



Article

Identification of New Key Genes and Their Association with Breast Cancer Occurrence and Poor Survival Using In Silico and In Vitro Methods

Rafat Ali ¹, Armiya Sultan ^{1,†} , Romana Ishrat ², Shafiul Haque ^{3,4,5} , Nida Jamil Khan ^{1,*} and Miguel Angel Prieto ^{6,*}

¹ Department of Biosciences, Jamia Millia Islamia (A Central University), New Delhi 110025, India

² Center for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia (A Central University), New Delhi 110025, India

³ Research and Scientific Studies Unit, College of Nursing and Allied Health Sciences, Jazan University, Jazan 45142, Saudi Arabia

⁴ Gilbert and Rose-Marie Chagoury School of Medicine, Lebanese American University, Beirut P.O. Box 36, Lebanon

⁵ Centre of Medical and Bio-Allied Health Sciences Research, Ajman University, Ajman P.O. Box 346, United Arab Emirates

⁶ Nutrition and Bromatology Group, Department of Analytical Chemistry and Food Science, Faculty of Science, Universidade de Vigo, E32004 Ourense, Spain

* Correspondence: njkhan@jmi.ac.in (N.J.K.); mprieto@uvigo.es (M.A.P.)

† Current address: Department of Biotechnology, Jamia Millia Islamia (A Central University), New Delhi 110025, India.

Abstract: Breast cancer is one of the most prevalent types of cancer diagnosed globally and continues to have a significant impact on the global number of cancer deaths. Despite all efforts of epidemiological and experimental research, therapeutic concepts in cancer are still unsatisfactory. Gene expression datasets are widely used to discover the new biomarkers and molecular therapeutic targets in diseases. In the present study, we analyzed four datasets using R packages with accession number GSE29044, GSE42568, GSE89116, and GSE109169 retrieved from NCBI-GEO and differential expressed genes (DEGs) were identified. Protein–protein interaction (PPI) network was constructed to screen the key genes. Subsequently, the GO function and KEGG pathways were analyzed to determine the biological function of key genes. Expression profile of key genes was validated in MCF-7 and MDA-MB-231 human breast cancer cell lines using qRT-PCR. Overall expression level and stage wise expression pattern of key genes was determined by GEPIA. The bc-GenExMiner was used to compare expression level of genes among groups of patients with respect to age factor. OncoLnc was used to analyze the effect of expression levels of *LAMA2*, *TIMP4*, and *TMTC1* on the survival of breast cancer patients. We identified nine key genes, of which *COL11A1*, *MMP11*, and *COL10A1* were found up-regulated and *PCOLCE2*, *LAMA2*, *TMTC1*, *ADAMTS5*, *TIMP4*, and *RSPO3* were found down-regulated. Similar expression pattern of seven among nine genes (except *ADAMTS5* and *RSPO3*) was observed in MCF-7 and MDA-MB-231 cells. Further, we found that *LAMA2*, *TMTC1*, and *TIMP4* were significantly expressed among different age groups of patients. *LAMA2* and *TIMP4* were found significantly associated and *TMTC1* was found less correlated with breast cancer occurrence. We found that the expression level of *LAMA2*, *TIMP4*, and *TMTC1* was abnormal in all TCGA tumors and significantly associated with poor survival.

Keywords: breast cancer; differentially expressed genes; down regulated genes; poor survival; up regulated genes

1. Introduction

Breast cancer is one of the most prevalent types of cancer diagnosed globally. Its burden has been rising over the past decades accounting for 1 in 8 cancer diagnoses



Citation: Ali, R.; Sultan, A.; Ishrat, R.; Haque, S.; Khan, N.J.; Prieto, M.A. Identification of New Key Genes and Their Association with Breast Cancer Occurrence and Poor Survival Using In Silico and In Vitro Methods.

Biomedicines **2023**, *11*, 1271. <https://doi.org/10.3390/biomedicines11051271>

Academic Editor: Khalil Helou

Received: 31 January 2023

Revised: 20 February 2023

Accepted: 7 March 2023

Published: 25 April 2023



Copyright: © 2023 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/>).

worldwide and a total of 2.3 million new cases in both sexes combined [1,2]. Breast cancer continues to have a significant impact on the global number of cancer deaths. Estimation reflected that about 685,000 women died from breast cancer in 2020, corresponding to 16% or 1 in every 6 cancer deaths in women [1]. Breast cancer cases are predicted to reach 4.4 million in 2070 [3]. The incidence of breast cancer varies around the world, with the highest rates typically found in more developed countries [4,5]. However, the number of breast cancer cases is also exponentially increasing in low- and middle-income countries [6]. Breast cancer incidence is strongly linked to human development and is higher among nations with the highest human development index [2]. In females, the worldwide age-standardized incidence rate is estimated to be 48/100,000, ranging from less than 30/100,000 in Sub-Saharan Africa to more than 70/100,000 in Western Europe and North America [7]. Although the relative incidence of breast cancer is highest in the most developed parts of the world, since less developed regions have much bigger populations means more than half of all breast cancer cases are identified in low- and middle-income nations, creating a significant considerable disease burden [2,7]. Despite the high incidence of the disease, early detection and improved treatment options have led to a decrease in breast cancer deaths in many countries. However, access to early detection and treatment options remains a concern in low- and middle-income countries, and more efforts are needed to address this issue.

The reported worldwide differences in the incidence of breast cancer must be viewed in the context of the disease's recognized risk factors. There are several risk factors that have been associated with an increased risk of developing breast cancer. Some of the most well-known major risk factors include older age, breast density and family history of breast cancer, early age menarche, late age at first full-term pregnancy, shorter breastfeeding periods, use of hormonal menopausal therapy, use of oral contraceptives, high fat dietary, high body mass index, physical inactivity, obesity and exposure to tobacco [4,8,9]. Highest incidence rate of breast cancer in developed countries is likely due to a combination of factors such as an aging population, lifestyle factors, and improved access to healthcare and cancer screening programs [10]). Increasing incidence of breast cancer in low- and middle-income countries is likely attributed to changes in lifestyle and reproductive patterns, such as an increase in the use of hormone replacement therapy (HRT) and a decrease in the number of children born to women [4]. Additionally, increased urbanization and the adoption of western lifestyles in these countries may also contribute to the rise in breast cancer incidence [6]. It's worth noting that having one or more of these risk factors does not mean that a woman will develop breast cancer, and many women who develop the disease do not have any known risk factors.

Early diagnosis and treatment are crucial for a positive outcome of any disease. The treatment of breast cancer depends on several factors, including the stage and type of cancer, as well as the patient's overall health. Despite modern advances in target therapy method, the result of treating breast cancer is still unsatisfactory. Thus, understanding the molecular mechanisms of breast cancer progression and identifying novel potential prognostic biomarkers and molecular targets are urgently needed. This will also give deep insight for the diagnosis and treatment at every stage of breast cancer. Currently, in-silico techniques followed by in-vitro or in-vivo validations are widely used to identify the key regulatory genes and to determine their structural and relational aspects with disease [11]. This approach offers an ideal way to screen large gene expression profiles of normal and clinical populations to understand the genomic mechanisms contributing for the development and progression of different diseases [12]. Nowadays, high-throughput microarray technology and bioinformatics analysis is widely used to find gene expression variations between malignant and non-cancerous tissues, assess differentially expressed genes (DEGs), and uncover the pathways that contribute to carcinogenesis and cancer development. New biomarkers and therapeutic targets have been discovered from cancer-associated gene expression profiles which showed reliable outcome in clinical studies [13–16].

In the present study, NCBI-GEO database was accessed to retrieve four cancer-associated gene expression profiles. These datasets were analyzed by integrated in-silico methods to determine the DEGs associated with breast cancer. Key regulatory genes having high correlation with sample clinical characteristics were extracted and further validated using MCF-7 and MDA-MB-231 human breast cancer cell lines.

2. Materials and Methods

2.1. Retrieval of Datasets and Extraction of Differentially Expressed Genes (DEGs)

The NCBI-GEO (<http://www.ncbi.nlm.nih.gov/geo/> (accessed on 3 March 2021)) database is a free public database of microarrays. It is used for gene expression datasets and platform records. Four datasets with accession number GSE29044, GSE42568, GSE89116, and GSE109169 were retrieved from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/> (accessed on 3 March 2021)) and analyzed using R packages. DEGs between the cancer and normal samples were identified by applying statistical parameters. The DEGs with $FC \geq 1.5$ for up regulated and $FC \leq -1.5$ for down regulated and adjust $p < 0.05$ were considered as the cut-off criteria.

2.2. Construction of Protein-Protein Interaction (PPI) Network

The identified DEGs were further used as input in network building. The PPI network was constructed by using the STRING database. A total of 161 DEGs extracted from all the four datasets were used to make the PPI network. Further, analysis of the network and their visualization process was performed by Cytoscape software (version 3.7.1) [17].

2.3. Characterization of Networks Topological Properties

Topological parameter behaviors were explored to determine the structural properties of complex networks by using the Network Analyser plugin in Cytoscape [18]. The Topological properties analyzed are explained below:

2.4. Probability of Degree Distribution

The degree distribution of a PPI network is a probability distribution of a node to have degree k . The ' k ' represents the number of links of a node that connects with other nodes. For example, If $G = (N, E)$ describes a graph of a network, where N and E represent the node and edges of the network respectively. The, the degree distribution probability ($P(k)$) of a network is measured by the equation [19,20]:

$$P(k) = nk/N \quad (1)$$

where, nk = number of nodes having degree k and N = total number of nodes in the network.

2.5. Betweenness Centrality

In a node of a PPI network, Betweenness centrality characterizes the prominence of information that is flowing through one node to another by following a shortest path [21,22]. The geodesic paths are shown from node i to node j by ' $d_{ij}(v)$ ' which passes through node ' v ' and ' d_{ij} '. The Betweenness centrality of a node v can be measured by the following equation:

$$CB(v) = \sum_{i,j, i \neq j \neq k} (d_{ij}(v)/d_{ij}) \quad (2)$$

2.6. Closeness Centrality

Closeness centrality characterizes how quickly the information is travelling through the network i.e., from one node of the PPI network to another node [23]. The Closeness centrality of the node i is described as the reciprocal average length of the geodesic paths between the node and all other nodes connected to it in the network. Closeness Centrality is measured by the following equation:

$$CC(k) = n / (\sum_j d_{ij}) \quad (3)$$

where, d_{ij} in a PPI network presents length of the geodesic path between nodes i and j . n presents total number of nodes in the PPI network connected to node i .

2.7. Community Detection: Leading Eigen Vector Approach

To characterize the modular nature, attributes, and organizing principle of the hierarchical network, the activities of the created network were defined at various levels of hierarchy [24]. In this work, the Leading Eigen Vector technique (LEV) [25,26] from the package 'igraph' [27] was utilized in R to discover communities. The LEV technique is the most promising for community discovery since it computes the Eigen value for each connection, demonstrating the significance of each link rather than nodes. We discovered modules from the entire network and then sub-modules from the modules at each level of organisation to retrieve just the theme (Figure S1).

