohsiung 813, Taiwan
  - <sup>8</sup> Institute of Biomedical Sciences, National Sun Yat-Sen University, Kaohsiung 804, Taiwan
  - <sup>9</sup> Division of Breast Surgery, Department of Surgery; Center for Cancer Research, Kaohsiung Medical University Chung-Ho Memorial Hospital, Kaohsiung 807, Taiwan
  - <sup>10</sup> Graduate Institute of Clinical Medicine, Kaohsiung Medical University, Kaohsiung 807, Taiwan
  - <sup>11</sup> National Sun Yat-Sen University-Kaohsiung Medical University Joint Research Center, Kaohsiung Medical University, Kaohsiung 807, Taiwan
  - <sup>12</sup> National Chiao Tung University-Kaohsiung Medical University Joint Research Center, Kaohsiung Medical University, Kaohsiung 807, Taiwan
- \* Correspondence: nigel6761@gmail.com (C.-J.L.); mifeho@kmu.edu.tw (M.-F.H.);  
Tel.: +886-7-342-2121 (ext. 4048) (C.-J.L.); +886-7-312-1101 (ext. 6060) (M.-F.H.)
- † These authors contributed equally to this paper.

Received: 25 June 2019; Accepted: 31 July 2019; Published: 2 August 2019



**Abstract:** Breast cancer is one of the most common malignancies. However, the molecular mechanisms underlying its pathogenesis remain to be elucidated. The present study aimed to identify the potential prognostic marker genes associated with the progression of breast cancer. Weighted gene coexpression network analysis was used to construct free-scale gene coexpression networks, evaluate the associations between the gene sets and clinical features, and identify candidate biomarkers. The gene expression profiles of GSE48213 were selected from the Gene Expression Omnibus database. RNA-seq data and clinical information on breast cancer from The Cancer Genome Atlas were used for validation. Four modules were identified from the gene coexpression network, one of which was found to be significantly associated with patient survival time. The expression status of 28 genes formed the black module (basal); 18 genes, dark red module (claudin-low); nine genes, brown module (luminal), and seven genes, midnight blue module (nonmalignant). These modules were clustered into two groups according to significant difference in survival time between the groups. Therefore, based on betweenness centrality, we identified *TXN* and *ANXA2* in the nonmalignant module, *TPM4* and *LOXL2* in the luminal module, *TPRN* and *ADCY6* in the claudin-low module, and *TUBA1C* and *CMIP* in the basal module as the genes with the highest betweenness, suggesting that they play a central role in information transfer in the network. In the present study, eight candidate biomarkers were identified for further basic and advanced understanding of the molecular pathogenesis of breast cancer by using co-expression network analysis.

**Keywords:** breast cancer; weighted gene coexpression network analysis; prognosis; GEO; TCGA

## 1. Introduction

The incidence of breast cancer continues to increase worldwide, and currently, breast cancer is a serious disease among women. In 2017, the estimated number of new cases of invasive breast cancer was 252,710, among which 2470 men were diagnosed. In addition, 63,410 women were diagnosed with breast cancer in situ, and around 40,610 women and 460 men were expected to die of breast cancer. Presently, Asia has become a high-risk region for breast cancer, ranking first among malignant tumors in females [1,2]. Recently, with the continuous efforts and progress of modern medicine, the treatment of breast cancer has become more effective, and the mortality rate of breast cancer has been significantly reduced. However, the recurrence and metastasis of breast cancer have still not been addressed comprehensively, and have become the greatest challenges in clinical treatment [3,4].

Researchers are using genetic studies to understand the functions of tumor-related genes and roles of tumor cell signaling pathways. Therefore, we tried to understand and screen the key genes related to breast cancer recurrence to predict the recurrence factors for breast cancer accurately, in order to reduce the risk of cancer recurrence and metastasis. In addition, patients can be provided with more effective and individualized treatment, which in turn reduces breast cancer mortality and improves long-term survival.

Both bioinformatics and system biology are strong interdisciplinary fields that combine the collection, storage, processing, and dissemination of biological information, summarize life sciences and computer science, and collect and analyze genetic data [5,6].

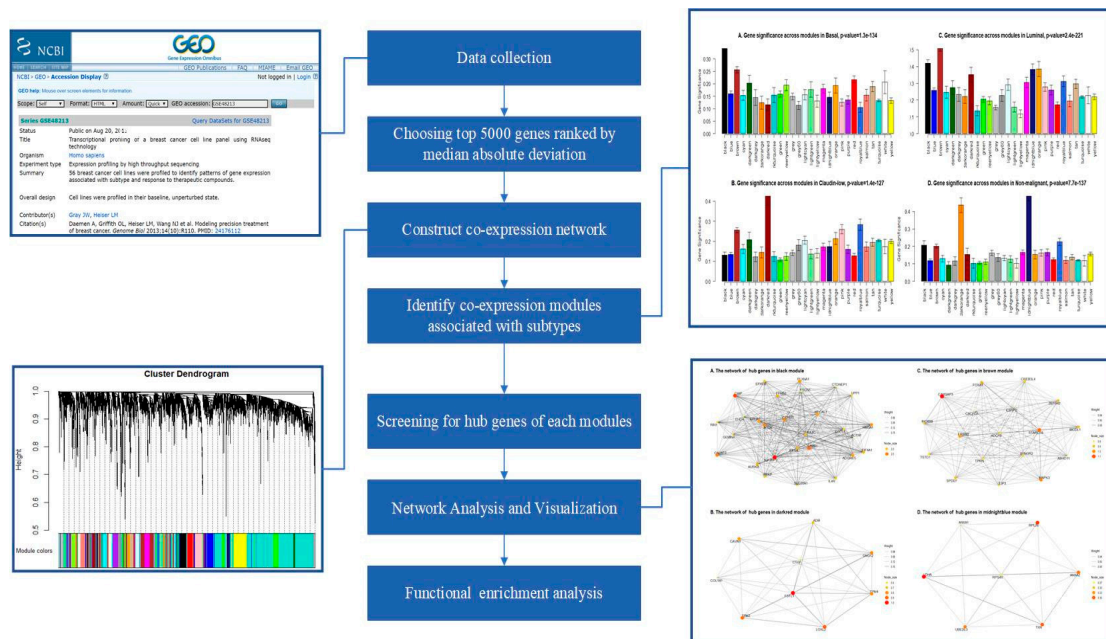
The rapid technological breakthrough of genome-wide sequencing has provided a new perspective for the study of clinical issues and related pathological mechanisms underlying various cancers. Most studies have focused on the screening of differentially expressed genes, but not enough to understand the high degree of interconnectivity between genes, i.e., genes with similar expression patterns may be functionally related [7]. In this study, we used coexpression analysis as a powerful technique for constructing free-scale gene coexpression networks. Weighted gene coexpression network analysis has been widely used to analyze large-scale data sets and find modules of highly correlated genes. Weighted gene correlation network analysis (WGCNA) [8] has been successfully used to evaluate the associations between gene sets and clinical features and identify candidate biomarkers. Therefore, we described the correlation patterns among genes using a systems biology method based on WGCNA and identified novel biomarkers associated with breast cancer prognosis. We used WGCNA to examine a previously published dataset (GSE48213, [9]). This dataset is a transcriptional profile of a breast cancer cell line based on 56 samples used to identify the gene expression patterns associated with subtype and response to therapeutic compounds.

