Identification of Key Genes in Breast Cancer Using Bioinformatics Analysis

A Project Work Submitted to the Department of Statistics of Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh. In partial fulfillment of requirements for the degree of Bachelor of Science (Honors) in Statistics

Submitted By

Md Amanat Ullah Arman
ID: 17STA003
Year: 4th
Semester: 2nd
Session: 2017-18

Tasnia Akter Maya
ID: 17STA064
Year: 4th
Semester: 2nd
Session: 2017-18

Supervised By

Md. Matiur Rahaman, PhD
Assistant Professor
Department of Statistics,
Bangabandhu Sheikh Mujibur
Rahman Science and Technology
University

Department of Statistics

Bangabandhu Sheikh Mujibur Rahman Science and Technology University,
Gopalganj-8100, Bangladesh.

July 2023

Identification of Key Genes in Breast Cancer Using Bioinformatics Analysis

A Project Work Submitted to the Department of Statistics of Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh. In partial fulfillment of requirements for the degree of Bachelor of Science (Honors) in Statistics

Submitted By

Md Amanat Ullah Arman
ID: 17STA003
Year: 4th
Semester: 2nd
Session: 2017-18

Tasnia Akter Maya
ID: 17STA064
Year: 4th
Semester: 2nd
Session: 2017-18

Supervised By

Md. Matiur Rahaman, PhD
Assistant Professor
Department of Statistics,
Bangabandhu Sheikh Mujibur
Rahman Science and Technology
University

Department of Statistics

Bangabandhu Sheikh Mujibur Rahman Science and Technology University,
Gopalganj-8100, Bangladesh.

July 2023

Table of Contents

Table of Contents i

List of Tables ii

List of Figures ii

List of Abbreviations iii

Abstract iv

Acknowledgment v

Chapter 1: Introduction 1

1.1 Background of the Study 2

1.2 Basic Information about Breast Cancer 4

1.2.1 Breast Cancer 4

1.3 Risk Factors for Breast Cancer 5

1.4 Symptoms of Breast Cancer 7

1.5 Causes of Breast Cancer 7

1.6 Breast Cancer Stages 8

1.7 Diagnosis of Breast Cancer 10

1.8 Prevention of Breast Cancer 10

1.9 Treatment Options for Breast Cancer 11

1.10 Objectives of the Project 11

1.11 Problems Identification for the Project 11

1.12 Limitations of the Project 11

1.13 Layout 12

Chapter 2: Materials and methods 13

2.1. Data sources and descriptions 14

2.2 Identification of DEGs 14

2.2.1 Microarray data 14

2.2.2 GEO2R 14

2.2.3 Bioconductor 15

2.2.4 Enter a Series accession number 15

2.2.5 Define Sample groups 16

2.2.6 Perform the analysis 16

2.3. Protein-protein interaction (PPI) network analysis based on validated DEGs 16

2.3.1 Cytoscape 16

2.3.2 CytoHubba 17

2.4 Expression Analyses of the Hub Genes in Breast Cancer 18
2.5. Functional and pathway enrichment analysis of DEGs highlighting cKGs 19
Chapter 3: Result 20
3.1 Identification of DEGs for BC Patients 21
3.2 PPI network analysis of DEGs and identification of cKGs 23
3.3 Expression Analyses of the Hub Genes in Breast Cancer 26
3.4 GO and KEGG pathway enrichment analysis of DEGs highlighting cKGs 27
Chapter 4: Discussion and Conclusion 29
4.1 Discussion 30
4.2 Conclusion 33
4.3 Areas of Further Research 33
References 34
APPENDIX 38

List of Tables

Table 1: Selection of cKGs 25
Table 2: Top 14 in network string ranked by Degree, MCC and MNC method 26

Table A1: Biological Process (BP) 38
Table A2: Cellular Componenet (CC) 41
Table A3: Mollecular Function (MF) 46
Table A4: KEGG Pathways 49

List of Figures

Figure 1.1: Snap Shot of GEO2R 16

Figure 3.1: Screening of the overlapping DEGs GSE42568 datasets 21
Figure 3.2: Mean difference (MD) plot of DEGs. 22
Figure 3.3: Heatmaps of the DEGs. 22
Figure 3.4: PPI network analysis results of DEGs for identification cKGs 24
Figure 3.5: Expression and survival analyses of cKGs in BC. 27

List of Abbreviations

BC- Breast Cancer

KGs-Key Genes

NCBI- National Center of Biotechnology Information

LIMMA-Linear Models for Microarray Data

GEO-Gene Expression Omnibus

DEGs-Differential Expression Genes

PPI- Protein-protein interaction

cHubGs – Common Hub Genes

cKGs – Common Key Genes

MNC-maximum Neighborhood Component

MCC- Maximal Clique Centrality

GO- Gene Ontology

KEGG- Kyoto encyclopedia of genes and genomes

BP-Biological Process

MF-Molecular Function

CC-Cellular Component

Abstract

Breast cancer (BC) is one of the most malignant tumors and the leading cause of cancer-related death in women worldwide. So, an in-depth investigation on the molecular mechanisms of BC progression is required for diagnosis, and prognosis. In this study, we identified 1528 up-regulated and 1749 down-regulated differentially expressed genes (DEGs) by analyzing gene expression profiles with NCBI accession numbers GSE42568 using GEO2R web tool. Then we select top 200 up-regulated and 200 down- regulated genes between BC and control. Then we constructed protein-protein interaction (PPI) network of though the STRING database and identified 8 cHubGs (FN1,CCNB1,CCNB2,ASPM,CENPF, ITGB4, SOX9, NT5E) as a set of key genes (KGs) by cytoHubba plugin in Cytoscape. We validation the expression of cHubGs of BC. Then we investigated the pathogenetic processes of DEGs highlighting cKGs by GO terms and KEGG pathway enrichment analysis. Several BC-causing crucial biological processes, molecular functions, cellular components, and pathways were significantly enriched by the estimated DEGs including at-least one cKGs. Therefore, the findings of this study might be useful resources for BC diagnosis, and prognosis.