2.8. Genes Tracing across the Networks

Identifying the primary forces that influence the regulation of the PPI network is one special problem [28]. This was accomplished by gene tracing using the LCV method in CYTOSCAPE. This gene tracing was done up to the motif level in several modules/sub-modules derived through clustering. The network's regulator was identified by tracking the most significant and influential nodes within the network's construction.

2.9. Gene Ontology and Pathway Analysis of Key Genes

The DAVID (<https://david.ncifcrf.gov/> (accessed on 7 June 2021)) and KEGG databases were utilized for GO and pathway analysis to explore the function and associated pathways of the key genes [29,30]. Gene ontology (GO) analysis annotates genes and gene products using functions such as molecular function, biological pathways, and cellular components [31]. KEGG is a collection of genomic and enzymatic techniques, as well as an online library of biological chemical energy [32]. KEGG is a resource for comprehensive gene function analysis as well as associated high-level genome functional information. DAVID can give full biological function annotation information for high-throughput gene expression [33]. As a result, we used DAVID online tools to perform GO and KEGG pathway analyses on the key genes at the functional level. A $p < 0.05$ was considered statistically significantly different.

2.10. GEPIA Analysis

GEPIA (<http://gepia.cancer-pku.cn/detail.php> (accessed on 7 June 2021)) is a specialized web server for the analysis of RNA-seq data of 9736 tumors and 8587 normal samples from the TCGA (<http://portal.gdc.cancer.gov/> (accessed on 7 June 2021)) and the GTEx (<http://gtexportal.org/home/> (accessed on 7 June 2021)) projects [34]. Using the GEPIA web server, expression level, survival, and expression level at different stages of key genes was studied. The predictive value of all key genes was assessed throughout the TCGA dataset using the GEPIA web server's default parameters. The default values for all parameters were used, and the cut-off value was set at median = 50 percent. $p < 0.05$ was used to indicate a statistically significant difference in the HR.

2.11. bc-GenExMiner Analysis

The Breast Cancer Gene Expression Miner v4.4, (<http://bcgenex.ico.unicancer.fr/BC-GEM/GEM-Accueil.php?js=1> (accessed on 7 June 2021)) a DNA microarray and RNA-seq database may be used to look at gene expression and predict prognosis. We looked at the relationship between gene expression of the key genes and clinic pathological parameters like age, and specific region of breast cancer patient sample using microarray data. Furthermore, we conducted a prognostic analysis of the genes as well.

2.12. UALCAN Analysis

The online cancer transcriptome database UALCAN (<http://ualcan.path.uab.edu/> (accessed on 7 June 2021)) is meant to enable simple access to publicly accessible cancer transcriptome data (TCGA and MET500 transcriptome sequencing) [35]. UALCAN is a comprehensive, user-friendly, and interactive web resource for analyzing cancer OMICS data. UALCAN enables researchers to access Level 3 RNA-seq data from The Cancer Genome Atlas (TCGA) and perform gene expression and survival analysis on about 20,500 protein-coding genes in 33 different tumor types [35]. It's written in PERL-CGI and has high-resolution visuals created using JavaScript and CSS. The Clinical Proteomic Tumor Analysis Consortium (CPTAC) Confirmatory/Discovery dataset is now available in UALCAN for protein expression analysis. The level of expression of key genes in normal breast tissue and primary invasive breast cancer was compared using this database.

2.13. OncoLnc Analysis

2.13.1. Survival Analysis

An overall survival analysis for patients with breast cancer was performed using the OncoLnc program (www.oncolnc.org (accessed on 7 June 2021)). OncoLnc is an interactive online application that allows users to explore the survival data of 8647 individuals from 21 cancer studies in The Cancer Genome Atlas (TCGA), as well as TCGA's mRNA and miRNA RNA-Seq expression data. The software allows you to create Kaplan-Meier graphs that are stratified by gene expression levels. In survival analysis, log-rank *p*-values were collected. High and low groups were defined as the 80th (upper) and 20th (lower) percentiles, respectively. On 10 November 2021, the survival rate curves were produced using OncoLnc (<http://www.OncoLnc.org/> (accessed on 7 June 2021)) [36]. The upper and lower quartiles were used to divide the high and low expression groups.

2.13.2. Cell Lines, Culture and Validation of Key Regulatory Genes by qRT-PCR

The in-vitro validation of the key regulatory genes was done by using human breast cancer cell lines namely MCF-7 and MDA-MB-231. Cells were procured from National Center for Cell Science (NCCS), Pune, India. Cells were grown and maintained in DMEM media supplemented with 10% FBS and 1% penicillin/streptomycin salt solution at 37 °C in a humidified 5% CO₂ incubator. RNA was extracted from 70 to 75% confluent cells using a Trizol reagent (Ambion, Carlsbad, CA, USA). RNA was quantified using nanodrop, and 1000 ng of RNA was reverse transcribed into cDNA using Verso cDNA Synthesis Kit (Thermo Fisher Scientific, Waltham, MA, USA). qRT-PCR was conducted using SYBRTM Green Master Mix (Thermo Fisher Scientific, USA) and The Applied Biosystems[®] QuantStudioTM 6 Flex Real-Time PCR System to determine the expression profile of nine key regulatory genes obtained through bioinformatics analysis. 18s was used as an endogenous control to normalize the target genes. Thereafter, relative fold change in expression level was calculated for all the nine key regulatory genes. Primer list of respective genes is presented in Table S1.

3. Results

3.1. Characteristics of Datasets Used to Extract Common DEGs

Four datasets having accession number GSE29044, GSE42568, GSE89116, and GSE109169 were obtained from freely accessible NCBI-GEO database. Detailed information of the datasets is presented in Table 1. GSE29044 dataset was based on the GPL570 platform containing 6 samples of early tumour patients and 5 samples of early normal aged between 20–35 years old and 25 samples of late tumour and 7 samples of late normal whose age was greater than 55 years old. GSE89116 dataset was based on the GPL6947 platform containing 11 samples of early tumour patients (max. age 38 years) and 4 samples of early normal (max. age 35 years) and 13 samples of late tumour (max. age 80 years) and 5 samples of late normal (max. age 80 years). GSE109169 dataset was based on the GPL570 platform containing 5 samples of early tumour patients and 5 samples of early normal aged less

than 40 years and 20 samples of late tumour and 20 samples of late normal aged more than 40 years. GSE42568 dataset was based on the GPL570 platform containing 104 tumour samples of patients aged between 31 to 89 years old at the time of diagnosis and 17 normal samples with no defined age.

Table 1. Detailed information of datasets used to extract differentially expressed genes.

| Datasets | Total Samples | Normal (No. of Samples and Age) | | Disease (No. of Samples and Age) | | Up Regulated | Down Regulated | Country | Platform | Author |
|-----------|---------------|---------------------------------|--------------------------|----------------------------------|---------------------------|--------------|----------------|--------------|----------|----------|
| GSE29044 | 43 | 5 (Age 20–35 years) | 7 (Age > 55 years) | 6 (Age 20–35 years) | 25 (Age > 55 years) | 490 | 716 | Saudi Arabia | GPL570 | Colak D |
| GSE89116 | 33 | 4 (Age upto 35 years) | 5 (Age upto 80 years) | 11 (Age upto 38 years) | 13 (Age upto 80 years) | 186 | 542 | India | GPL6947 | Malvia S |
| GSE109169 | 50 | 5 (Age upto 40 years) | 20 (Age > 40 years) | 5 (Age upto 40 years) | 20 (Age > 40 years) | 219 | 358 | Taiwan | GPL570 | Chang JW |
| GSE42568 | 121 | 17 (Age not mentioned) | | 104 (Age 31–89 years) | | 896 | 1202 | Ireland | GPL570 | Clarke C |

The microarray expression profiles are widely utilized to study the gene expression on a genome-wide scale. There are few algorithms available that are used to correct the batch effects before analysing the microarray data. We employed the Empirical Bayes method built-in function in LIMMA, in combination with the fit2 function. Effective in-silico methods are required for the integration of Meta-analyses-based microarray data. These in silico methods are used to merge efficiently various microarray datasets without considering the impact of demographics, experimental designs, and specimen sources [19].

3.2. Identification of Common DEGs

Only those DEGs that surpassed the cut-off criteria of $FC \geq 1.5$ for up regulated genes and $FC \leq -1.5$ for down regulated genes and adjust $p < 0.05$ in all the four data series were considered as the significant DEGs Figure 1. Total 161 common DEGs were extracted from all the four datasets by Venn diagram Figure 1e,f. Among these 161 common DEGs, 44 genes were up regulated and 117 genes were down regulated. List of all the common DEGs is presented in Table 2. Volcano plot represented DEGs in breast cancer tissues and non-tumor samples in datasets. (a) GSE29044 (b) GSE42568 (c) GSE89116 (d) GSE109169 (e) Venn diagram represented the down regulated overlapping DEGs among GSE29044, GSE42568, GSE89116, and GSE109169 datasets. (f) Venn diagram represented the up-regulated overlapping DEGs among GSE29044, GSE42568, GSE89116, and GSE109169 datasets.

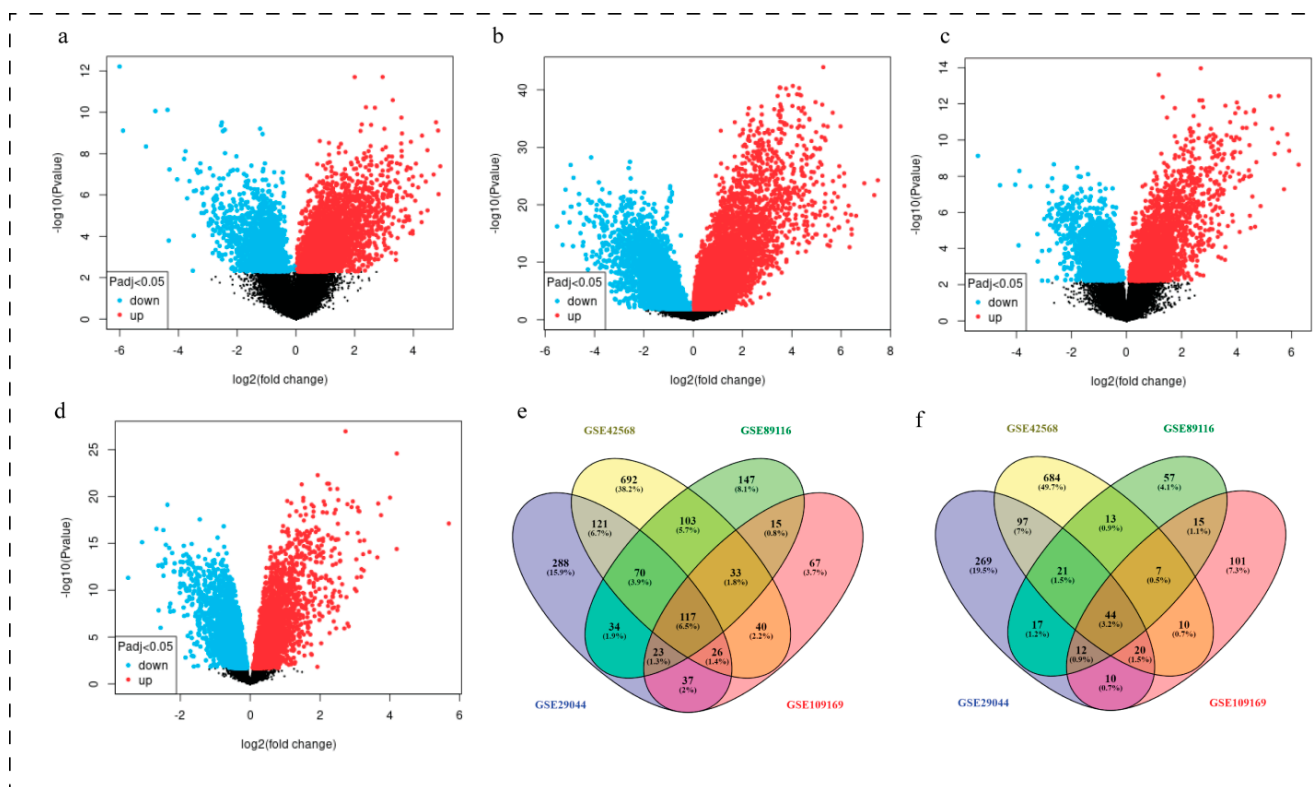


Figure 1. DEGs in breast cancer tissues. Volcano plot presenting DEGs in breast cancer tissues in datasets. (a) GSE29044 (b) GSE42568 (c) GSE89116 (d) GSE109169 (e) Venn diagram represented the down regulated overlapping DEGs among GSE29044, GSE42568, GSE89116, and GSE109169 datasets. (f) Venn diagram represented the up-regulated overlapping DEGs among GSE29044, GSE42568, GSE89116, and GSE109169 datasets.