Finally, we used three databases, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology, and The Cancer Genome Atlas (TCGA), and completely constructed related genes and coexpression networks. We also identified key genes to facilitate further studies into their potential as biomarkers as well as therapeutic targets in breast cancer. We performed a comprehensive bioinformatics analysis to further investigate the functions, pathways, regulatory mechanisms, and drug candidates of these genes.

## 2. Materials and Methods

### 2.1. Dataset Download

The pipeline of this study is presented in Figure 1. The gene expression profiles of GSE48213 were stored in the Gene Expression Omnibus (GEO) database. It is a transcriptional profiling of a breast cancer cell line based on 56 samples associated with subtypes (basal, claudin-low, luminal, nonmalignant, and unknown).



**Figure 1.** Flowchart of the study’s design, illustrating data preparation, preprocessing, and analysis.

### 2.2. Data Preprocessing

Firstly, normalized data were downloaded from the GEO database, and then we converted each Ensembl gene to Ensembl Gene ID with the biomaRt package [10,11] in R. However, a large number of these genes did not have different expression between samples. Therefore, the data was processed by focusing on the 4655 most variant genes for WGCNA analysis by using a robust method called median absolute deviation (MAD) [12]. The remaining genes were not used for analysis, because those showed no or very low changes in expression between different samples.

### 2.3. Gene Coexpression Network Construction

The expression data profiles of these 4655 genes were constructed to a scale-free coexpression network using the WGCNA package in R [8,13].

The adjacency coefficient  $a_{ij}$  was calculated as follows:

$$s_{ij} = |cor(x_i, x_j)| a_{ij} = S_{ij}^{\beta}$$

where  $x_i$  and  $x_j$  are vectors of expression value for gene  $i$  and  $j$ .  $s_{ij}$  represents the Pearson’s correlation coefficient of the two vectors.  $a_{ij}$  is the adjacency coefficient between gene  $i$  and gene  $j$ . In the presented study, the power of  $\beta = 6$  (scale free  $R^2 = 0.88$ ) was selected as the soft-thresholding parameter to ensure a scale-free network. In the coexpression network, genes with high absolute correlations were clustered into the same module. The WGCNA method not only considers the association between two connected genes, but also takes associated genes into account. Modules were also identified via hierarchical clustering of the weighting coefficient matrix. To further identify functional modules in the coexpression network with these 4655 genes, the topological overlap measure, representing the overlap in shared neighbors, was calculated using the adjacency coefficient. It calculates the weighting coefficient  $W_{ij}$  from  $a_{ij}$  as follows:

$$W_{ij} = \frac{\sum_{k=1}^N a_{ik} \cdot a_{kj} + a_{ij}}{\min(k_i, k_j) + 1 - a_{ij}}$$

where  $a$  is the weighted adjacency coefficient given  $a_{ij}$  and  $\beta = 6$  is the soft thresholding power. According to the  $W_{ij}$  measure with a minimum size (gene group) of 30 for the gene dendrogram, average linkage hierarchical clustering was conducted, and genes with similar expression profiles were classified into the same gene modules using the DynamicTreeCut algorithm.

In this study, an appropriate soft threshold power (soft power = 6) was selected to construct a standard scale-free network using the pickSoftThreshold function [14,15]. Then, the network construction and module detection were performed by the one-step function “blockwiseModules” [15]. It can handle the splitting into blocks automatically, and just needed to specify the largest number of genes that can fit in a block. The R function of blockwiseModules has many parameters; parameters were implemented with the following: power = 6, maxBlockSize = 6000, minModuleSize = 30, and networkType = “unsigned” in our study.

#### 2.4. Identification of Clinically Significant Modules

To identify key modules that are significantly associated with clinical subtypes of breast cancer, the expression profiles of each module were summarized by the module eigengene (ME) and the correlation between the module and clinical subtypes was calculated. The associations of individual genes with each subtype (basal, claudin-low, luminal, and nonmalignant) were quantified by Gene Significance (GS) value. Module significance (MS) refers to the average GS of all genes in the module. Modules with an absolute MS ranking first or second in all modules are considered candidates relevant to clinical traits. Finally, the module highly correlated with clinical subtypes of breast cancer was selected for further analysis.

#### 2.5. Candidate Hub Genes Identification

Hub genes in the coexpression network were highly interconnected with nodes in a module, and they have been shown to have significant function. In this study, the connectivity of genes was measured by the absolute value of the Pearson’s correlation (module membership > 0.8). The module membership (MM) was defined as the correlation of gene expression profile with module eigengenes (MEs). Also, we identified hub genes in modules that were highly related to certain clinical traits, which were measured by the absolute value of the Pearson’s correlation. Thus, the intramodular hub genes were chosen based on GS (gene significance) > 0.2 and MM > 0.8. The common hub genes in the coexpression networks were regarded as “real” hub genes for further analysis.

#### 2.6. Network Analysis and Visualization

Gene coexpression modules identified by clustering are often large; therefore, it is important to identify which gene in each module best explains its behavior. A widely used approach is to identify highly connected genes in a coexpression network, called hub genes [16]. In this study, we identified hub genes by betweenness centrality. Genes with high betweenness centrality are important as shortest-path connectors through a network [17]. Connectivity is often used to measure network robustness and indicates how many genes need to be removed from the network before the remaining genes are disconnected.