Acknowledgment

At first we would like to express our gratitude to Almighty Allah, source of all power and knowledge, for giving us strength and ability to accomplish our project work and also thank to our family members, who always beside us for better supports, encouragement and advice to go ahead.

We also like to express our deep appreciation to our supervisor for his valuable time and brilliant contribution to this study until its completion.

Our exclusive thank to all of our honorable teachers, department of statistics, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj-8100, for their valuable suggestions and comments.

Chapter 1

Introduction

Chapter 1
Introduction

1.1 Background of the Study

A type of cancer that forms in the cells of the breast tissue is known as breast cancer. It occurs when the cells in the breast tissue begin to grow uncontrollably and form a tumor[1]. Breast cancer can develop in both men and women, although it is more commonly found in women. There are different types of breast cancer, and it can also occur in different parts of the breast. Early detection and treatment of breast cancer can improve the chances of successful treatment and recovery.

Breast cancer is a serious and potentially life-threatening disease that affects the breast tissue. It is the most common cancer diagnosed in women worldwide, but it can also affect men[2].

Breast cancer develops when cells in the breast tissue grow and divide uncontrollably. This abnormal growth can form a lump or tumor, which may be felt as a lump in the breast or detected by a mammogram. Breast cancer can also spread to other parts of the body, such as the lymph nodes, bones, or liver.

There are several types of breast cancer, including ductal carcinoma, lobular carcinoma, and inflammatory breast cancer. Some types of breast cancer are hormone-sensitive, meaning they grow in response to hormones like estrogen and progesterone. Genetic factors, such as the BRCA1 and BRCA2 genes, can also increase the risk of developing breast cancer[3].

Symptoms of breast cancer can include a lump or thickening in the breast, changes in breast size or shape, nipple discharge or inversion, and skin changes on the breast, such as redness or dimpling[4].

Early detection is important for successful treatment and recovery from breast cancer. Women are encouraged to perform regular breast self-exams, receive clinical breast exams, and have regular mammograms as recommended by their healthcare provider. Treatment for breast cancer may include surgery, radiation therapy, chemotherapy, or targeted therapy, depending on the type and stage of the cancer.

Percent of Breast Cancer Deaths by Age (2010-2014) According to the SEER statistics, between the years of 2010 and 2014, the average age of diagnosis of breast cancer in women is most often diagnosed between the ages of 55 and 64 years, Aug 26, 2018[5].

Overweight and obesity, a global phenomenon, affects more than 1 billion adults, with 300 million being clinically obese. Obesity has a major impact on the risk and prognosis of some of the more common forms of cancer, but also provides us with one of the few preventive interventions capable of making a significant impact on the cancer problem. Weight increase and obesity in menopausal females have been identified as the most important prognostic risk factors for breast cancer in postmenopausal women. Several studies have reported that at diagnosis of breast cancer, obese women exhibit an increase in lymph-node involvement and a higher propensity to develop distant metastases.

Developing breast cancer when you're a teenager is extremely rare. It's also uncommon in women in their 20s and 30s. The vast majority of breast cancers are diagnosed in women over the age of 50[6].

"Breast cancer is not a fatal in and of itself “said Zuckerman "what makes it fatal is if it goes into other parts of the body and gets into the lymph nodes, lungs, and other organs, "she said." also if it gets into the blood or the bones, it can kill a person that's the risk a person metastasized cancer. The average 5-year survival rate for people with breast cancer is 90%. The average 10-year survival rate is 83% if the cancer is located only in the breast; the 5-year relative survival rate of people with breast cancer is 99%. 62% of case are diagnosed at this stage.

Breast cancer is the most common cancer in women worldwide, with nearly 1.7 million new cases diagnosed in 2012. Around the world represented about 12% of all new cancer and 25% off all cancers in women. Breast cancer has ranked number one cancer among Indian females with age adjusted rate as high as 25.8 per 100,000 and mortality 12.7 per 100,000 women. Data reports from various latest national Cancer registries were compared for incidence mortality rates.

In 2018, an estimated 1,735,350 new cases of cancer will be diagnosed in the United States and 609,640 people will die from the disease [7].

Belgium is most highly affected country in the world age specific rate per 100,000 in worldwide is 111.9, and Denmark, France, Netherlands are 105.0, 104.5, 99.0

respectively. Exercise is a compelling methodology to enhance personal satisfaction and physical wellness in bosom disease survivors; be that as it may, few examinations have concentrated on the early survivorship time frame, minorities, physically idle and stout ladies, or tried a consolidated exercise program and estimated bone wellbeing. Here, we report the impacts of a 16-week high-impact and opposition practice intercession on patient revealed results, physical wellness, and bone wellbeing in ethnically assorted, physically latent, overweight or hefty bosom malignancy survivors. Breast cancer begins when cells in the bosom start to develop wild. These cells ordinarily frame a tumor that can frequently be seen on a x-beam or felt as a protuberance. The tumor is threatening (disease) if the cells can develop into (attack) encompassing tissues or spread (metastasize) to far off regions of the body. Bosom growth happens as a rule in ladies.

1.2 Basic Information about Breast Cancer

1.2.1 Breast Cancer

Worldwide, Breast cancer is the most common type of cancer in women and the second highest in terms of mortality rates. Men can also develop breast cancer, although incidences are rare. About 90% of all breast cancer cases start in the tissue of the milk ducts or in the lobules the supply milk to the ducts. It is possible to diagnose breast cancer at an early stage. The early detection of cancer greatly increases the chances of successful treatment.

There are several types of breast cancer, which are broken into two main categories: “invasive” and “noninvasive” or in situ. While invasive cancer has separated from the breast ducts or glands to other part of the breast noninvasive cancer has not spread from the original tissue.

These two categories are used to describe the most common type of breast cancer which include:

Ductal carcinoma in situ: Ductal carcinoma in situ(DCIS) is a noninvasive condition. With DCIS, the cells that line the ducts in your breast change and look cancerous. However, DCIS cells haven’t invaded the surrounding breast tissue[8].