Table 2. All common differentially expressed genes (DEGs).

| |
|---|
| <p><i>IGF1, LIFR, SFRP1, TSHZ2, SPTBN1, TMEM47, MME, NTRK2, PLAGL1, DMD, MAMDC2, LPL, HLF, PROS1, FABP4, ABCA9, ADAMTS5, RBMS3, APOLD1, SPRY2, ANGPTL1, EBF1, ITM2A, MEOX2, CAV1, TGFB3, OGN, ANKRD29, CRYAB, CLIP4, CD36, EDNRB, CAV2, FAM13A, SYNM, ADH1B, FHLL1, CDO1, TFPI, GPAM, CHRDL1, ABCA8, LYVE1, CACHD1, ITIH5, PLSCR4, SORBS1, LARP6, RUNX1T1, GPC3, PCDH18, AKAP12, PDK4, ACACB, ECM2, SPRY1, PAMR1, CXCL12, ADIPOQ, EMCN, FGF2, ATP1A2, ADH1C, PCOLCE2, ABCA6, SIK2, PCDH9, TMEM100, ARHGAP20, CHL1, MAOA, MTURN, LEP, SLIT3, IGSF10, GSTM5, VIT, RBP7, GPD1, CIDEC, MYZAP, RSPO3, FAXDC2, CFD, FBLN5, LDB2, GNG11, TEK, RBP4, FAM107A, ENPP2, PLIN1, CPED1, TIMP4, GSN, LMOD1, ALDH1A1, LAMA2, GPX3, AOC3, MT1M, JAM2, FXYP1, PPARG, APOD, DPT, G0S2, SH3BGL2, VGLL3, DLC1, PCK1, C7, PDGFD, ANK2, GHR, TMTTC1, CCDC3, PRC1, RRM2, TPX2, CDKN3, DTL, KIF4A, FAM83D, UBE2T, NUF2, MMP11, COL10A1, CCNB1, CDK1, DEPDC1, AURKA, CENPF, KIF23, KIF11, ANLN, FN1, EZH2, TRIP13, PBK, DLGAP5, UHRF1, TK1, CCNB2, MELK, CENPU, ATAD2, HMMR, ECT2, NUSAP1, ASPM, FANCI, SQLE, TOP2A, KIF20A, RACGAP1, GJB2, TLCD1, COL11A1, LRRC15, CKS2</i></p> |
|---|

3.3. Protein-Protein Interaction (PPI) Network of DEGs

The PPI network was constructed by using the String online database and was imported into Cytoscape v. 3.8.0, which supports the visualization of bipartite graph of gene-gene linking/interaction/regulation, reflecting gene-disease associations. This also provides gene-centric views of the network data [20]. The Probe Ids of common DEGs were mapped to their corresponding gene symbols to create the native network Figure 2. The PPI network showed 449 nodes and 18214 edges. The network was characterized by several properties such as average number of neighbours was found 80.811, network diameter was

7, characteristic path length was 2.228, clustering coefficient was 0.662, network density was 0.180, network heterogeneity was 0.783, network centralization was 0.375, connected component was 1 and analysis time of 0.484 s.

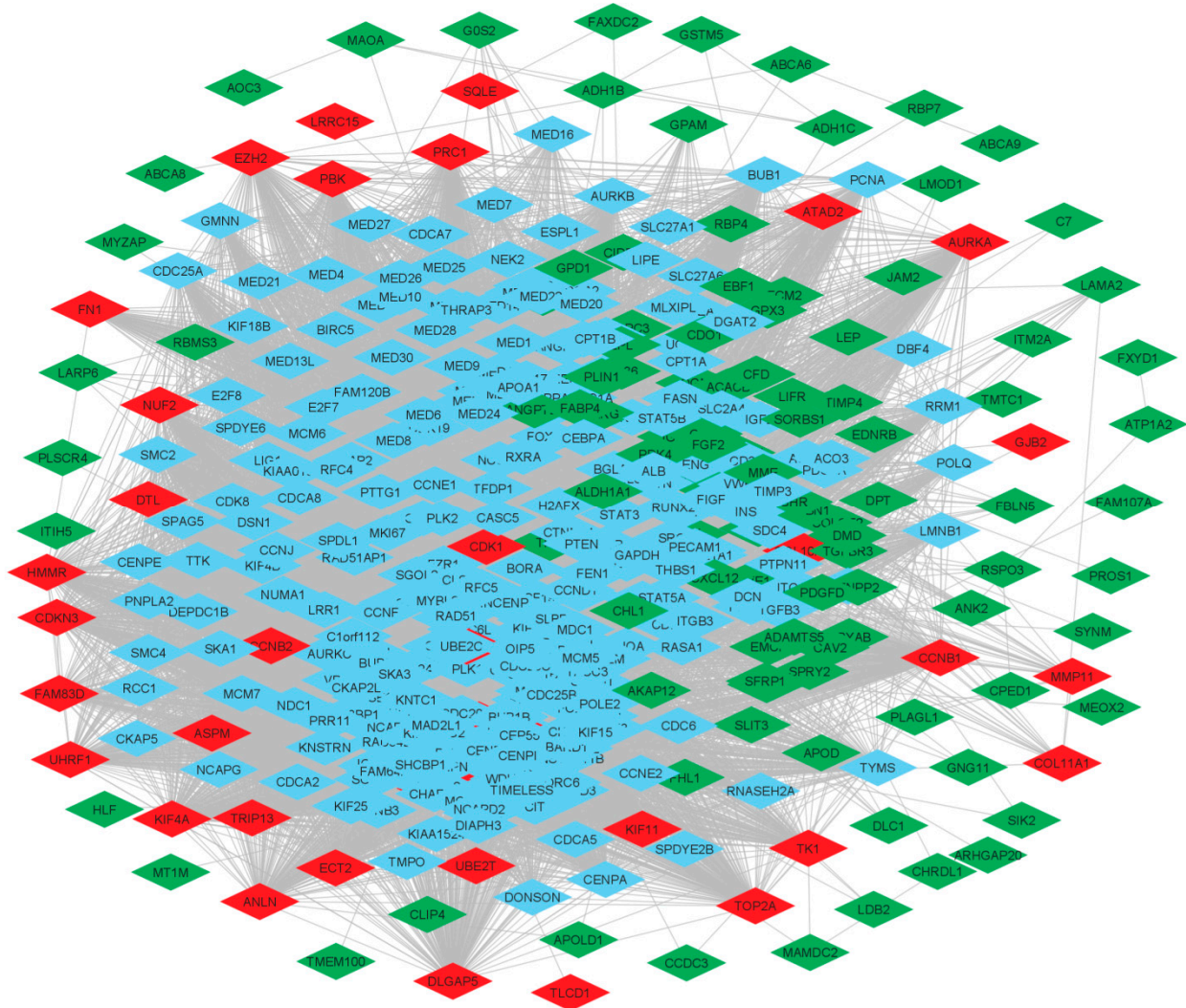


Figure 2. Protein-protein interaction network showing down-regulated (green) DEGs and up-regulated (red) DEGs. Blue colour is presenting supporting genes.

3.4. Community Detection by Leading Eigen Vector Method

To characterise the modular nature, attributes, and organising principle of the hierarchical network, the activities of the created network were defined at various levels of hierarchy. For this, Leading Eigen Vector (LEV) method from the package ‘igraph’ was utilised in R to discover communities. The LEV technique is the most promising for community discovery since it computes the Eigen value for each connection, demonstrating the significance of each link rather than nodes. We discovered modules from the entire network and then sub-modules from the modules at each level of organisation to retrieve just them (Figure 3).