The hub genes detected by WGCNA can be analyzed with other tools. We performed network analysis by using the R package tidygraph (Version: 1.1.2; CRAN: MIT, MA, USA). The node centrality in networks is useful to detect genes with important functional roles. Specifically, the weighted degree of a node was used to identify potential biologically meaningful genes based on betweenness centrality [18].

#### 2.7. Pathway Enrichment Analysis and Gene Ontology

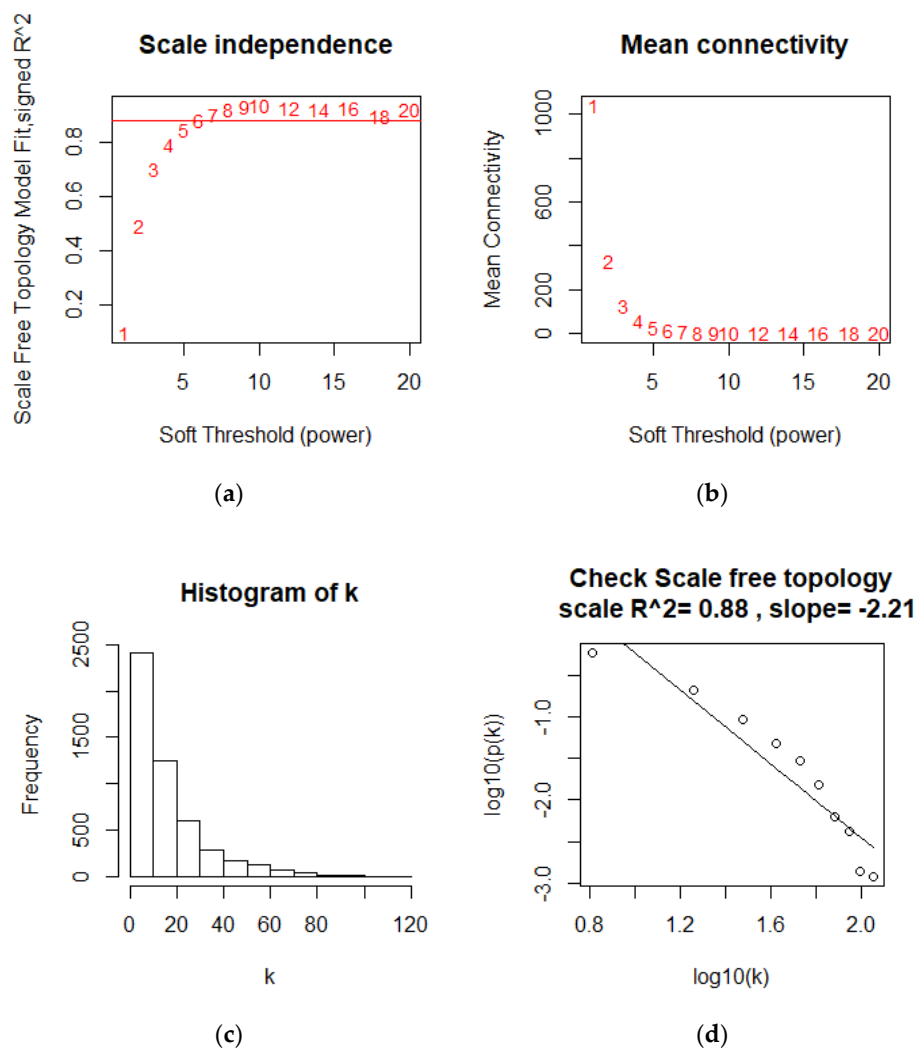
To investigate a comprehensive set of functionally annotated hub genes, Gene Ontology term enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed by using “FunRich”. FunRich is a functional enrichment and interaction network analysis

tool. GO enrichment analysis and KEGG pathway analysis were performed with the FunRich functional enrichment analysis tool (version 3.1.3; ExoCarta: Victoria, Australia). The top ten terms with  $p < 0.05$  and more than two enrichments were selected [19].

### 3. Results

#### 3.1. Weighted Coexpression Network Construction and Key Modules Identification

Fifty-six samples with subtype data were included in the coexpression analysis. The samples of GSE48213 were clustered using the average linkage hierarchical clustering method (Figure 1). In this study, the power of  $\beta = 6$  (scale-free  $R^2 = 0.88$ ) was selected as the soft thresholding parameter to ensure a scale-free network (Figure 2).



**Figure 2.** Determination of soft-thresholding power in WGCNA analysis. (a) Analysis of the scale-free fit index for various soft-thresholding powers  $\beta$ . (b) Analysis of the mean connectivity for various soft-thresholding powers. (c) Histogram of connectivity distribution when  $\beta = 6$ . (d) Checking the scale free topology when  $\beta = 6$ .