Lobular carcinoma in situ: Lobular carcinoma in situ (LCIS) is cancer that grows in the milk- producing glands of your breast. Like DCIS, the cancer cells haven’t yet invaded the surrounding tissue.[9]

Invasive ductal carcinoma: Invasive ductal carcinoma (lDC) is the most common type of breast cancer. This type of breast cancer begins in your breast's milk ducts and then invades nearby tissue in the breast. Once the breast cancer has spread to the tissue outside your milk ducts, it can begin to spread to other nearby organs and tissue.

Invasive lobular carcinoma: If breast cancer is diagnosed as ILC, it has already spread to nearby tissue and organs. There are also some less common types of breast cancer include.

Paget disease of the nipple: These type of breast cancer begins in the breast ducts, but yes it grows, it beings to effect the skin and areola of the nipple.

Phyllodes tumor: This is very real type of press cancer grows in the connective tissue of the breast.

Angiosarcoma: This is cancer that grows on the blood vessel or lymph vessels in the breast.

1.3 Risk Factors for Breast Cancer

A breast cancer risk factor is anything that makes it more likely you'll get breast cancer. But having one or even several breast cancer risk factors doesn’t necessarily mean you'll develop breast cancer. Many women who develop breast cancer have no known risk factors other than simply being women.

Factors that are associated with an increased risk of breast cancer include:

Being female: Women are much more likely than men are to develop breast cancer.

Increasing age: Your risk of breast cancer increases as you age.

A personal history of breast conditions: If you've had a breast biopsy that found lobular carcinoma in situ (LCIS) or atypical hyperplasia of the breast, you have an increased risk of breast cancer.

A personal history of breast cancer: If you've had breast cancer in one breast, you have an increased risk of developing cancer in the other breast.

A family history of breast cancer: If your mother, sister or daughter was diagnosed with breast cancer, particularly at a young age, your risk of breast cancer is increased still, the majority of people diagnosed with breast cancer have no family history of the disease.

Inherited genes that increase cancer risk: Certain gene mutations that increase the risk of breast cancer can be passed from parents to children. The most well-known gene mutations are referred to as BRCA1 and BRCA2. These genes can greatly increase your risk of breast cancer and other cancers, but they don't make cancer inevitable.

Radiation exposure: If you received radiation treatments to your chest as a child or young adult, your risk of breast cancer is increased[10].

Obesity: Being obese increases your risk of breast cancer[11].

Beginning your period at a younger age: Beginning your period before age 12 increases your risk of breast cancer[12].

Beginning menopause at an older age: If you began menopause at an older age, you're more likely to develop breast cancer.

Having your first child at an older age: Women who give birth to their first child after age 30 may have an increased risk of breast cancer.

Having never been pregnant: Women who have never been pregnant have a greater risk of breast cancer than do women who have had one or more pregnancies.

Postmenopausal hormone therapy: Women who take hormone therapy medications that combine estrogen and progesterone to treat the signs and symptoms of menopause have an increased risk of breast cancer. The risk of breast cancer decreases when women stop taking these medications.

Drinking alcohol: Drinking alcohol increases the risk of breast cancer[13].

1.4 Symptoms of Breast Cancer

In some cases, women with breast cancer may have no symptoms, while some abnormalities they experience may not be cancerous. However, it is important to seek medical attention when the following symptoms appear.

A breast lump or tissue thickening that feels different than surrounding tissue and has developed recently.

a) Breast pain.

b) Red, pitted skin over your entire breast.

c) Swelling in all or part of your breast.

d) A nipple discharge other than breast milk.

e) Bloody discharge from your nipple.

f) Peeling, scaling, or flaking of skin on your nipple or breast.

g) A sudden, unexplained change in the shape or size of your breast.

h) Inverted nipple

i) Changes to the appearance of the skin on your breasts.

j) A lump or swelling under your arm.

k) Irritation or dimpling of breast skin.

l) Redness or flaky skin in the nipple area or the breast.

m) Nipple discharge other than breast milk, including blood.

n) Any change in the size or the shape of the breast.

o) Pain in any area of the breast

1.5 Causes of Breast Cancer

Doctors know that breast cancer occurs when some breast cells begin to grow abnormally. These cells divide more rapidly than healthy cells do and continue to accumulate, forming a lump or mass Cells may spread (metastasize) through your breast to your lymph nodes or to other parts of your body.

Breast cancer most often begins with cells in the milk-producing ducts (invasive ductal carcinoma) Breast cancer may also begin in the glandular tissue called lobules (invasive lobular carcinoma) or in other cells or tissue within the breast[14].

Researchers have identified hormonal, lifestyle and environmental factors that may increase your risk of breast cancer. But it's not clear why some people who have no risk factors develop cancer, yet other people with risk factors never do. It's likely that breast cancer is caused by a complex interaction of your genetic makeup and your environment.

Several studies are looking at the effect of exercise, weight gain or loss, and diet on risk[15].

Studies on the best use of genetic testing for breast cancer mutations continue at a rapid pace[16].

Scientists are exploring how common gene variations (small changes in genes that are not as significant as mutations) may affect breast cancer risk. Gene variants typically have only a modest effect on risk, but when taken together they could possibly have a large impact.

Possible environmental causes of breast cancer have also received more attention in recent years. While much of the science on this topic is still in its earliest stages, this is an area of active research[17].

1.6 Breast Cancer Stages

Breast cancer can be divided into stages based on how severe it is. Cancers that have grown and invaded nearby tissues and organs are at a higher stage than cancers that are still contained to the breast. Breast cancer has five main stages. They are given below:

Stage 0 breast cancer:

Stage 0 is DCIS Cancer cells in DCIS remain confined to the ducts in the breast and have not spread into nearby tissue.

Stage 1 breast cancer:

There are two types of stage I breast cancer:

a**) Stage 1A:** The primary tumor is 2 centimeters wide or less and the lymph nodes are not affected.

b**) Stage 1B:** Cancer is found in nearby lymph nodes, and either there is no tumor in the breast, or the tumor is smaller than 2 centimeters.