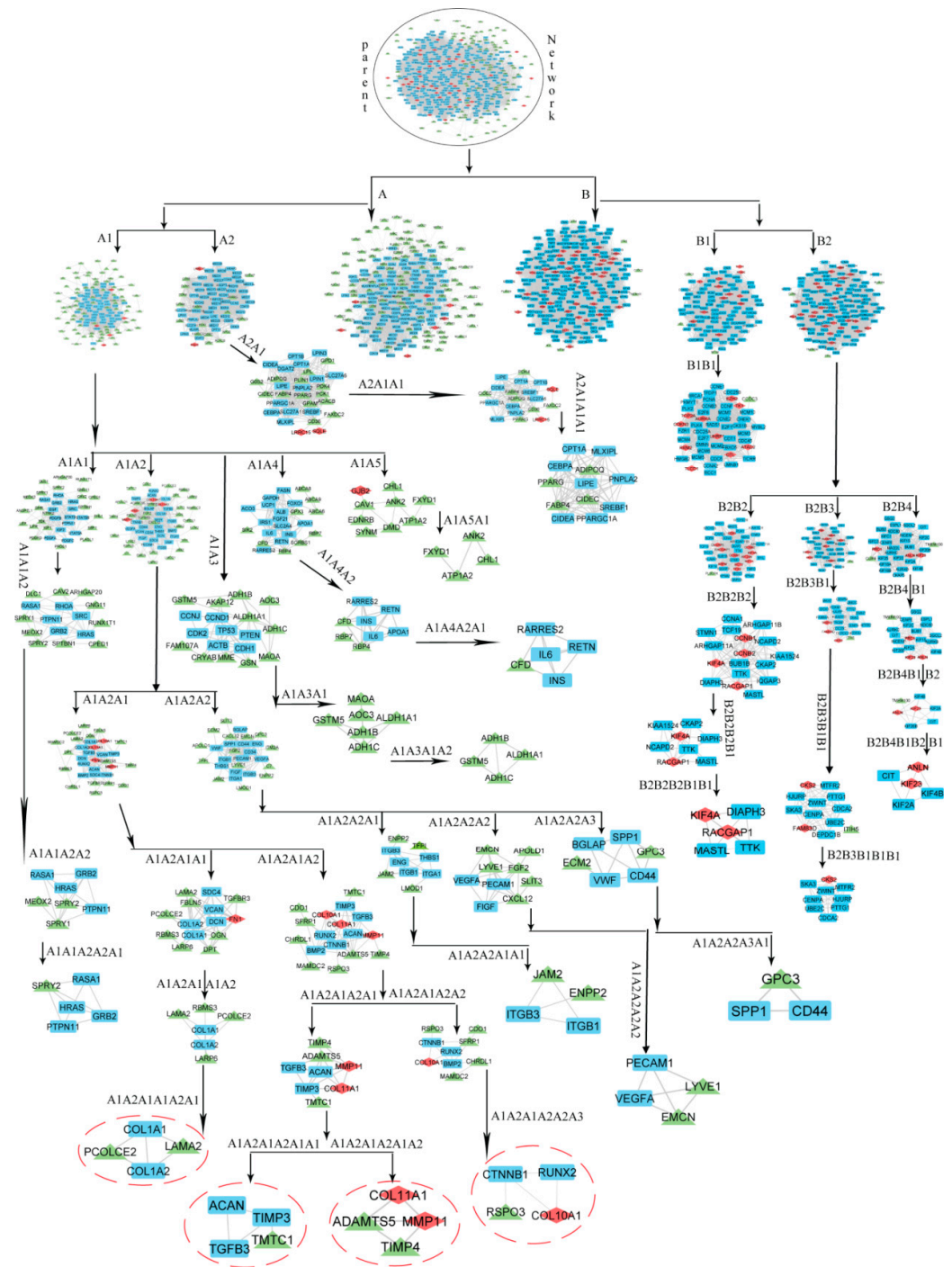


Figure 3. Network/modules/sub-modules at different levels of network.

The modules from the native network along with sub-modules from modules at each level of organization were identified until only motifs remained i.e., unbreakable part of the network.

3.5. Identification of Key Regulators and Properties of Breast Cancer Network

In the constructed PPI Network, we found two communities that were further broken down into sub-community and sub-sub-communities up to seventh level. The analysis of modular structure along its arrangement was carried out by the Newman and Girvan standard community finding techniques [37]. These techniques were employed at different

organizational levels (Figure 3). We found that our PPI network is organized hierarchically through seven different levels.

The leading hubs (nodes) are considered as essential regulators based on changes in the activities of proteins/genes along with their regulating mechanisms. However, all the leading hubs cannot be considered as key regulators for the progression of disease. Only those that regulate the network from top to bottom (where the PPI network cannot be further divided into sub-community and form motif) are considered as important leading hubs. These leading hubs are termed as Key Regulators (KRs). Because these KRs are deeply rooted reaching to motif level (fundamental regulating unit) through different community or sub-community levels of the organization in the PPI network. KRs act as the backbone for network's stability and capacitate it to tackle any unacceptable changes. Topological properties of any PPI network assist to gain deep insight of a network, its behaviour, somehow function, characteristics, and how and what the network is [38]. Therefore, we described some topological properties namely Betweenness Centrality, Degree Centrality, and Closeness Centrality of our network at almost the last level from the parent network (Figure S2). Betweenness Centrality is a way of detecting the amount of influence a node has over the flow of information in a graph. It is often used to find nodes that serve as a bridge from one part of a graph to another. The measure of Degree Centrality presents popular nodes within a graph. It measures the number of incoming or outgoing (or both) relationships from a node, depending on the orientation of a relationship projection. Closeness centrality detects nodes that can spread information very efficiently through a graph. The closeness centrality of a node measures its average farness (inverse distance) to all other nodes. Nodes with a high closeness score have the shortest distances to all other nodes [21]. The gene tracing was done up to the motif level in several modules/sub modules derived through clustering. The network's regulator was identified by tracking the most significant and influential nodes within the network's construction. Finally, we have found nine key regulators namely *PCOLCE2*, *LAMA2*, *TMTC1*, *ADAMTS5*, *TIMP4*, *RSPO3*, *COL11A1*, *MMP11*, and *COL10A1*.

3.6. Gene Ontology and KEGG Pathway Analysis of Key DEGs

All the DEGs were uploaded to the DAVID database (<https://david.ncifcrf.gov/> (accessed on 7 June 2021)) for GO analysis. Results showed that the nine key genes were involved in several GO biological processes namely extracellular structure organization, external encapsulating structure organization, extracellular matrix organization, extracellular matrix disassembly, cellular component disassembly, collagen fibril organization, and supramolecular fibre organization (Table S2). Results showed GO Cellular Components associated with key genes include endoplasmic reticulum lumen, intracellular organelle lumen, collagen-containing extracellular matrix, basement membrane, and Golgi lumen (Table S3). GO Molecular Function analysis showed the key genes were associated with the molecular functions namely metalloendopeptidase activity, metallopeptidase activity, metalloendopeptidase inhibitor activity, and endopeptidase activity (Table S4). Results showed that the key genes were enriched in protein digestion and absorption and viral myocarditis according to the KEGG pathway analysis (Table S5).

3.7. Gene Expression Profiling of Key DEGs

GEPIA (Gene Expression Profiling Interactive Analysis) is a web-based tool for analyzing gene expression data. GEPIA was chosen to analyse hub genes, their overall expression level comparison to normal tissues and stage wise expression pattern of breast cancer scenario. The box plot (Figure 4) of all nine hub genes demonstrates that the genes were abnormally expressed in breast cancer as compared to normal breast tissue. The genes namely *COL11A1*, *MMP11* and *COL10A1* were found up regulated and *PCOLCE2*, *LAMA2*, *TMTC1*, *ADAMTS5*, *TIMP4* and *RSPO3* were found down regulated in breast cancer. The details LogFC, *p*-value, and Adj. *p*. Value of all the nine key genes is presented in Table 3.

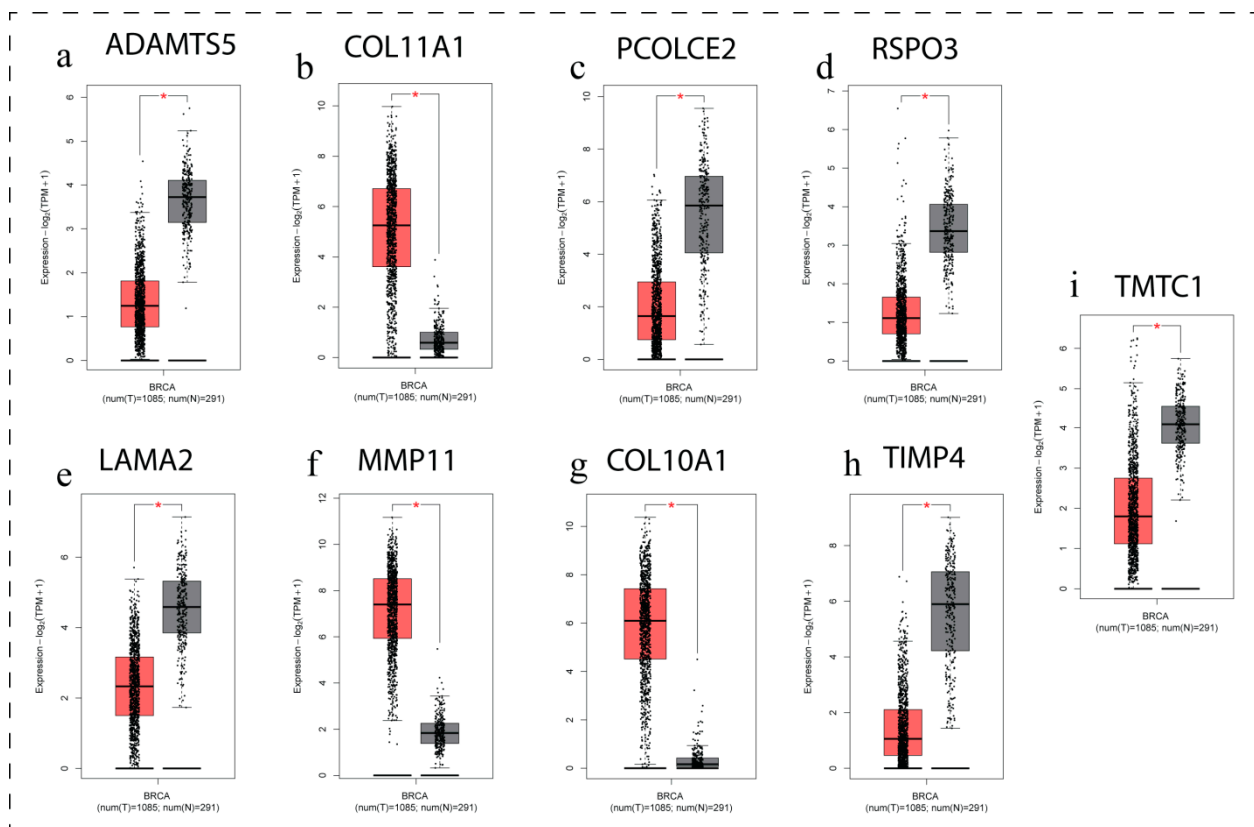


Figure 4. Comparisons of the expression of the nine genes between breast cancer and normal breast tissues in TCGA and GTEx based on GEPIA. The Y axis represents the \log_2 (TPM + 1) for gene expression. The Gray bar indicates the normal tissues, and the red bar shows the breast cancer tissues. These figures were derived from GEPIA. TPM: transcripts per kilobase million. The box plots (a–i) of all nine hub genes demonstrate that the genes were abnormally expressed in breast cancer as compared to normal breast tissue. (a) ADAMTS5—down-regulated, (b) COL11A1—up-regulated, (c) PCOLCE2—down-regulated, (d) RSPO3—down-regulated, (e) LAMA2—down-regulated, (f) MMP11—up-regulated, (g) COL10A1—up-regulated, (h) TIMP4—down-regulated, and (i) TMTC1—down-regulated. * $p < 0.05$.

Table 3. List of key up- and down-regulated genes with LogFC, p -value, and Adj. p Value.

| Genes | LogFC | p Value | Adj. p Value |
|----------------|-------|------------------------|------------------------|
| <i>PCOLCE2</i> | −4.45 | 2.44×10^{-11} | 1.67×10^{-7} |
| <i>LAMA2</i> | −1.89 | 3.21×10^{-4} | 1.26×10^{-2} |
| <i>TMTC1</i> | −2.49 | 2.97×10^{-3} | 4.98×10^{-2} |
| <i>ADAMTS5</i> | −2.31 | 1.80×10^{-4} | 9.13×10^{-3} |
| <i>TIMP4</i> | −2.54 | 1.40×10^{-4} | 7.60×10^{-3} |
| <i>RSPO3</i> | −2.58 | 5.07×10^{-5} | 4.03×10^{-3} |
| <i>COL11A1</i> | 5.29 | 9.91×10^{-14} | 3.60×10^{-12} |
| <i>MMP11</i> | 4.62 | 4.13×10^{-6} | 8.01×10^{-4} |
| <i>COL10A1</i> | 3.49 | 6.29×10^{-13} | 2.01×10^{-11} |

The expression-stage plot analysis (violin plots) revealed that three genes namely *LAMA2*, *TMTC1* and *TIMP4* among these nine genes were found significantly associated ($p < 0.05$) with different stages of breast cancer (Figure 5).