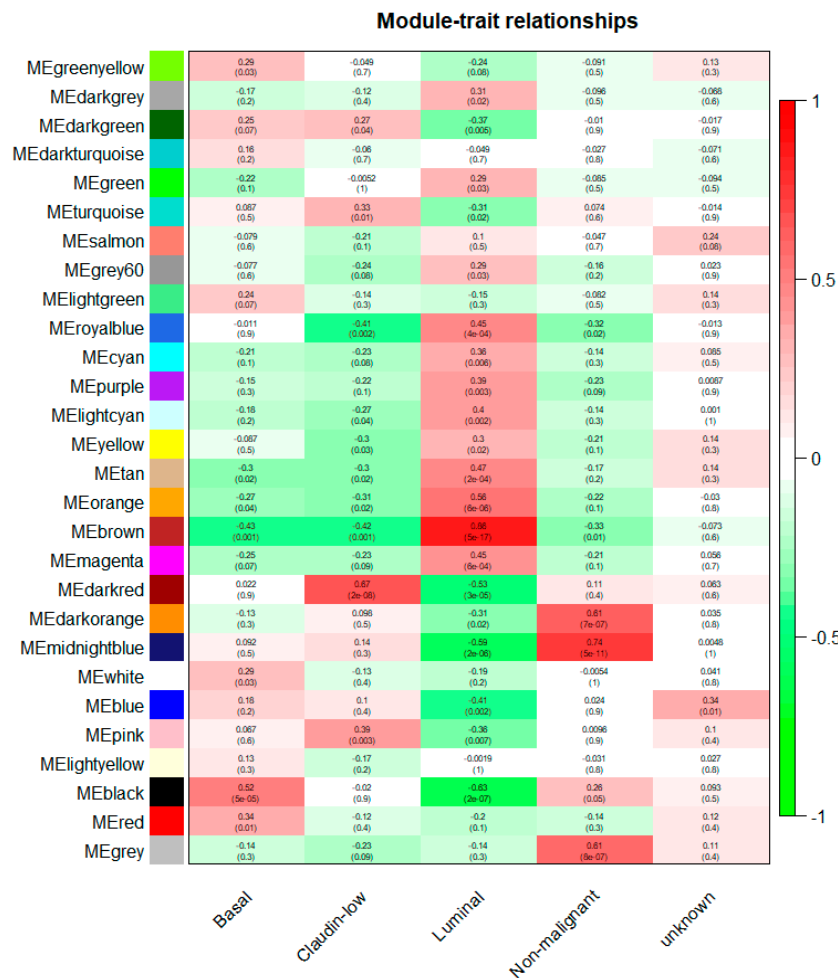
A total of 27 modules were identified based on average linkage hierarchical clustering (Figure 3a). Furthermore, module eigengenes (MEs) of the black, dark red, brown, and midnight blue modules were found to have the highest correlation with the subtypes (basal, claudin-low, luminal, and nonmalignant, respectively; Figures 3b and 4). These modules were selected as clinically significant modules for the identification of hub genes.





### 3.2. Identification of Candidate Genes with High Weighted Degree Score

Highly connected hub genes were defined by module connectivity (module membership > 0.8) and clinical trait relationship (gene significance > 0.2). Next, the hub genes of each module were visualized as networks by using the R package tidygraph, and the top candidate genes were screened and ordered in ranks by using weighted degree scores for further analysis (Figure 5).



**Figure 5.** Heatmap of the correlation between module eigengenes and clinical subtypes of breast cancer. Each column contained the corresponding correlation and *p* value.

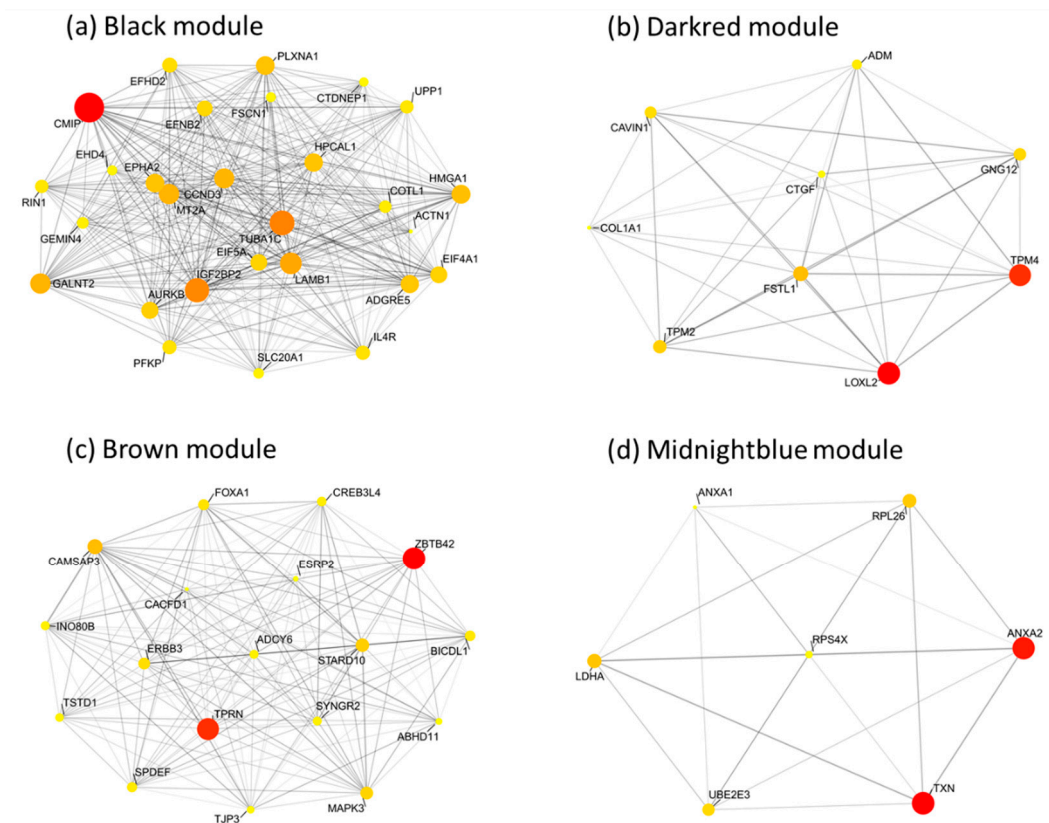
### 3.3. Pathway Enrichment Analysis and Gene Ontology

The genes in the clinically significant modules were categorized using the FunRich tool. Findings with higher scores are more significant than those with low scores. The GO database provides information on the associated biological processes, molecular functions, and cellular components. The GO analysis results revealed that the black module (basal) genes in the BP group were mainly enriched in cell communication, signal transduction, and cell growth and/or maintenance; the genes in the MF group were mainly enriched in molecular function unknown, translation regulator activity, and receptor activity; the genes in the CC group were significantly enriched in cytoplasm, exosomes, and nucleus. The brown module (luminal) was enriched in the BP, including regulation of nucleobase, nucleoside, nucleotide, and nucleic acid metabolism. The genes in the MF group were mainly enriched in molecular function unknown and transcription factor activity. The genes in the CC group were mainly enriched in cytoplasm, nucleus, and plasma membrane; the dark red module (claudin-low) genes in the BP group were mainly enriched in cell growth and maintenance. The genes in the MF group were

mainly enriched in structural constituent of cytoskeleton and extracellular matrix structural constituent. The genes in the CC group were enriched in extracellular, cytoplasm, and extracellular space; the midnight blue module (nonmalignant) genes in the BP were mainly enriched in protein metabolism. The gene in the MF was enriched in calcium ion binding. The genes in the CC group were enriched in centrosome, exosomes, cytosol, and nucleolus. The hub genes were also over-represented in these top KEGG pathways: the black module (basal) included the integrin family cell surface interactions and beta-1 integrin cell surface interactions; brown module (luminal), mesenchymal-to-epithelial transition; dark red module (claudin-low), signal transduction; and midnight blue module (nonmalignant), peptide chain elongation. Among the functional and pathway enrichment analyses, cell division and cell cycle were the most significantly enriched.