Stage 2 breast cancer:

Stage 2 breast cancers are also divided into two categories-

a) **Stage 2A:** The tumor is smaller than 2 centimeters and has spread to 1-3 nearby lymph nodes, or it's between 2 and 5 centimeters and hasn't spread to any lymph nodes.

b**) Stage 2B:** The tumor is between 2 and 5 centimeters and has spread to 1-3 axillary (armpit) lymph nodes, or it's larger than 5 centimeters and hasn't spread to any lymph nodes.

Stage 3 breast cancer:

There are three main types of stage 3 breast cancer.

a**) Stage 3A:** This stage can have several types of cancer:

The cancer has spread to 4-9 axillary lymph nodes or has enlarged the internal mammary lymph nodes, and the primary tumor can be any size.

The tumor is bigger than 5 centimeters and small groups of cancer cells are found in the lymph nodes.

Tumors are greater than 5 centimeters and the cancer has spread to 1-3 axillary lymph nodes or any breastbone nodes.

b) **Stage 3B:** A tumor has invaded the chest wall or skin and may or may not have invaded unto 9 lymph nodes.

c**) Stage 3C:** Cancer is found in 10 or more axillary lymph nodes, lymph nodes near the collarbone, or internal mammary nodes.

Stage 4 breast cancer

Stage 4 breast cancer can have a tumor of any size, and its cancer cells have spread to nearby and distant lymph nodes, as well as distant organs.

1.7 Diagnosis of Breast Cancer

Diagnosis of breast cancer is performed when an abnormal lump is found (from self-examination or x-ray) or a tiny speck of calcium is seen (of an x-ray)[18].

The method of cancer diagnosis varies depending on several factors, such as age, current medications, type of cancer, severity of symptoms and earlier test results Diagnosis of breast cancer can be performed in the following ways:

Diagnostic Radiology

a) Diagnostic Mammography

b) Ultrasound

c) MRI

[1] Biopsy

[II]Surgical Pathology

[III] Blood Test

Additional Tests

a) Chest X-ray

b) Bone Scan

c) CT scan to generate 3D images of the organs to check for the spread of cancer

1.8 Prevention of Breast Cancer

Some preventions of breast cancer are listed below:

a) Making changes in our daily life may help reduce our risk of breast cancer

b) Ask our doctor about breast cancer screening. Discuss with our doctor when to begin breast cancer screening exams and tests, such as clinical breast exams and mammograms

c) Become familiar with our breasts through breast self-exam for breast awareness

d) Drink alcohol in moderation, if at all. Limit the amount of alcohol we drink to no more than one drink a day, if we choose to drink.

e) Exercise most days of the week

f) Limit postmenopausal hormone therapy. Combination hormone therapy may increase the risk of breast cancer

g) Maintain a healthy weight.

h) Choose a healthy diet.

1.9 Treatment Options for Breast Cancer

There are several treatments for breast cancer, like as

a) Surgery

b) Radiotherapy

c) Chemotherapy

d) Hormonal Therapy

e) Targeted Therapy

1.10 Objectives of the Project

The main objectives are given below:

a) Computational identification of genomic biomarkers for BC.

b) To perform the pathways analysis by the identified significant metabolites.

c) To integrate the pathway based analysis and statistical analysis.

1.11 Problems Identification for the Project

Although Breast cancer (BC) is a malignancy cancer of female with a high incidence, besides it is quite challenging to detect biomarkers.

1.12 Limitations of the Project

There is no microarray data generation opportunity in our country. Although there is some data-generating technology in this country have no proper experts for the extraction of data. The study has been done based on the secondary dataset from website. So, the limitations of secondary data remain in the study. Furthermore, all of these studies have been done in the dry laboratory. Thus, using the integration of wet and dry laboratories this study would be an avenue for further research.

1.13 Layout

Chapter 1: This chapter includes the basic concept of background of the study, basic information about BC, objectives, and limitations of the project.

Chapter 2: This chapter includes the method and methodology for the identification of DE genes from microarray data, PPI networking, KEGG pathway analysis, expression of genes.

Chapter 3: This chapter includes the results of DE genes from microarray data, PPI networking, KEGG pathway analysis, expression of genes.

Chapter 4: This chapter includes a discussion, conclusion, and areas of further research of the project.

Chapter 2

Materials and methods

Chapter 2
Materials and methods

To reach the goal of this study, we considered both raw-data (gene expression profiles) and meta-data associated with breast cancer. Integrated bioinformatics and statistical approaches were used to analyze the datasets to explore cKGs highlighting their functions, pathways, regulatory factors, prognosis power.

2.1. Data sources and descriptions

Collection of gene expression profiles for exploring KGs. The microarray gene expressions profile dataset with accession number GSE42568 was downloaded from the National Center of Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database. The dataset was generated based on two different breast cancer cell lines (BT-20 and MDA-MB-231), either overexpression (BT-20) or knock-down (MDA-MB-231). Untreated cell lines served as controls. The whole genome expression profiles were consisted of 104 treated (case) and 17 control samples with 54,676 probes.

2.2 Identification of DEGs

2.2.1 Microarray data

GEO2R uses GEOquery and limma to perform differential expression analysis using original submitter-supplied processed data tables as input. GEOquery parses GEO data into R data structures that can be used by other R packages. limma (Linear Models for Microarray Analysis) is a statistical test for identifying differentially expressed genes in microarray data. It handles a wide range of experimental designs and data types and applies multiple-testing corrections on P-values to help correct for the occurrence of false positives.

2.2.2 GEO2R

GEO2R is an interactive web tool that allows users to compare two or more groups of Samples in a GEO Series in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by P-value, and as a collection of graphic plots to help visualize differentially expressed genes and assess data set quality. GEO2R uses a variety of R packages from the Bioconductor project. Bioconductor is an open-source software project based on the R programming language that provides tools for the analysis of high-throughput genomic data.