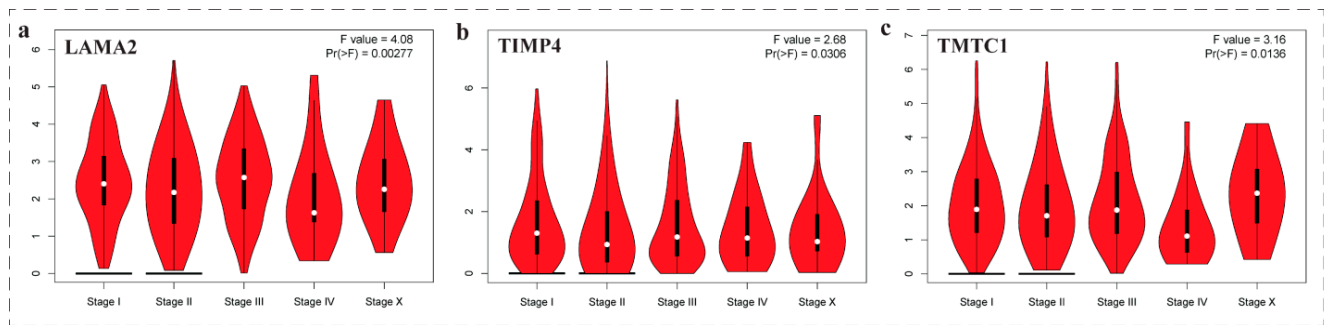


Figure 5. The expression-stage plot of three genes associated with breast cancer. The plots were achieved by the GEPIA web server. The expression-stage plot analysis (violin plots a–c) revealed that three genes namely (a) LAMA2, (b) TIMP4, and (c) TMTC1 among these nine genes were found significantly associated ($p < 0.05$) with different stages of breast cancer.

3.8. Expression Level of Genes among Groups of Patients with Respect to Age Factor

We have used bc-GenExMiner web dependent tool to compare expression level of genes among groups of patients with respect to age factor. We found that the genes namely LAMA2, TMTC1, and TIMP4 were significantly expressed among different age groups of patients, i.e., lower 21 age to higher 97 age groups as indicated by the violin plots (Figure 6). Moreover, we further deeply investigated the role of LAMA2, TMTC1 and TIMP4 genes in breast cancer prognosis. We found that LAMA2 and TIMP4 were significantly associated and TMTC1 gene was less correlated with breast cancer occurrence (Figure 7).

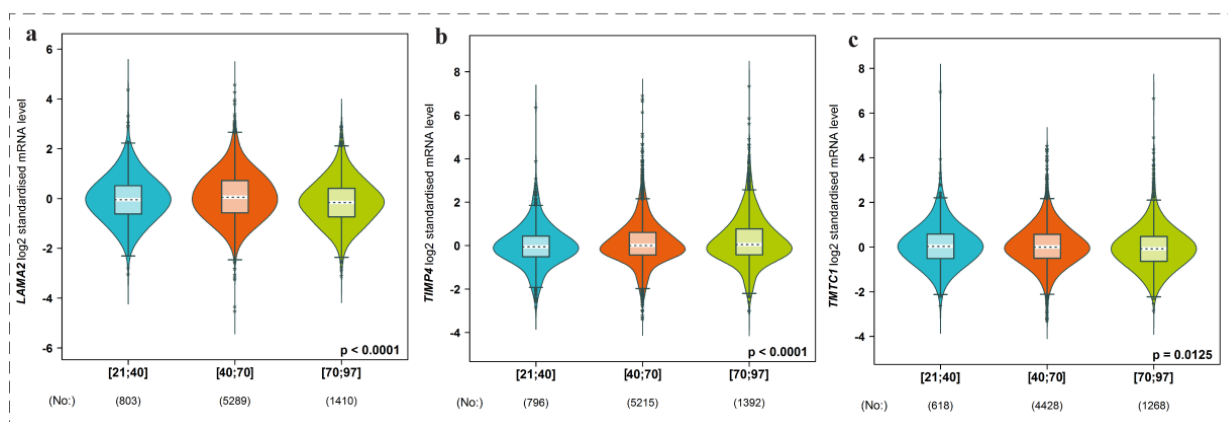


Figure 6. Violin plot showing gene expression among groups of patients categorized according to age (a–c). We found that the genes namely (a) LAMA2, (b) TMTC1, and (c) TIMP4 were significantly expressed among different age groups of patients, i.e., lower 21 age to higher 97 age groups as indicated by the violin plots.

3.9. Pan-Cancer View of LAMA2, TIMP4, and TMTC1 Expression Level Using UALCAN Analysis

UALCAN is a comprehensive, user-friendly, and interactive web resource for analyzing cancer OMICS data. UALCAN enables researchers to access Level 3 RNA-seq data from The Cancer Genome Atlas (TCGA) and perform gene expression and survival analysis on about 20,500 protein-coding genes in 33 different tumor types [8]. Expression levels of LAMA2, TIMP4, and TMTC1 across TCGA tumours are shown in (Figure 8). We found that the expression level of LAMA2, TIMP4, and TMTC1 was higher in all TCGA tumours.

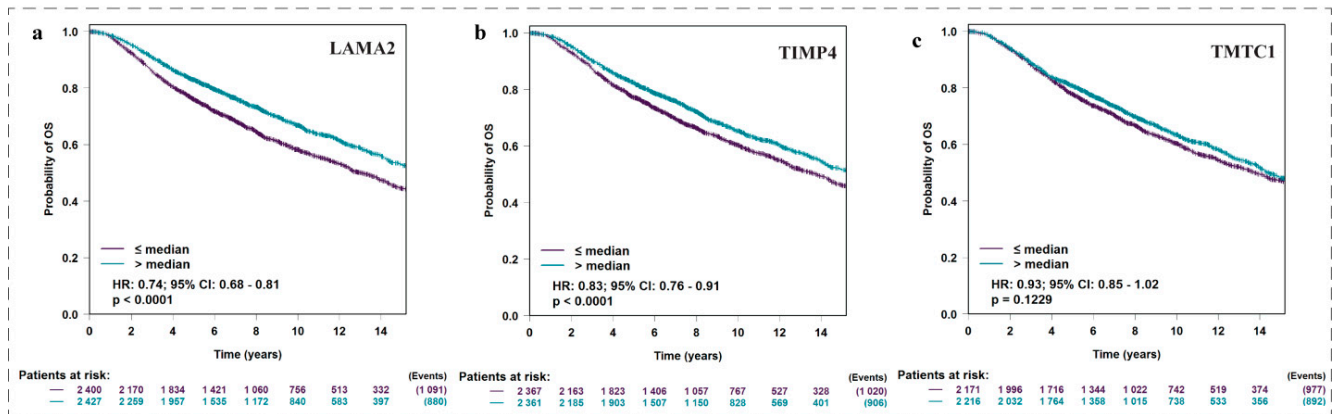


Figure 7. Association of genes with breast cancer occurrence (a–c). (a) LAMA2 and (b) TIMP4 were found significantly associated and (c) TMTC1 gene was found less correlated with breast cancer occurrence.

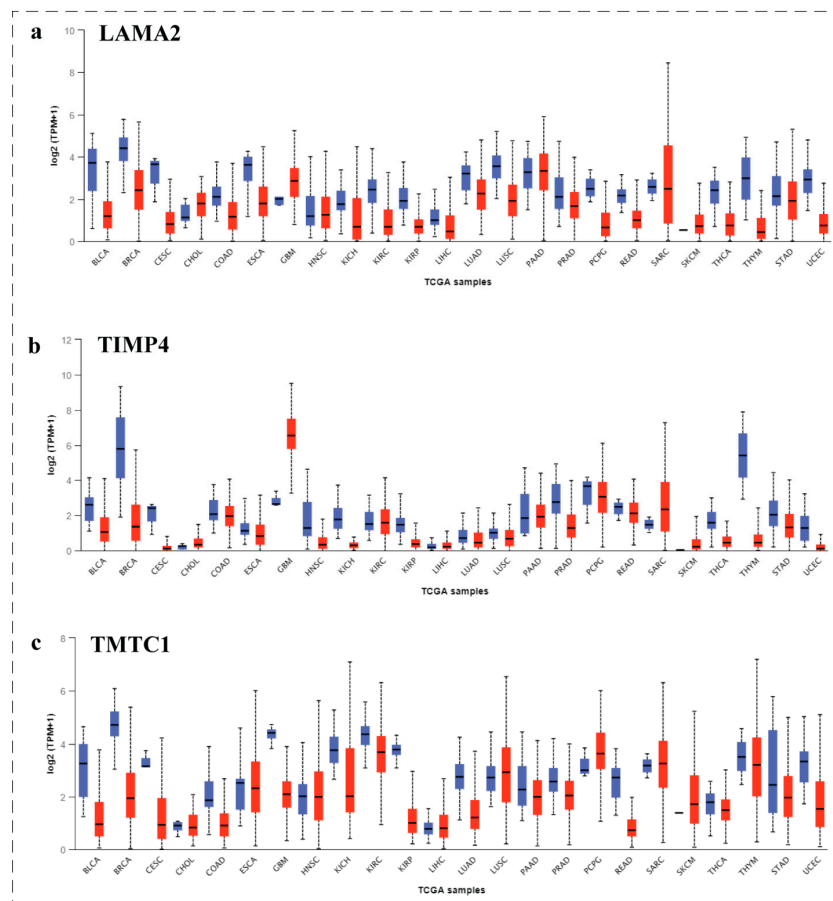


Figure 8. Pan-cancer view of *LAMA2*, *TIMP4*, and *TMTC1* expression level (a–c). We found that the expression level of (a) *LAMA2*, (b) *TIMP4*, and (c) *TMTC1* was higher in all TCGA tumors.

3.10. Effect of Expression Levels of *LAMA2*, *TIMP4*, and *TMTC1* on the Survival of Breast Cancer Patient Oncolnc Analysis

Oncolnc was used to analyse the effect of expression levels of *LAMA2*, *TIMP4*, and *TMTC1* on the survival of breast cancer patients. Results are presented in the (Figure 9). Results showed that all the three genes were significantly ($p < 0.05$) associated with poor survival.