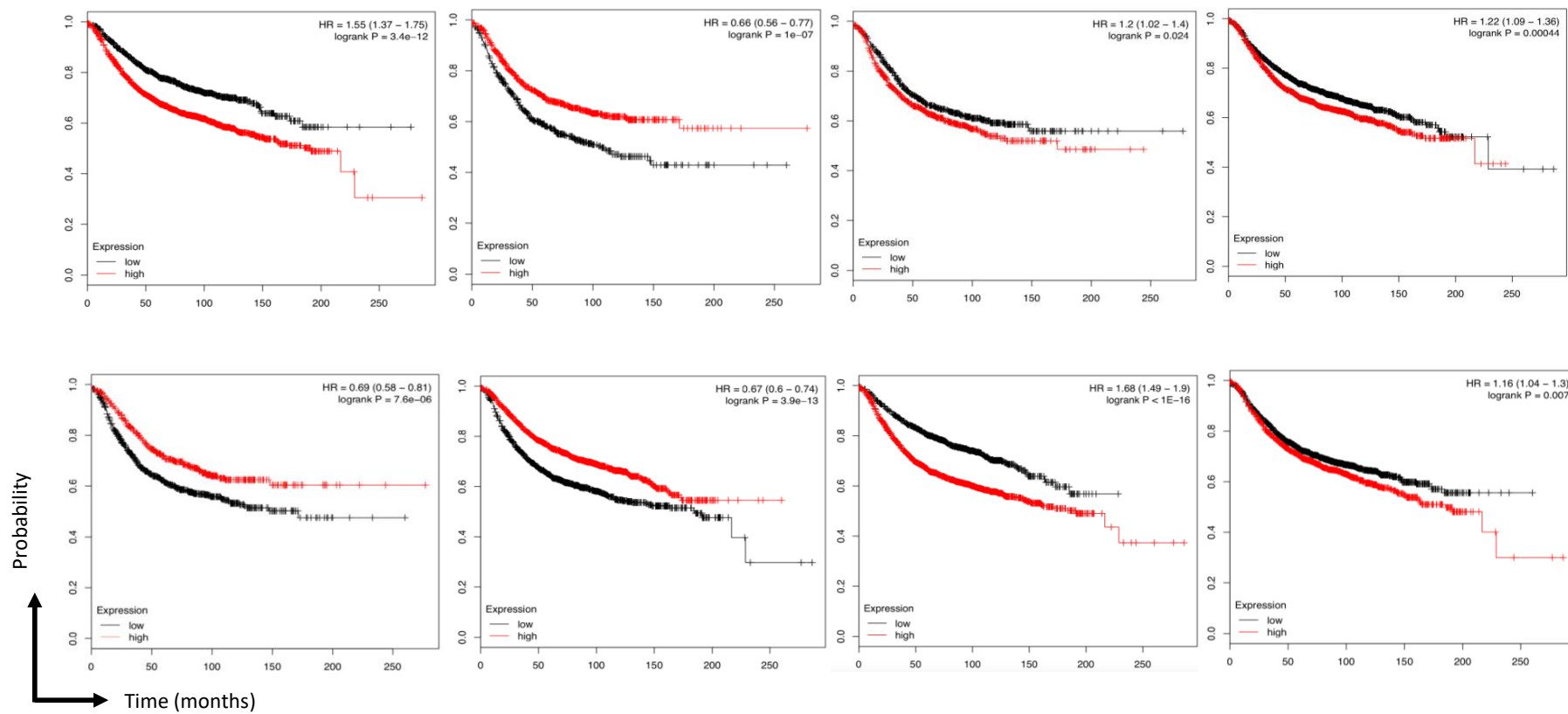
### 3.4. Identification and Validation of Hub Genes

Based on the cut-off criteria ( $|MM| > 0.8$  and  $|GS| > 0.2$ ), 42 genes with high connectivity in the four modules (including basal, luminal, claudin-low, and nonmalignant) were identified as hub genes, and centrality measures, mainly betweenness centrality, were used. Hub genes with high betweenness centrality are important as the shortest path connectors through a network. Connectivity is typically used to measure network robustness and indicates how many genes need to be removed from the network before the remaining genes are disconnected (Figure 6). Moreover, based on the TCGA data, the expression levels of these eight genes were significantly higher in breast cancer tissues. As tumor progression always affects tumor prognosis, we also validated the eight hub genes by investigating their roles in breast cancer prognosis, including overall survival time (Figure 7).



**Figure 6.** The network of hub genes in the (a) black module, (b) darkred module, (c) brown module, and (d) midnight blue module. Nodes represent genes and node size indicates weighted degree score. Edges are colored by weight.





**Figure 7.** Survival analysis of eight hub genes in breast cancers. Overall survival analysis *TXN*, *ANXA2*, *TPM4*, *LOXL2*, *TPRN*, *ADCY6*, *TUBA1C* and *CMIP*. (Red lines represent the samples with a highly expressed gene and the blue line represents samples with a lowly expressed gene. HR: hazard ratio).

#### 4. Discussion

The progression and prognosis of breast cancer are quite variable. Although many prognostic models have been proposed, most are based on clinical parameters and lack accuracy. In the era of precision medicine, there is an urgent need for better cancer-specific prognosis and progression biomarkers to provide accurate clinical information that could significantly enhance decision-making for patient management [20].

This study is based on the gene expression profile of breast cancer, and our main aim was to analyze the molecular mechanisms underlying breast cancer. We analyzed the data from both single- and multigene layers to find differential genes involved in breast cancer development.