2.2.3 Bioconductor

Current statistical inference problems in biomedical and genomic data analysis routinely involve the simultaneous test of thousands, or even millions, of null hypotheses. These testing problems share the following general characteristics: inference for high-dimensional multivariate distributions, with complex and unknown dependence structures among variables; a broad range of parameters of interest, for example, regression coefficients and correlations; many null hypotheses, in the thousands or even millions; and complex dependence structures among test statistics.

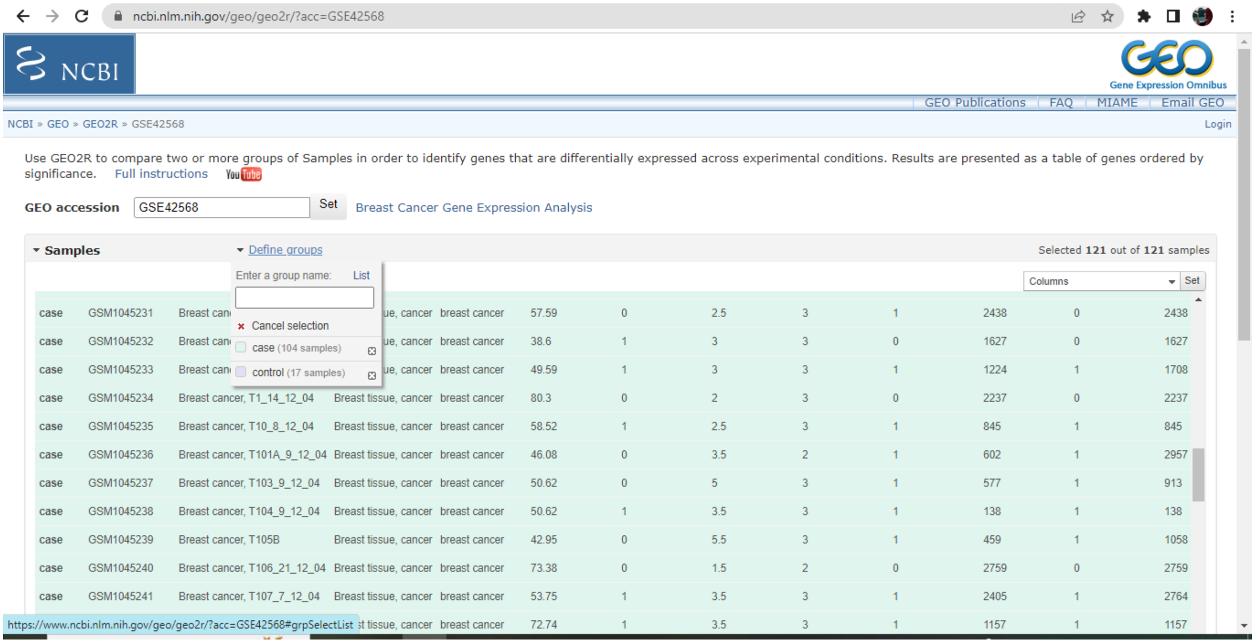
The Bioconductor project started in 2001 and is an open-source, open-development software project to provide tools for the analysis and comprehension of high throughput genomic data. It is based primarily on the R programming language, and most of the Bioconductor components are distributed as R packages. It provides widespread access to a broad range of powerful statistical and graphical methods for the analysis of genomic data. Bioconductor’s software can be used for microarray analysis (data import, quality assessment, normalization, differential expression analysis, clustering, classification, and many more applications); annotation (using microarray probe, gene, pathway, gene ontology, homology, and other annotations); high throughput assays (importing, transforming, editing, analyzing and visualizing various types of assays); and transcription factors analysis (finding candidate binding sites for known transcription factors via sequence matching) (www.Bioconductor.org).

2.2.4 Enter a Series accession number

If you followed a link from a Series record, the GEO accession box will already be populated. Otherwise, enter a Series accession number in the box, e.g., GSE42568. If the Series is associated with multiple microarray Platforms, you will be asked to select the Platform of interest.

2.2.5 Define Sample groups

In the Samples panel, click 'Define groups' and enter names for the groups of Samples you plan to compare, e.g., test and control. Up to 10 groups can be defined. At least two groups must be defined in order to perform the analysis. Groups can be removed using the [X] feature next to the group name. The order in which you define the groups has a bearing on downstream results. For 2 group comparisons, typically it is appropriate to define the test group first, then define the control group - that way, the log fold change direction will follow convention and be positive for genes upregulated in test Samples compared to controls, and negative for downregulated genes



Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. Full instructions

GEO accession GSE42568 Set Breast Cancer Gene Expression Analysis

Selected 121 out of 121 samples

☐ case (104 samples)
☐ control (17 samples)

case	GSM1045231	Breast can	ue, cancer	breast cancer	57.59	0	2.5	3	1	2438	0	2438
case	GSM1045232	Breast can	ue, cancer	breast cancer	38.6	1	3	3	0	1627	0	1627
case	GSM1045233	Breast can	ue, cancer	breast cancer	49.59	1	3	3	1	1224	1	1708
case	GSM1045234	Breast cancer, T1_14_12_04	Breast tissue, cancer	breast cancer	80.3	0	2	3	0	2237	0	2237
case	GSM1045235	Breast cancer, T10_8_12_04	Breast tissue, cancer	breast cancer	58.52	1	2.5	3	1	845	1	845
case	GSM1045236	Breast cancer, T101A_9_12_04	Breast tissue, cancer	breast cancer	46.08	0	3.5	2	1	602	1	2957
case	GSM1045237	Breast cancer, T103_9_12_04	Breast tissue, cancer	breast cancer	50.62	0	5	3	1	577	1	913
case	GSM1045238	Breast cancer, T104_9_12_04	Breast tissue, cancer	breast cancer	50.62	1	3.5	3	1	138	1	138
case	GSM1045239	Breast cancer, T105B	Breast tissue, cancer	breast cancer	42.95	0	5.5	3	1	459	1	1058
case	GSM1045240	Breast cancer, T106_21_12_04	Breast tissue, cancer	breast cancer	73.38	0	1.5	2	0	2759	0	2759
case	GSM1045241	Breast cancer, T107_7_12_04	Breast tissue, cancer	breast cancer	53.75	1	3.5	3	1	2405	1	2764
			t tissue, cancer	breast cancer	72.74	1	3.5	3	1	1157	1	1157

https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE42568#grpSelectList

Figure 1.1: Snap Shot of GEO2R

2.2.6 Perform the analysis

After Samples have been assigned to groups, click the Analyze button to run the analysis with default parameters.