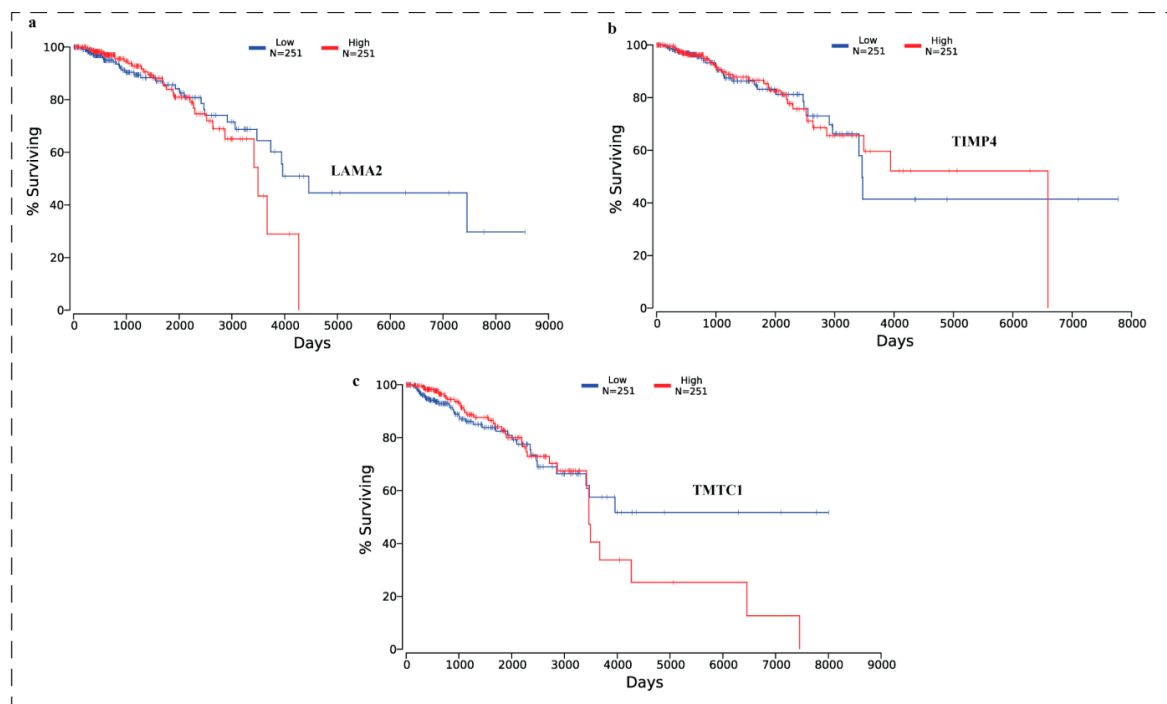


Figure 9. Analysis of the prognostic value of three differentially expressed genes in breast cancer patients using The Cancer Genome Atlas data. All the three genes (a) LAMA2, (b) TIMP4, and (c) TMTC1 were found significantly ($p < 0.05$) associated with poor survival.

3.11. Expression of Key DEGs in Human Breast Cancer Cell Lines Using qRT-PCR

The qRT-PCR analysis showed similar type of expression profiles of key regulatory genes as revealed by the bioinformatics analysis except the two genes namely ADAMTS5 and RSPO3. Results are presented as (Figure 10a–g). The genes namely COL11A1, MMP11 and COL10A1 (Figure 10a–c) were found up-regulated and PCOLCE2, LAMA2, TMTC1, and TIMP4 (Figure 10d–g) were found down-regulated in breast cancer cell lines also. We could not determine the expression of two genes namely ADAMTS5 and RSPO3.

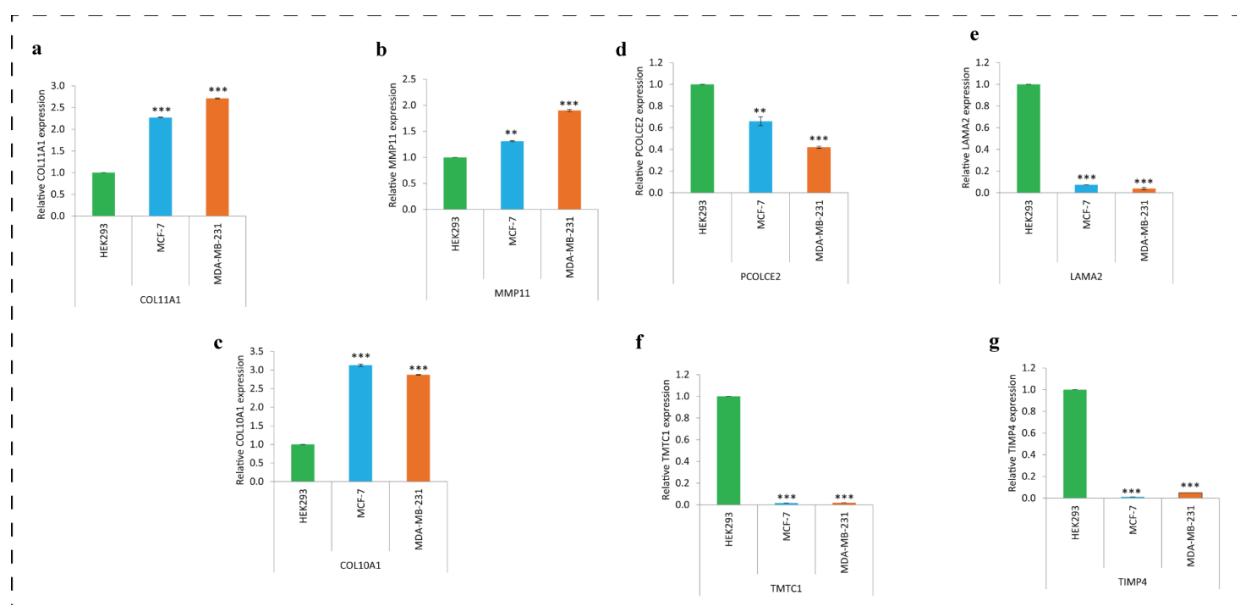


Figure 10. Validation of key regulatory genes in human breast cancer cell lines using qRT-PCR. In the figure, (a–c) presents up-regulated and (d–g) presents down-regulated genes, ** $p < 0.01$, *** $p < 0.001$.

4. Discussion

Breast cancer associated higher mortality rate reflect the need of identification and discovering new biomarkers, therapeutic molecules, and molecular therapeutic targets which will pave for the development of early diagnosis and effective treatment. Identification of imperative gene targets associated with the cancer phenotypes is essential for the development of successful therapy. Currently, at larger scale in-silico techniques are implemented to discover the key regulatory genes. Analysis of gene expression profiles from different databases provides a plinth to quantify and differentiate the gene expression level between normal and tumor samples. Hence, the main aim of this proposed investigation was in-silico identification and in-vitro validation of key regulatory genes associated with breast cancer phenotypes. We conducted an integrative bioinformatics analysis by comparing the normal and breast cancer samples from four transcriptomic datasets. Initially, 161 common differentially expressed genes were extracted from all the four datasets. Among these differentially expressed genes, 44 genes were up-regulated, and 117 genes were found down-regulated. Then we constructed the PPI network, which showed 449 nodes and 18214 edges. To characterize the modular nature, attributes, and organizing principle of the hierarchical network, the activities of the created network were defined at various levels of hierarchy. The modules from the native network along with sub-modules from modules at each level of organization were identified until only motifs remained i.e., unbreakable part of the network. We found that our PPI network is organized hierarchically through seven different levels. We traced these modules to find out the leading hubs or key regulators. Only those that regulate the network from top to bottom (where the PPI network cannot be further divided into sub-community and form motif) are considered as important leading hubs or key regulators. We identified nine key regulators namely *PCOLCE2* (Procollagen C-Endopeptidase Enhancer 2), *LAMA2* (Laminin Subunit Alpha 2), *TMTC1* (transmembrane O-mannosyltransferase targeting cadherins 1), *ADAMTS5* (ADAM Metallopeptidase with Thrombospondin Type 1 Motif 5), *TIMP4* (TIMP-Metallopeptidase Inhibitor 4), *RSPO3* (R-Spondin 3), *COL11A1* (Collagen Type XI Alpha 1 Chain), *MMP11* (matrix metalloproteinase-11), and *COL10A1* (Collagen Type X Alpha 1 Chain). These nine genes were termed as network's key regulators or organizers, as reflected by the PPI network analysis. Among these key genes *COL11A1*, *MMP11* and *COL10A1* were highly expressed and *PCOLCE2*, *LAMA2*, *TMTC1*, *ADAMTS5*, *TIMP4*, and *RSPO3* were having lower expression levels in breast cancer samples. Similar expression profile of seven among nine genes (except *ADAMTS5* and *RSPO3*) was validated in MCF-7 and MDA-MB-231 human breast cancer cell lines. These validation observations reflect that these key genes may pave the way for effective therapeutics of breast cancer. *COL11A1* has been reported markedly associated with head and neck, oral cavity/pharynx, breast, oesophagus, lung, colon, stomach, ovary, and pancreas cancers [39,40]. It has been found well correlated with adverse clinical outcomes in breast cancer [41], recurrence in glioblastoma and ovarian cancer [42,43], and poor survival of kidney and ovarian cancer patients [37,44]. Distinct roles of *MMP11* in cancer development, progression and therapeutics have been reported (reviewed in [45]). *MMP11* has been suggested as a novel target antigen for cancer immunotherapy [46]. High *MMP11* expression has been found associated with poor survival of breast cancer patients [47]. Reports highlight that the expression of *COL10A1* is markedly increased in colon, esophagus, and breast cancer and contributes to cell proliferation, migration, invasion and tumor vasculature [48–50]. Previous studies suggest that expression pattern of *COL10A1* might act as a potential diagnostic predictor for early breast cancer [51]. *PCOLCE2* has been implicated in the colorectal cancer [5] and gastric cancer [52], however, the expression pattern of *PCOLCE2* is poorly understood. Decreased expression of *LAMA2* has been reported in various cancers [53]. *LAMA2* has also been found well correlated with tumor sites and to predict poor survival in pancreatic cancer [54]. *TMTC1* has been also found associated with gastric cancer and has been suggested to act as serve as predictive biomarker for gastric cancer treatment [55]. It is well documented that *ADAMTS5* shows tumor type specific functions. *ADAMTS5* has

been reported to act as a tumor suppressor gene in breast cancer [56] and hepatocellular carcinoma [57]. *TIMP4* has been found associated with breast cancer to modulate the ER- α Signalling in MCF7 Breast Cancer Cells [58]. *RSPO3* has been implicated in ovarian cancer and has been suggested as candidate marker to predict ovarian cancer aggressiveness [59].

Next, we mapped all these key genes to GO analysis. We found that these key regulators were involved in several GO biological processes namely extracellular structure organization, external encapsulating structure organization, extracellular matrix organization, extracellular matrix disassembly, cellular component disassembly, collagen fibril organization, and supramolecular fibre organization. GO Cellular Components associated with key genes include endoplasmic reticulum lumen, intracellular organelle lumen, collagen-containing extracellular matrix, basement membrane, and Golgi lumen. GO Molecular Function analysis showed the key genes were associated with the molecular functions namely metalloendopeptidase activity, metallopeptidase activity, metalloendopeptidase inhibitor activity, and endopeptidase activity. These key regulators were enriched in protein digestion and absorption and viral myocarditis according to the KEGG pathway analysis. All these findings reflect that these key genes are associated with crucial biological functioning and hence can be implicated for the therapeutics of cancer. Further, we found that the genes namely *LAMA2*, *TMTC1*, and *TIMP4* were significantly expressed among different age groups of patients. *LAMA2* and *TIMP4* were found significantly associated and *TMTC1* gene was found less correlated with breast cancer occurrence. We found that the expression level of *LAMA2*, *TIMP4*, and *TMTC1* was higher in all TCGA tumours and significantly associated with poor survival.