##### (1) Construction and analysis of the gene coexpression network

First, we used WGCNA, which is commonly used to construct gene coexpression modules, to explain the mechanism of breast cancer on a multigene level [21,22]. The coexpression module aggregates genes with similar expression levels into one class, but the biological pathways involved in these modules and the changes in expression levels are unknown. To address these problems, we mainly performed differential correlation analysis, functional enrichment analysis, and differential gene correlation analysis of the modules [23]. First, the module was subjected to differential correlation analysis. The gene expression spectrum was used to analyze the differences in the average expression levels of the genes in the module. Simultaneously, displacement analysis was used to calculate the false discovery rate (FDR) and four modules with significant differences in expression were found [24,25]. Next, functional analysis of the significantly different modules was performed, and the biological significance of each module was analyzed to evaluate the biological processes and functional pathways involved in breast cancer. This study is based on the gene expression profile of breast cancer to analyze the molecular mechanism underlying breast cancer. We analyzed the data from both single- and multigene layers to find differential genes involved in breast cancer development.

##### (2) Screening of differential genes

In this study, we downloaded the counts for breast cancer RNA sequencing and fragments per kilobase per million mapped reads gene expression data from the TCGA cancer database [22,26]. By using the cluster analysis method to remove the outlier samples, genes were restricted to protein-encoding genes and the data were regularized for preprocessing and constructing an accurate gene expression profile. Next, the differential gene analysis method was used to screen differential genes, which directly revealed the relationship between gene expression changes and breast cancer development. Our results showed a high coincidence rate with the differentially expressed genes associated with breast cancer. Noncoincident parts included the genes *TXN*, *ANXA2*, *TPM4*, *LOXL2*, *TPRN*, *ADCY6*, *TUBA1C*, and *CMIP*.

WGCNA was performed to identify the gene coexpression modules associated with breast cancer progression. The modules were identified, and several central genes were obtained from the modules. Therefore, *TXN*, *ANXA2*, *TPM4*, *LOXL2*, *TPRN*, *ADCY6*, *TUBA1C*, and *CMIP* in the common networks may be the key genes, indicating that they are highly correlated with clinical features and important biological processes.

The eight genes predicted by us have been reported for other cancers, but not for breast cancer. *ANXA2* is a member of the annexin family. Members of this calcium-dependent phospholipid-binding protein family play a role in the regulation of cell growth and signal transduction pathways. *ANXA2* has been reported to be associated with cervical, gastric, colorectal, ovarian, and liver cancers [27,28].

*TPM4*, a gene that encodes a member of the tropomyosin family of actin-binding proteins, is involved in the contractile system of streak and smooth muscles as well as the cytoskeleton of nonmuscle cells. It has been reported to be associated with lung cancer metastasis and cervical and ovarian cancers [29–31].

*LOXL2* is a gene that encodes a member of the lysyl-oxidase gene family. Prototype members of this family are essential for the biosynthesis of connective tissue, which encodes an extracellular copper-dependent amine oxidase. *LOXL2* can also confer additional effects on developmental regulation, aging, tumor (prostate and liver cancer) inhibition, cell growth control, and chemotaxis to each member of the family [32,33].

The locus *TPRN* encodes a sensory epithelial protein. Mutations of this locus are associated with autosomal recessive deafness. To date, few cancer-related functions of *TPRN* have been reported. It has great potential as a biomarker for breast cancer.

*ADCY6* encodes a member of the adenylate cyclase protein family, and it is required for the synthesis of cAMP. *ADCY6* promotes increased phosphorylation of various proteins, including *AKT*. It plays a role in regulating  $\text{Ca}^{2+}$  uptake and storage in the cardiac sarcoplasmic reticulum and is required for normal ventricular contraction [34].

*TUBA1C*, a tubulin, is the main component of microtubules. It binds two molecules of GTP, one at the exchangeable site on the beta strand and the other at a nonexchangeable site on the alpha chain. The  $\alpha$ -2-HS-glycoprotein precursor and tubulin  $\beta$  chain may be involved in the pathogenesis of colorectal cancer and serve as potential biomarkers for serological diagnosis [35].

*CMIP* is a DNA-binding, leucine-containing transcription factor that acts as a homodimer or heterodimer. It plays a role in the T-cell signaling pathway. Depending on the binding site and binding partner, *CMIP* may encode a transcriptional activator or repressor protein. This protein plays a role in the regulation of several cellular processes, including cell development in embryonic lens, increased T-cell susceptibility to apoptosis, and chondrocyte terminal differentiation. *CMIP* has been reported to be associated with gastric cancer and glioma [36,37].

To further study the mechanisms underlying regulation of tumorigenesis, we performed GO and KEGG pathway and GSEA analyses. We found that the hub genes showed significant differences in cell membrane, cytoskeleton, immune cells, and ion regulation. KEGG analysis revealed that cell structure is the most important pathway (Table 1). Interestingly, we found that the central genes are often rich in cellular structures and ion signaling pathways (Supplementary Figure S1). Previous studies have demonstrated that calcium regulates the cytoskeleton, and thus promotes the migration and metastasis of cancer cells [38]. Therefore, the eight hub genes may be involved in the progression of breast cancer and may influence its prognosis through calcium-regulated immune and molecular signaling pathways, contributing to the poor prognosis of breast cancer.

**Table 1.** The KEGG pathways in the enrichment analysis of *TXN*, *ANXA2*, *TPM4*, *LOXL2*, *TPRN*, *ADCY6*, *TUBA1C* and *CMIP*.