2.3. Protein-protein interaction (PPI) network analysis based on validated DEGs

2.3.1 Cytoscape

Cytoscape is an open source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. Although Cytoscape was originally designed for biological research, now it is a general platform for complex network analysis and visualization. Cytoscape core distribution provides a basic set of features for data integration, analysis, and visualization. Additional features are available as Apps (formerly called Plugins). Apps are available for network and molecular profiling analyses, new layouts, additional file format support, scripting, and connection with databases. They may be developed by anyone using the Cytoscape open API based on Java™ technology and App community development is encouraged. Most of the Apps are freely available from Cytoscape App Store[19].

2.3.2 CytoHubba

Background: Network is a useful way for presenting many types of biological data including protein-protein interactions, gene regulations, cellular pathways, and signal transductions. We can measure nodes by their network features to infer their importance in the network, and it can help us identify central elements of biological networks[20].

Results: We introduce a novel Cytoscape plugin cytoHubba for ranking nodes in a network by their network features. CytoHubba provides 11 topological analysis methods including Degree, Edge Percolated Component, Maximum Neighborhood Component, Density of Maximum Neighborhood Component, Maximal Clique Centrality and six centralities (Bottleneck, EcCentricity, Closeness, Radiality, Betweenness, and Stress) based on shortest paths. Among the eleven methods, the new proposed method, MCC, has a better performance on the precision of predicting essential proteins from the yeast PPI network[21].

Conclusions: CytoHubba provide a user-friendly interface to explore important nodes in biological networks. It computes all eleven methods in one stop shopping way. Besides, researchers are able to combine cytoHubba with and other plugins into a novel analysis scheme. The network and sub-networks caught by this topological analysis strategy will lead to new insights on essential regulatory networks and protein drug targets for experimental biologists. According to cytoscape plugin download statistics, the accumulated number of cytoHubba is around 6,700 times since 2010

Determination of hub genes from the PPI network by using the Cytohubba plugin in Cytoscape. We applied five algorithms of the Cytohubba plugin to obtain the hub genes. Here (A) maximum neighborhood component (MNC), (B) betweenness, (C) degree, (D) edge percolated component (EPC), and (E) maximal clique centrality (MCC). Red to yellow color gradients indicate the higher ranking of hub genes[22].

Degree: Degree shows the number of connected nodes with the individual node. So higher degree indicates a characteristics of hub. Similarly, the more central (closeness centrality) a node is, the closer it is to all other nodes. So higher closeness centrality reflects the tendency of a node to be a hub[23].

MCC: To increase the sensitivity and specificity, we propose MCC to discover featured nodes. The intuition behind MCC is that essential proteins tend to be clustered in a yeast protein-protein interaction network [24]. Given a node *v*, the MCC of *v* is defined as$MCC(v) = \sum_{C \in S(v)} (|C| - 1)!$, where $S(v)$ is the collection of maximal cliques which contain *v*, and $(|C|-1)!$ is the product of all positive integers less than $|C|$. If there is no edge between the neighbors of the node *v*, then $MCC(v)$ is equal to its degree.

MNC - Maximum Neighborhood Component Definition The neighborhood of a node v, nodes adjacent to v, induce a subnetwork N(v). The score of node v, MNC(v), is defined to be the size of the maximum connected component of N(v). The neighborhood N(v) is the set of nodes adjacent to v and does not contain node v. MNC(v) = |V(MC(v))| where MC(v) is a maximum connected component of the G[N(v)] and G[N(v)] is the induced subgraph of G by N(v)[25].

We performed PPI network analysis to explore breast cancer causing hub-genes. To construct the PPI network for host signatures, genes data were collected from the STRING database , and the Cytoscape software was used to construct the network based on “NetworkAnalyzer” for visualization the PPI network of DEGs. cytoHubba (Version 0.1) plugin in Cytoscape were used to explore important nodes/hubs of PPI network. Three top- logical measures among degree, MCC and MNC were used to identify top fourteen key gene.

2.4 Expression Analyses of the Hub Genes in Breast Cancer

We examined the expression of hub genes in BC tissue samples compared to normal tissue samples using GEPIA [26] database. GEPIA is a new web based tool that uses gene expressions from TCGA database to compare the expression profiles of genes between normal and cancer samples. In GEPIA, there were 1085 BC tumor tissue samples and 291 normal tissue samples from the TCGA database. The default cutoffs |Log2Fold Change | > 1.5 and p-value<0.05 were considered as statistically significant. There were survival data of 179 patients with BC in the GEPIA database.

2.5. Functional and pathway enrichment analysis of DEGs highlighting cKGs

Gene ontology (GO) functional and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis were used to determine the significantly terms (biological processes (BP), molecular functions (MF) and cellular components (CC)) and KEGG pathways of the DEGs and highlighting cKGs. Functional enrichment analysis tool DAVID (version: 6.8) were used to analyze GO enrichment and KEGG pathway for the DEGs[27].