While bioinformatics-based transcript profiling is a powerful tool for characterizing the molecular features of breast cancer subtypes, there are limitations associated to this approach. Bioinformatics-based gene expression profiling may reveal distinct gene expression patterns that are associated with different subtypes of breast cancer such as in hormone receptor-positive breast cancer, HER2-positive breast cancer, and Triple-negative breast cancer. Generally, in hormone receptor-positive breast cancer, there is overexpression of estrogen receptor (ER) and/or progesterone receptor (PR) genes, increased expression of genes involved in cell proliferation and cell cycle regulation, such as *Ki-67* and *Cyclin D1*, high expression of genes involved in estrogen signaling, such as *GATA3* and *FOXA1*, and Low expression of genes involved in immune response and inflammation, such as *TNF* and *IL6*. In HER2-positive breast cancer, there is overexpression of the *HER2* gene and other genes in the HER2 signaling pathway, increased expression of genes involved in cell proliferation and survival, such as *MYC* and *BCL2*, high expression of genes involved in DNA repair and genomic stability, such as *BRCA1* and *BRCA2*, and low expression of genes involved in immune response and inflammation, such as *TNF* and *IL6*. In triple-negative breast cancer, there is low expression of hormone receptors (*ER* and *PR*) and *HER2* gene, increased expression of genes involved in cell cycle regulation, such as *Cyclin B1* and *CDC20*, high expression of genes involved in DNA damage repair, such as *RAD51* and *BRCA1*, and high expression of genes involved in immune response and inflammation, such as *IFN γ* and *TNF α* . It's important to note that these transcript profile characteristics in bioinformatics-based gene expression profiling may not be absolute and can vary among individual tumors within a given subtype. In addition, bioinformatics analysis based identified genes could be also general prognostic markers for other types of tumors, but not specific to breast tumors only. Nonetheless, the bioinformatics bases gene expression profiling has helped to identify potential therapeutic targets and develop personalized treatment strategies for many diseases.

5. Conclusions

In conclusion, our study identified nine key regulators, of which *COL11A1*, *MMP11* and *COL10A1* were up regulated and *PCOLCE2*, *LAMA2*, *TMTC1*, *ADAMTS5*, *TIMP4* and *RSPO3* were down regulated in breast cancer samples as compared to control samples. Expression level of *LAMA2*, *TIMP4*, and *TMTC1* was higher in all different stages of TCGA

breast cancer samples and significantly expressed among different age groups of patients (younger to older age group). *LAMA2* and *TIMP4* were significantly associated and *TMTC1* gene was less correlated with breast cancer occurrence. Survival analysis of the genes showed significant association of *LAMA2*, *TIMP4*, and *TMTC1* were significantly associated with poor survival.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomedicines11051271/s1>, Figure S1: A pictorial flow chart of network breakdown.; Figure S2: The behaviors of topological properties: Betweenness (CB(k)), closeness (CC(k)), degree distributions (P(k)) at different levels of organization.; Table S1: List of qRT-PCR primers of key genes.; Table S2: GO Biological Process; Table S3: GO Cellular Component; Table S4: GO Molecular Function; Table S5: KEGG pathways.

Author Contributions: R.A., R.I. and N.J.K. conceptualized the work. R.A. and A.S. did the data curation. R.A. and M.A.P. prepared the figures. R.A., A.S. and R.I. wrote the manuscript. R.A., A.S., M.A.P. and S.H. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The current study was supported by the funding provided to RA as Senior Research Fellow (SRF) by Indian Council of Medical Research (ICMR), Government of India (Ref no- BMI/11/(35)/2020). The authors also express their thanks to MICINN for providing Ramón y Cajal grant to M.A. Prieto (RYC-2017-22891). The funding body had no role in the study design; the collection, analysis, and interpretation of data; or in writing the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Bio-Safety Committee (IBSC), Jamia Millia Islamia (A Central University), New Delhi, India, (Approval Code: P1/7-24.11.20). This project was not found to carry any Biosafety hazards as per the DBT guidelines.

Data Availability Statement: All related data are available within the article.

Acknowledgments: The authors wish to express their thanks to the Indian Council of Medical Research (ICMR), Government of India for providing fellowship to RA as Senior Research Fellow (SRF) (Ref no- BMI/11/(35)/2020). The authors also express their thanks to MICINN for providing Ramón y Cajal grant to M.A. Prieto (RYC-2017-22891).

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

References

1. Arnold, M.; Morgan, E.; Rungay, H.; Mafra, A.; Singh, D.; Laversanne, M.; Vignat, J.; Gralow, J.R.; Cardoso, F.; Siesling, S.; et al. Current and Future Burden of Breast Cancer: Global Statistics for 2020 and 2040. *Breast* **2022**, *66*, 15–23. [[CrossRef](#)] [[PubMed](#)]
2. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
3. Soerjomataram, I.; Bray, F. Planning for Tomorrow: Global Cancer Incidence and the Role of Prevention 2020–2070. *Nat. Rev. Clin. Oncol.* **2021**, *18*, 663–672. [[CrossRef](#)]
4. Lei, S.; Zheng, R.; Zhang, S.; Wang, S.; Chen, R.; Sun, K.; Zeng, H.; Zhou, J.; Wei, W. Global Patterns of Breast Cancer Incidence and Mortality: A Population-based Cancer Registry Data Analysis from 2000 to 2020. *Cancer Commun.* **2021**, *41*, 1183–1194. [[CrossRef](#)] [[PubMed](#)]
5. Yao, H.; Li, C.; Tan, X. An Age Stratified Analysis of the Biomarkers in Patients with Colorectal Cancer. *Sci. Rep.* **2021**, *11*, 22464. [[CrossRef](#)]
6. Francies, F.Z.; Hull, R.; Khanyile, R.; Dlamini, Z. Breast Cancer in Low-Middle Income Countries: Abnormality in Splicing and Lack of Targeted Treatment Options. *Am. J. Cancer Res.* **2020**, *10*, 1568–1591.
7. Wilkinson, L.; Gathani, T. Understanding Breast Cancer as a Global Health Concern. *BJR* **2022**, *95*, 20211033. [[CrossRef](#)]
8. Britt, K.L.; Cuzick, J.; Phillips, K.-A. Key Steps for Effective Breast Cancer Prevention. *Nat. Rev. Cancer* **2020**, *20*, 417–436. [[CrossRef](#)]
9. Zhang, Y.; Zhou, Y.; Mao, F.; Yao, R.; Sun, Q. Ki-67 Index, Progesterone Receptor Expression, Histologic Grade and Tumor Size in Predicting Breast Cancer Recurrence Risk: A Consecutive Cohort Study. *Cancer Commun.* **2020**, *40*, 181–193. [[CrossRef](#)]
10. Momenimovahed, Z.; Salehiniya, H. Epidemiological Characteristics of and Risk Factors for Breast Cancer in the World. *BCTT* **2019**, *11*, 151–164. [[CrossRef](#)]

11. Jarada, T.N.; Rokne, J.G.; Alhajj, R. A Review of Computational Drug Repositioning: Strategies, Approaches, Opportunities, Challenges, and Directions. *J. Cheminform.* **2020**, *12*, 46. [[CrossRef](#)] [[PubMed](#)]
12. Alam, A.; Abubaker Bagabir, H.; Sultan, A.; Siddiqui, M.F.; Imam, N.; Alkhanani, M.F.; Alsulimani, A.; Haque, S.; Ishrat, R. An Integrative Network Approach to Identify Common Genes for the Therapeutics in Tuberculosis and Its Overlapping Non-Communicable Diseases. *Front. Pharmacol.* **2022**, *12*, 770762. [[CrossRef](#)] [[PubMed](#)]
13. Kamel, H.F.M.; Al-Amodi, H.S.A.B. Exploitation of Gene Expression and Cancer Biomarkers in Paving the Path to Era of Personalized Medicine. *Genom. Proteom. Bioinform.* **2017**, *15*, 220–235. [[CrossRef](#)]
14. Mishra, A.; Verma, M. Cancer Biomarkers: Are We Ready for the Prime Time? *Cancers* **2010**, *2*, 190–208. [[CrossRef](#)]
15. Zarei Ghobadi, M.; Emamzadeh, R. Integration of Gene Co-Expression Analysis and Multi-Class SVM Specifies the Functional Players Involved in Determining the Fate of HTLV-1 Infection toward the Development of Cancer (ATLL) or Neurological Disorder (HAM/TSP). *PLoS ONE* **2022**, *17*, e0262739. [[CrossRef](#)] [[PubMed](#)]
16. Zeng, X.; Shi, G.; He, Q.; Zhu, P. Screening and Predicted Value of Potential Biomarkers for Breast Cancer Using Bioinformatics Analysis. *Sci. Rep.* **2021**, *11*, 20799. [[CrossRef](#)]
17. You, G.; Zu, B.; Wang, B.; Fu, Q.; Li, F. Identification of MiRNA–MRNA–TFs Regulatory Network and Crucial Pathways Involved in Tetralogy of Fallot. *Front. Genet.* **2020**, *11*, 552. [[CrossRef](#)]
18. Doncheva, N.T.; Assenov, Y.; Domingues, F.S.; Albrecht, M. Topological Analysis and Interactive Visualization of Biological Networks and Protein Structures. *Nat. Protoc.* **2012**, *7*, 670–685. [[CrossRef](#)]
19. Albert, R.; Barabási, A.-L. Statistical Mechanics of Complex Networks. *Rev. Mod. Phys.* **2002**, *74*, 47–97. [[CrossRef](#)]
20. Barabási, A.-L.; Albert, R. Emergence of Scaling in Random Networks. *Science* **1999**, *286*, 509–512. [[CrossRef](#)]
21. Brandes, U. A Faster Algorithm for Betweenness Centrality*. *J. Math. Sociol.* **2001**, *25*, 163–177. [[CrossRef](#)]
22. Mason, O.; Verwoerd, M. Graph Theory and Networks in Biology. *IET Syst. Biol.* **2007**, *1*, 89–119. [[CrossRef](#)] [[PubMed](#)]
23. Canright, G.; Engø-Monsen, K. Roles in Networks. *Sci. Comput. Program.* **2004**, *53*, 195–214. [[CrossRef](#)]
24. Farooqui, A.; Tazyeen, S.; Ahmed, M.M.; Alam, A.; Ali, S.; Malik, M.Z.; Ali, S.; Ishrat, R. Assessment of the Key Regulatory Genes and Their Interologs for Turner Syndrome Employing Network Approach. *Sci. Rep.* **2018**, *8*, 10091. [[CrossRef](#)] [[PubMed](#)]
25. Newman, M.E.J. Finding Community Structure in Networks Using the Eigenvectors of Matrices. *Phys. Rev. E* **2006**, *74*, 036104. [[CrossRef](#)]
26. Newman, M.E.J.; Girvan, M. Finding and Evaluating Community Structure in Networks. *Phys. Rev. E* **2004**, *69*, 026113. [[CrossRef](#)]
27. Csardi, G.; Nepusz, T. The Igraph Software Package for Complex Network Research. *InterJournal Complex Syst.* **2006**, *1695*, 1–9.
28. Ud-Dean, S.M.M.; Heise, S.; Klamt, S.; Gunawan, R. TRaCE+: Ensemble Inference of Gene Regulatory Networks from Transcriptional Expression Profiles of Gene Knock-out Experiments. *BMC Bioinform.* **2016**, *17*, 252. [[CrossRef](#)]
29. Fan, S.; Liang, Z.; Gao, Z.; Pan, Z.; Han, S.; Liu, X.; Zhao, C.; Yang, W.; Pan, Z.; Feng, W. Identification of the Key Genes and Pathways in Prostate Cancer. *Oncol. Lett.* **2018**, *16*, 6663–6669. [[CrossRef](#)]
30. Liu, M.; Xu, Z.; Du, Z.; Wu, B.; Jin, T.; Xu, K.; Xu, L.; Li, E.; Xu, H. The Identification of Key Genes and Pathways in Glioma by Bioinformatics Analysis. *J. Immunol. Res.* **2017**, *2017*, 1–9. [[CrossRef](#)]
31. Ning, P.; Wu, Z.; Hu, A.; Li, X.; He, J.; Gong, X.; Xia, Y.; Shang, Y.; Bian, H. Integrated Genomic Analyses of Lung Squamous Cell Carcinoma for Identification of a Possible Competitive Endogenous RNA Network by Means of TCGA Datasets. *PeerJ* **2018**, *6*, e4254. [[CrossRef](#)] [[PubMed](#)]
32. Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **1999**, *27*, 29–34. [[CrossRef](#)] [[PubMed](#)]
33. Dennis, G.; Sherman, B.T.; Hosack, D.A.; Yang, J.; Gao, W.; Lane, H.C.; Lempicki, R.A. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **2003**, *4*, P3. [[CrossRef](#)] [[PubMed](#)]
34. Salavaty, A.; Rezvani, Z.; Najafi, A. Survival Analysis and Functional Annotation of Long Non-coding RNAs in Lung Adenocarcinoma. *J. Cell. Mol. Med.* **2019**, *23*, 5600–5617. [[CrossRef](#)]
35. Chandrashekar, D.S.; Bashel, B.; Balasubramanya, S.A.H.; Creighton, C.J.; Ponce-Rodriguez, I.; Chakravarthi, B.V.S.K.; Varambally, S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* **2017**, *19*, 649–658. [[CrossRef](#)]
36. Thumkeo, D.; Watanabe, S.; Narumiya, S. Physiological Roles of Rho and Rho Effectors in Mammals. *Eur. J. Cell Biol.* **2013**, *92*, 303–315. [[CrossRef](#)]
37. Boguslawska, J.; Kedzierska, H.; Poplawski, P.; Rybicka, B.; Tanski, Z.; Piekielko-Witkowska, A. Expression of Genes Involved in Cellular Adhesion and Extracellular Matrix Remodeling Correlates with Poor Survival of Patients with Renal Cancer. *J. Urol.* **2016**, *195*, 1892–1902. [[CrossRef](#)]
38. Bonacich, P. Power and Centrality: A Family of Measures. *Am. J. Sociol.* **1987**, *92*, 1170–1182. [[CrossRef](#)]
39. Raglow, Z.; Thomas, S.M. Tumor Matrix Protein Collagen XI α 1 in Cancer. *Cancer Lett.* **2015**, *357*, 448–453. [[CrossRef](#)] [[PubMed](#)]
40. Vázquez-Villa, F.; García-Ocaña, M.; Galván, J.A.; García-Martínez, J.; García-Pravia, C.; Menéndez-Rodríguez, P.; Rey, C.G.; Barneo-Serra, L.; de los Toyos, J.R. COL11A1/(pro)Collagen 11A1 Expression Is a Remarkable Biomarker of Human Invasive Carcinoma-Associated Stromal Cells and Carcinoma Progression. *Tumor Biol.* **2015**, *36*, 2213–2222. [[CrossRef](#)]
41. Farmer, P.; Bonnefoi, H.; Anderle, P.; Cameron, D.; Wirapati, P.; Becette, V.; André, S.; Piccart, M.; Campone, M.; Brain, E.; et al. A Stroma-Related Gene Signature Predicts Resistance to Neoadjuvant Chemotherapy in Breast Cancer. *Nat. Med.* **2009**, *15*, 68–74. [[CrossRef](#)] [[PubMed](#)]