Gene Name	Subtype	KEGG Pathways
<i>TXN</i>	Nonmalignant	<ul style="list-style-type: none"> <li>Fluid shear stress and atherosclerosis</li> <li>NOD-like receptor signaling pathway</li> </ul>
<i>ANXA2</i>	Nonmalignant	<ul style="list-style-type: none"> <li>No Data Available</li> </ul>
<i>LOXL2</i>	Luminal	<ul style="list-style-type: none"> <li>No Data Available</li> </ul>
<i>TPM4</i>	Luminal	<ul style="list-style-type: none"> <li>Adrenergic signaling in cardiomyocytes</li> <li>Cardiac muscle contraction</li> <li>Dilated cardiomyopathy (DCM)</li> <li>Hypertrophic cardiomyopathy (HCM)</li> </ul>
<i>TUBA1C</i>	Basal	<ul style="list-style-type: none"> <li>Apoptosis</li> <li>Gap junction</li> <li>Phagosome</li> <li>Tight junction</li> </ul>
<i>CMIP</i>	Basal	No Data Available
<i>TPRN</i>	Claudin-low	No Data Available
<i>ADCY6</i>	Claudin-low	<ul style="list-style-type: none"> <li>Adrenergic signaling in cardiomyocytes</li> <li>Dilated cardiomyopathy (DCM)</li> <li>Gap junction</li> <li>Aldosterone synthesis and secretion</li> <li>Apelin signaling pathway</li> <li>Bile secretion</li> <li>cAMP signaling pathway</li> </ul>

## 5. Conclusions

Our study used weighted gene coexpression analysis to construct a gene coexpression network and to identify and validate network hub genes associated with the progression and poor prognosis of breast cancer. Eight hub genes, namely, *TXN*, *ANXA2*, *TPM4*, *LOXL2*, *TPRN*, *ADCY6*, *TUBA1C*, and *CMIP*, were identified and validated as associated with the progression and poor prognosis of breast cancer.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2077-0383/8/8/1160/s1>, Figure S1: Protein–Protein network of *TXN*, *ANXA2*, *TPM4*, *LOXL2*, *TPRN*, *ADCY6*, *TUBA1C*, and *CMIP* based on the STRING website. Orange is eight key genes, blue connects *ADCY6* and *ANXA2*.

**Author Contributions:** Conceptualization, C.C.N.W. and C.Y.L.; methodology, J.-H.C., P.C.-Y.S.; software, J.-H.C., P.C.-Y.S.; validation, C.C.-N.W., M.-Y.W.; formal analysis, J.J.P.T.; data curation, C.C.-N.W., J.-H.C., P.C.-Y.S.; writing—original draft preparation, C.C.-N.W., C.-J.L., M.-F.H.; writing—review and editing, C.C.-N.W., C.-J.L., M.-F.H.; visualization, M.-Y.W.; supervision, C.-J.L., M.-F.H.

**Funding:** The present study was funded by the Ministry of Science and Technology, Taiwan (grant no. 105-2314-B-037-038-MY3 and 107-2632-E-468-002), Kaohsiung Medical University Research Center Grant (grant no. KMU-TC108A04) and Show Chwan Memorial Hospital (grant no. RD108030).

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

## References

1. Waks, A.G.; Winer, E.P. Breast Cancer Treatment: A Review. *JAMA* **2019**, *321*, 288–300. [CrossRef]
2. Bhikoo, R.; Srinivasa, S.; Yu, T.C.; Moss, D.; Hill, A.G. Systematic review of breast cancer biology in developing countries (part 2): Asian subcontinent and South East Asia. *Cancers* **2011**, *3*, 2382–2401. [PubMed]
3. Tang, J.; Kong, D.; Cui, Q.; Wang, K.; Zhang, D.; Gong, Y.; Wu, G. Prognostic Genes of Breast Cancer Identified by Gene Coexpression Network Analysis. *Front. Oncol.* **2018**, *8*, 374. [CrossRef] [PubMed]
4. Fan, L.; Goss, P.E.; Strasser-Weippl, K. Current Status and Future Projections of Breast Cancer in Asia. *Breast Care* **2015**, *10*, 372–378. [CrossRef] [PubMed]
5. Li, J.; Zhou, D.; Qiu, W.; Shi, Y.; Yang, J.J.; Chen, S.; Wang, Q.; Pan, H. Application of Weighted Gene Coexpression Network Analysis for Data from Paired Design. *Sci. Rep.* **2018**, *8*, 622. [CrossRef] [PubMed]
6. Yuan, L.; Chen, L.; Qian, K.; Qian, G.; Wu, C.L.; Wang, X.; Xiao, Y. Coexpression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccRCC). *Genom. Data* **2017**, *14*, 132–140. [CrossRef]
7. Stratton, M.R.; Campbell, P.J.; Futreal, P.A. The cancer genome. *Nature* **2009**, *458*, 719. [PubMed]
8. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559.
9. Daemen, A.; Griffith, O.L.; Heiser, L.M.; Wang, N.J.; Enache, O.M.; Sanborn, Z.; Pepin, F.; Durinck, S.; Korkola, J.E.; Griffith, M.; et al. Modeling precision treatment of breast cancer. *Genome Biol.* **2013**, *14*, R110. [CrossRef]
10. Durinck, S.; Moreau, Y.; Kasprzyk, A.; Davis, S.; De Moor, B.; Brazma, A.; Huber, W. BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* **2005**, *21*, 3439–3440.
11. Durinck, S.; Spellman, P.T.; Birney, E.; Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **2009**, *4*, 1184–1191. [PubMed]
12. Del Re, A.C.; Hoyt, W. MAD: Meta-Analysis with Mean Differences. 2011. Available online: [https://www.researchgate.net/publication/215543946\\_MAd\\_Meta-Analysis\\_with\\_Mean\\_Differences](https://www.researchgate.net/publication/215543946_MAd_Meta-Analysis_with_Mean_Differences) (accessed on 25 June 2019).
13. Langfelder, P.; Horvath, S. Fast R Functions for Robust Correlations and Hierarchical Clustering. *J. Stat. Softw.* **2012**, *46*. [CrossRef]
14. Horvath, S.; Dong, J. Geometric Interpretation of Gene Coexpression Network Analysis. *PLoS Comput. Biol.* **2008**, *4*, e1000117. [CrossRef] [PubMed]
15. Zhang, B.; Horvath, S. A general framework for weighted gene coexpression network analysis. *Stat. Appl. Genet. Mol. Biol.* **2005**, *4*. [CrossRef] [PubMed]
16. Van Dam, S.; Vosa, U.; van der Graaf, A.; Franke, L.; de Magalhaes, J.P. Gene coexpression analysis for functional classification and gene–disease predictions. *Brief. Bioinform.* **2017**, *19*, 575–592.
17. Freeman, L.C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1978**, *1*, 215–239. [CrossRef]
18. Azuaje, F.J. Selecting biologically informative genes in coexpression networks with a centrality score. *Biol. Direct* **2014**, *9*, 12. [CrossRef]
19. Benito-Martin, A.; Peinado, H. FunRich proteomics software analysis, let the fun begin! *Proteomics* **2015**, *15*, 2555–2556.
20. Miller, J.W.; Royalty, J.; Henley, J.; White, A.; Richardson, L.C. Breast and cervical cancers diagnosed and stage at diagnosis among women served through the National Breast and Cervical Cancer Early Detection Program. *Cancer Causes Control* **2015**, *26*, 741–747. [CrossRef]
21. Qiu, J.; Du, Z.; Wang, Y.; Zhou, Y.; Zhang, Y.; Xie, Y.; Lv, Q. Weighted gene coexpression network analysis reveals modules and hub genes associated with the development of breast cancer. *Medicine* **2019**, *98*, e14345. [CrossRef]
22. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [CrossRef]
23. Clarke, C.; Madden, S.F.; Doolan, P.; Aherne, S.T.; Joyce, H.; O’Driscoll, L.; Gallagher, W.M.; Hennessy, B.T.; Moriarty, M.; Crown, J.; et al. Correlating transcriptional networks to breast cancer survival: A large-scale coexpression analysis. *Carcinogenesis* **2013**, *34*, 2300–2308. [CrossRef]