Chapter 3

Result

Chapter 3

Result

3.1 Identification of DEGs for BC Patients

The datasets GSE42568 was analyzed to identify DEGs between BC infections and control samples, and the DEGs in each dataset were presented using the volcano plots (Figure 3.1), where red and blue dots represented the up-regulated and down-regulated genes, respectively. In GSE42568, a total of 3277 DEGs with 1528 up-regulated and 1749 down-regulated genes were identified by the GEO2R web tool with |logFC| > 1.5 and *p*-value < 0.05.

$$\text{DEGs} = \begin{cases} \text{DEG (Upregulated)} \quad ; \text{ if adj: p: value } < \ 0{:}\,05 \ \mathit{and}\ aLog2FCg \ > \ +1.5 \\ \text{DEG (Downregulated)} \ ; \text{ if adj: p: value } < \ 0{:}\,05 \ \mathit{and}\ aLog2FCg \ < \ -1.5 \end{cases}$$

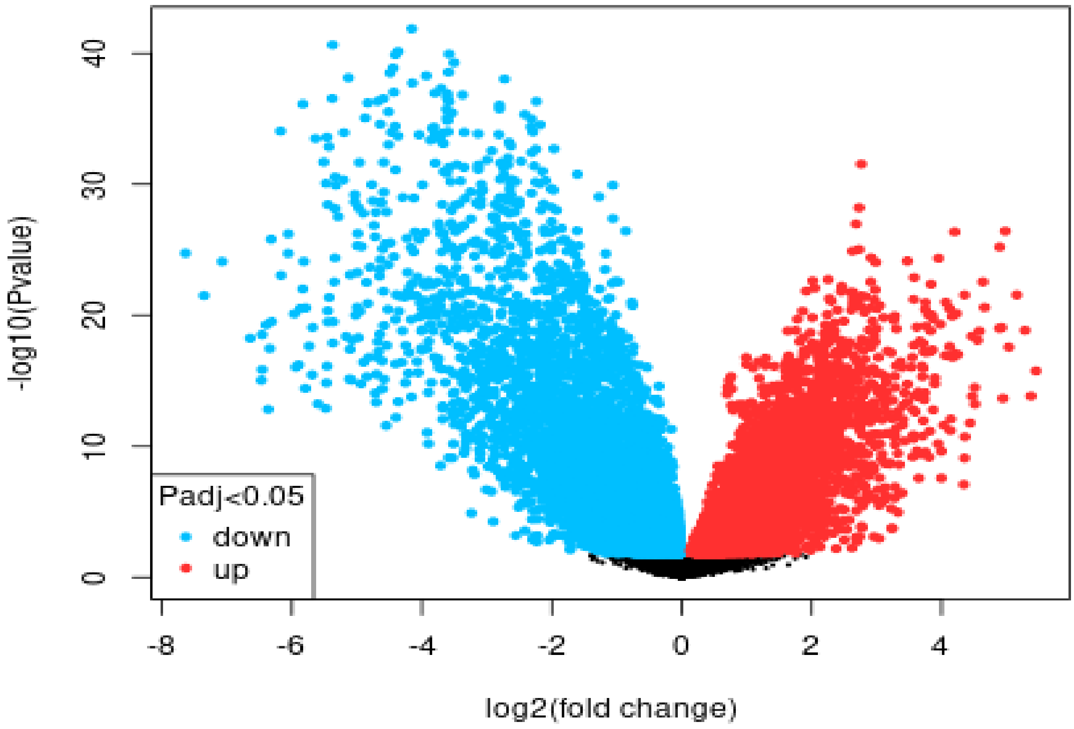


Figure 3.1: Screening of the overlapping DEGs GSE42568 datasets blue dots and red dots represented the significantly down-regulated and up-regulated DEGs, respectively.

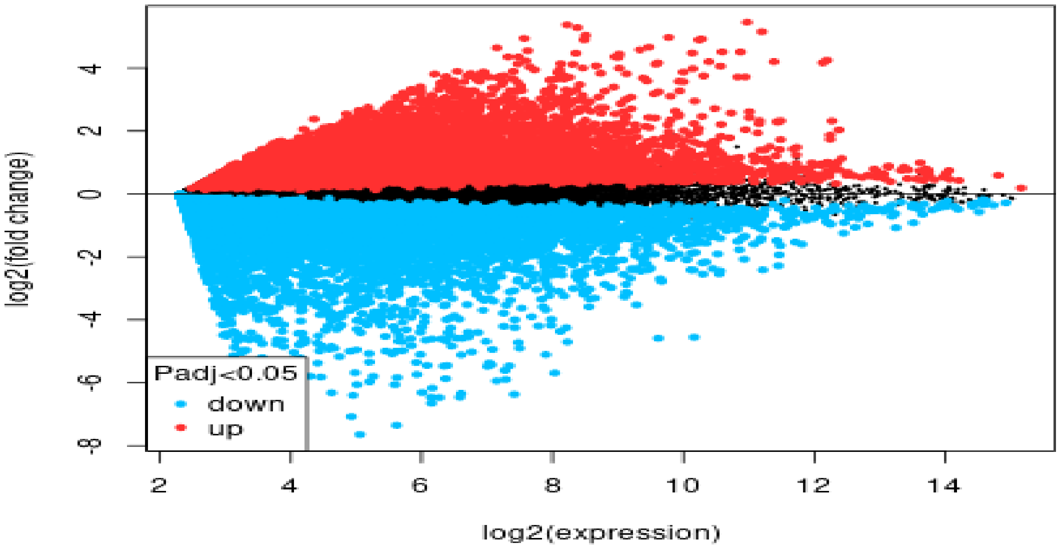


Figure 3.2: Mean difference (MD) plot of DEGs.The plot displays log2 fold change versus average log2 expression values of differentially expressed genes.