42. Cheng, W.-Y.; Kandel, J.J.; Yamashiro, D.J.; Canoll, P.; Anastassiou, D. A Multi-Cancer Mesenchymal Transition Gene Expression Signature Is Associated with Prolonged Time to Recurrence in Glioblastoma. *PLoS ONE* **2012**, *7*, e34705. [[CrossRef](#)] [[PubMed](#)]
43. Cheon, D.-J.; Tong, Y.; Sim, M.-S.; Dering, J.; Berel, D.; Cui, X.; Lester, J.; Beach, J.A.; Tighiouart, M.; Walts, A.E.; et al. A Collagen-Remodeling Gene Signature Regulated by TGF- β Signaling Is Associated with Metastasis and Poor Survival in Serous Ovarian Cancer. *Clin. Cancer Res.* **2014**, *20*, 711–723. [[CrossRef](#)] [[PubMed](#)]
44. Wu, Y.-H.; Chang, T.-H.; Huang, Y.-F.; Huang, H.-D.; Chou, C.-Y. COL11A1 Promotes Tumor Progression and Predicts Poor Clinical Outcome in Ovarian Cancer. *Oncogene* **2014**, *33*, 3432–3440. [[CrossRef](#)]
45. Zhang, X.; Huang, S.; Guo, J.; Zhou, L.; You, L.; Zhang, T.; Zhao, Y. Insights into the Distinct Roles of MMP-11 in Tumor Biology and Future Therapeutics (Review). *Int. J. Oncol.* **2016**, *48*, 1783–1793. [[CrossRef](#)]
46. Peruzzi, D.; Mori, F.; Conforti, A.; Lazzaro, D.; De Rinaldis, E.; Ciliberto, G.; La Monica, N.; Aurisicchio, L. MMP11: A Novel Target Antigen for Cancer Immunotherapy. *Clin. Cancer Res.* **2009**, *15*, 4104–4113. [[CrossRef](#)]
47. Kim, H.S.; Kim, M.G.; Min, K.-W.; Jung, U.S.; Kim, D.-H. High MMP-11 Expression Associated with Low CD8+ T Cells Decreases the Survival Rate in Patients with Breast Cancer. *PLoS ONE* **2021**, *16*, e0252052. [[CrossRef](#)]
48. Chapman, K.B.; Prendes, M.J.; Sternberg, H.; Kidd, J.L.; Funk, W.D.; Wagner, J.; West, M.D. COL10A1 Expression Is Elevated in Diverse Solid Tumor Types and Is Associated with Tumor Vasculature. *Future Oncol.* **2012**, *8*, 1031–1040. [[CrossRef](#)]
49. Huang, H.; Li, T.; Ye, G.; Zhao, L.; Zhang, Z.; Mo, D.; Wang, Y.; Zhang, C.; Deng, H.; Li, G.; et al. High Expression of COL10A1 Is Associated with Poor Prognosis in Colorectal Cancer. *OTT* **2018**, *11*, 1571–1581. [[CrossRef](#)]
50. Li, T.; Huang, H.; Shi, G.; Zhao, L.; Li, T.; Zhang, Z.; Liu, R.; Hu, Y.; Liu, H.; Yu, J.; et al. TGF-B1-SOX9 Axis-Inducible COL10A1 Promotes Invasion and Metastasis in Gastric Cancer via Epithelial-to-Mesenchymal Transition. *Cell Death Dis* **2018**, *9*, 849. [[CrossRef](#)]
51. Giussani, M.; Landoni, E.; Merlino, G.; Turdo, F.; Veneroni, S.; Paolini, B.; Cappelletti, V.; Miceli, R.; Orlandi, R.; Triulzi, T.; et al. Extracellular Matrix Proteins as Diagnostic Markers of Breast Carcinoma. *J. Cell Physiol.* **2018**, *233*, 6280–6290. [[CrossRef](#)] [[PubMed](#)]
52. Xiang, A.; Lin, X.; Xu, L.; Chen, H.; Guo, J.; Zhou, F. PCOLCE Is Potent Prognostic Biomarker and Associates With Immune Infiltration in Gastric Cancer. *Front. Mol. Biosci.* **2020**, *7*, 544895. [[CrossRef](#)] [[PubMed](#)]
53. Akhavan, A.; Griffith, O.L.; Soroceanu, L.; Leonoudakis, D.; Luciani-Torres, M.G.; Daemen, A.; Gray, J.W.; Muschler, J.L. Loss of Cell-Surface Laminin Anchoring Promotes Tumor Growth and Is Associated with Poor Clinical Outcomes. *Cancer Res.* **2012**, *72*, 2578–2588. [[CrossRef](#)] [[PubMed](#)]
54. Zhang, M.; Zeng, L.; Peng, Y.; Fan, B.; Chen, P.; Liu, J. Immune-Related Genes LAMA2 and IL1R1 Correlate with Tumor Sites and Predict Poor Survival in Pancreatic Adenocarcinoma. *Future Oncol.* **2021**, *17*, 3061–3076. [[CrossRef](#)] [[PubMed](#)]
55. Chen, X.; Zhang, Q.; Chekouo, T. Filtering High-Dimensional Methylation Marks With Extremely Small Sample Size: An Application to Gastric Cancer Data. *Front. Genet.* **2021**, *12*, 705708. [[CrossRef](#)]
56. Porter, S.; Scott, S.D.; Sassoon, E.M.; Williams, M.R.; Jones, J.L.; Girling, A.C.; Ball, R.Y.; Edwards, D.R. Dysregulated Expression of Adamalysin-Thrombospondin Genes in Human Breast Carcinoma. *Clin. Cancer Res.* **2004**, *10*, 2429–2440. [[CrossRef](#)]
57. Li, C.; Xiong, Y.; Yang, X.; Wang, L.; Zhang, S.; Dai, N.; Li, M.; Ren, T.; Yang, Y.; Gan, L.; et al. Lost Expression of ADAMTS5 Protein Associates with Progression and Poor Prognosis of Hepatocellular Carcinoma. *DDDT* **2015**, *9*, 1773. [[CrossRef](#)]
58. Lizarraga, F.; Espinosa, M.; Ceballos-Cancino, G.; Vazquez-Santillan, K.; Bahena-Ocampo, I.; Schwarz-Cruz y Celis, A.; Vega-Gordillo, M.; Garcia Lopez, P.; Maldonado, V.; Melendez-Zajgla, J. Tissue Inhibitor of Metalloproteinases-4 (TIMP-4) Regulates Stemness in Cervical Cancer Cells: TIMP-4 MODULATES CERVICAL CANCER CELLS STEMNESS. *Mol. Carcinog.* **2016**, *55*, 1952–1961. [[CrossRef](#)]
59. Gu, H.; Tu, H.; Liu, L.; Liu, T.; Liu, Z.; Zhang, W.; Liu, J. RSPO3 Is a Marker Candidate for Predicting Tumor Aggressiveness in Ovarian Cancer. *Ann. Transl. Med.* **2020**, *8*, 1351. [[CrossRef](#)]

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