24. Chen, C.; Cheng, L.; Grennan, K.; Pibiri, F.; Zhang, C.; Badner, J.A.; Gershon, E.S.; Liu, C. Two gene coexpression modules differentiate psychotics and controls. *Mol. Psychiatry* **2013**, *18*, 1308–1314. [[CrossRef](#)]
25. Feng, Y.; Li, Y.; Li, L.; Wang, X.; Chen, Z. Identification of specific modules and significant genes associated with colon cancer by weighted gene coexpression network analysis. *Mol. Med. Rep.* **2019**, *20*, 693–700.
26. Filloux, C.; Cedric, M.; Romain, P.; Lionel, F.; Christophe, K.; Dominique, R.; Abderrahman, M.; Daniel, P. An integrative method to normalize RNA-Seq data. *BMC Bioinform.* **2014**, *15*, 188. [[CrossRef](#)]
27. Shang, W.; Xie, Z.; Lu, F.; Fang, D.; Tang, T.; Bi, R.; Chen, L.; Jiang, L. Increased Thioredoxin-1 Expression Promotes Cancer Progression and Predicts Poor Prognosis in Patients with Gastric Cancer. *Oxidative Med. Cell. Longev.* **2019**, *2019*, 9291683. [[CrossRef](#)]
28. Cho, S.Y.; Kim, S.; Son, M.J.; Rou, W.S.; Kim, S.H.; Eun, H.S.; Lee, B.S. Clinical Significance of the Thioredoxin System and Thioredoxin-Domain-Containing Protein Family in Hepatocellular Carcinoma. *Dig. Dis. Sci.* **2019**, *64*, 123–136. [[CrossRef](#)]
29. Zhao, X.; Jiang, M.; Wang, Z. TPM4 promotes cell migration by modulating F-actin formation in lung cancer. *Onco Targets Ther.* **2019**, *12*, 4055–4063. [[CrossRef](#)]
30. Lomnytska, M.I.; Becker, S.; Bodin, I.; Olsson, A.; Hellman, K.; Hellstrom, A.C.; Mints, M.; Hellman, U.; Auer, G.; Andersson, S. Differential expression of ANXA6, HSP27, PRDX2, NCF2, and TPM4 during uterine cervix carcinogenesis: Diagnostic and prognostic value. *Br. J. Cancer* **2011**, *104*, 110–119. [[CrossRef](#)]
31. Tang, H.Y.; Beer, L.A.; Tanyi, J.L.; Zhang, R.; Liu, Q.; Speicher, D.W. Protein isoform-specific validation defines multiple chloride intracellular channel and tropomyosin isoforms as serological biomarkers of ovarian cancer. *J. Proteom.* **2013**, *89*, 165–178. [[CrossRef](#)]
32. De Jong, O.G.; van der Waals, L.M.; Kools, F.R.W.; Verhaar, M.C.; van Balkom, B.W.M. Lysyl oxidase-like 2 is a regulator of angiogenesis through modulation of endothelial-to-mesenchymal transition. *J. Cell. Physiol.* **2019**, *234*, 10260–10269. [[CrossRef](#)]
33. Shao, B.; Zhao, X.; Liu, T.; Zhang, Y.; Sun, R.; Dong, X.; Liu, F.; Zhao, N.; Zhang, D.; Wu, L.; et al. LOXL2 promotes vasculogenic mimicry and tumour aggressiveness in hepatocellular carcinoma. *J. Cell. Mol. Med.* **2019**, *23*, 1363–1374. [[CrossRef](#)]
34. Tarnawski, L.; Reardon, C.; Caravaca, A.S.; Rosas-Ballina, M.; Tusche, M.W.; Drake, A.R.; Hudson, L.K.; Hanes, W.M.; Li, J.H.; Parrish, W.R.; et al. Adenylyl Cyclase 6 Mediates Inhibition of TNF in the Inflammatory Reflex. *Front. Immunol.* **2018**, *9*, 2648. [[CrossRef](#)]
35. Fan, N.J.; Kang, R.; Ge, X.Y.; Li, M.; Liu, Y.; Chen, H.M.; Gao, C.F. Identification alpha-2-HS-glycoprotein precursor and tubulin beta chain as serology diagnosis biomarker of colorectal cancer. *Diagn. Pathol.* **2014**, *9*, 53. [[CrossRef](#)]
36. Zhang, J.; Huang, J.; Wang, X.; Chen, W.; Tang, Q.; Fang, M.; Qian, Y. CMIP is oncogenic in human gastric cancer cells. *Mol. Med. Rep.* **2017**, *16*, 7277–7286. [[CrossRef](#)]
37. Wang, B.; Wu, Z.S.; Wu, Q. CMIP Promotes Proliferation and Metastasis in Human Glioma. *BioMed Res. Int.* **2017**, *2017*, 5340160. [[CrossRef](#)]
38. Tsai, F.C.; Kuo, G.H.; Chang, S.W.; Tsai, P.J. Ca<sup>2+</sup> signaling in cytoskeletal reorganization, cell migration, and cancer metastasis. *BioMed Res. Int.* **2015**, *2015*, 409245. [[CrossRef](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).