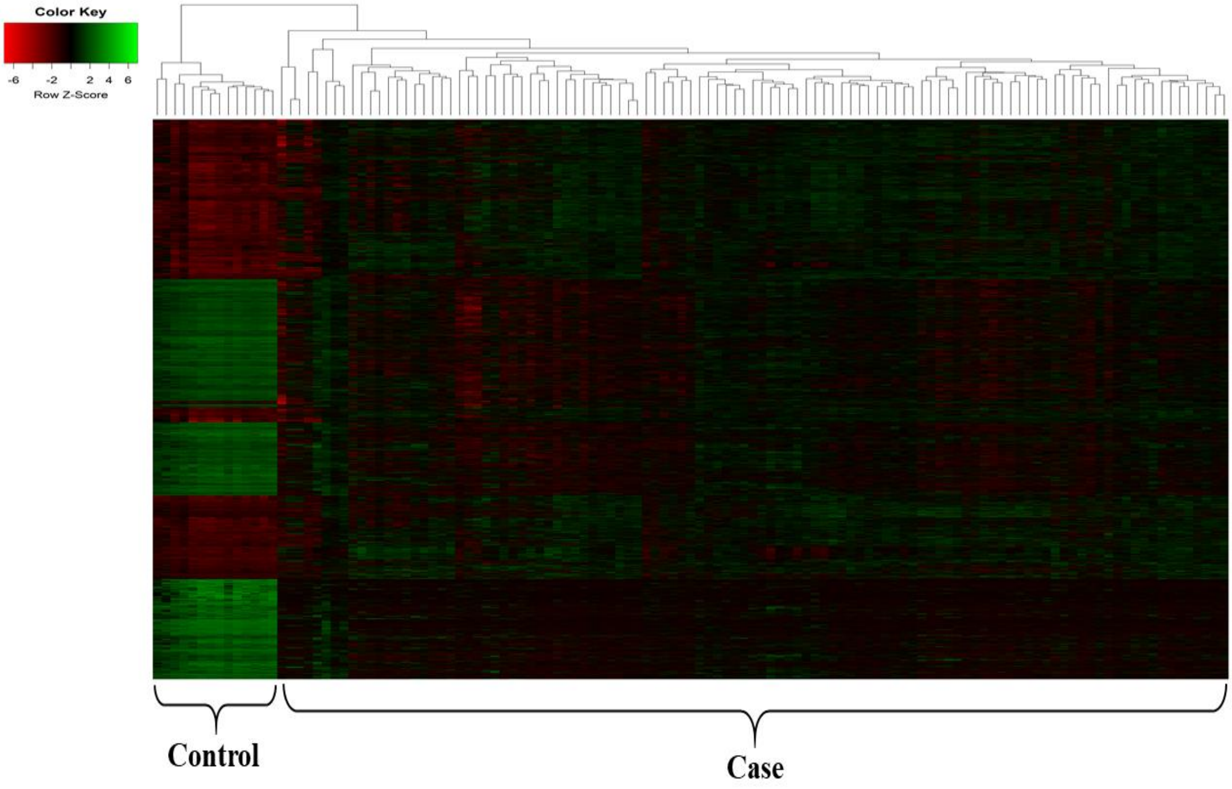
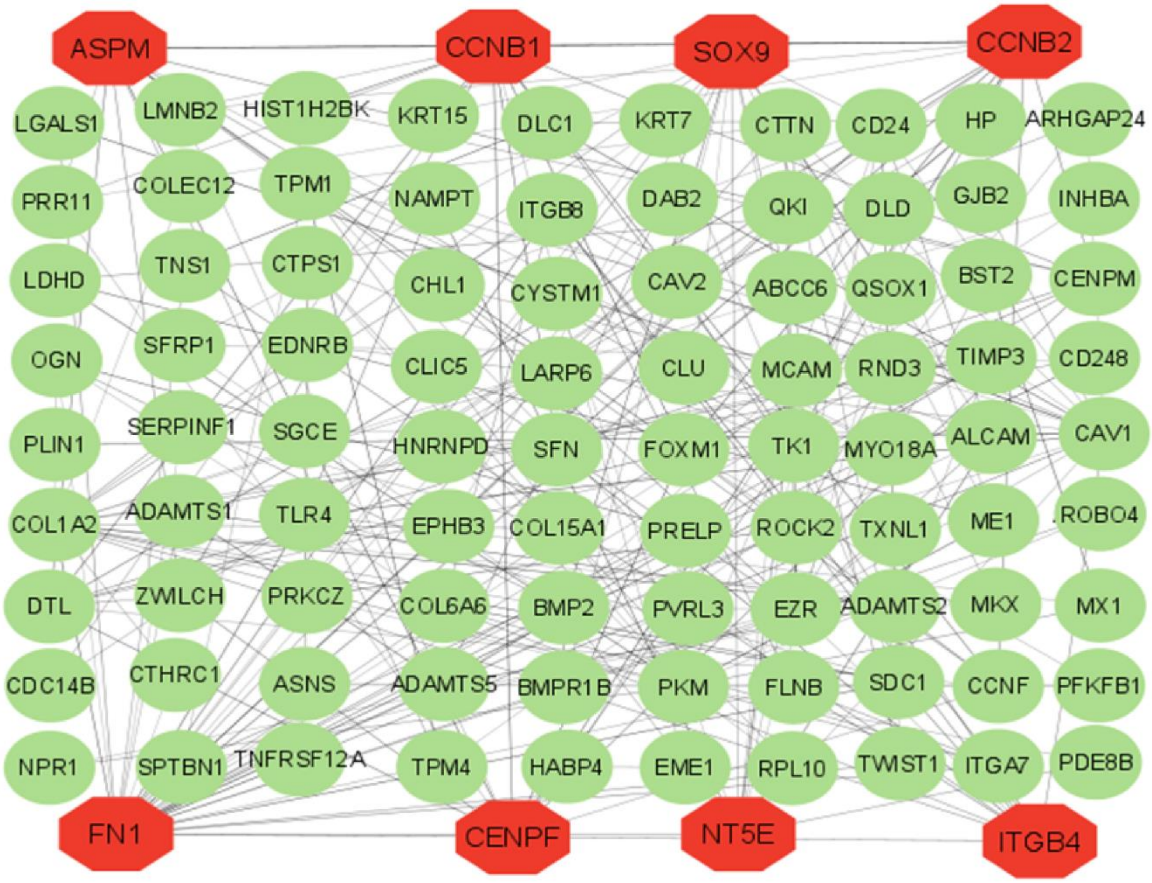


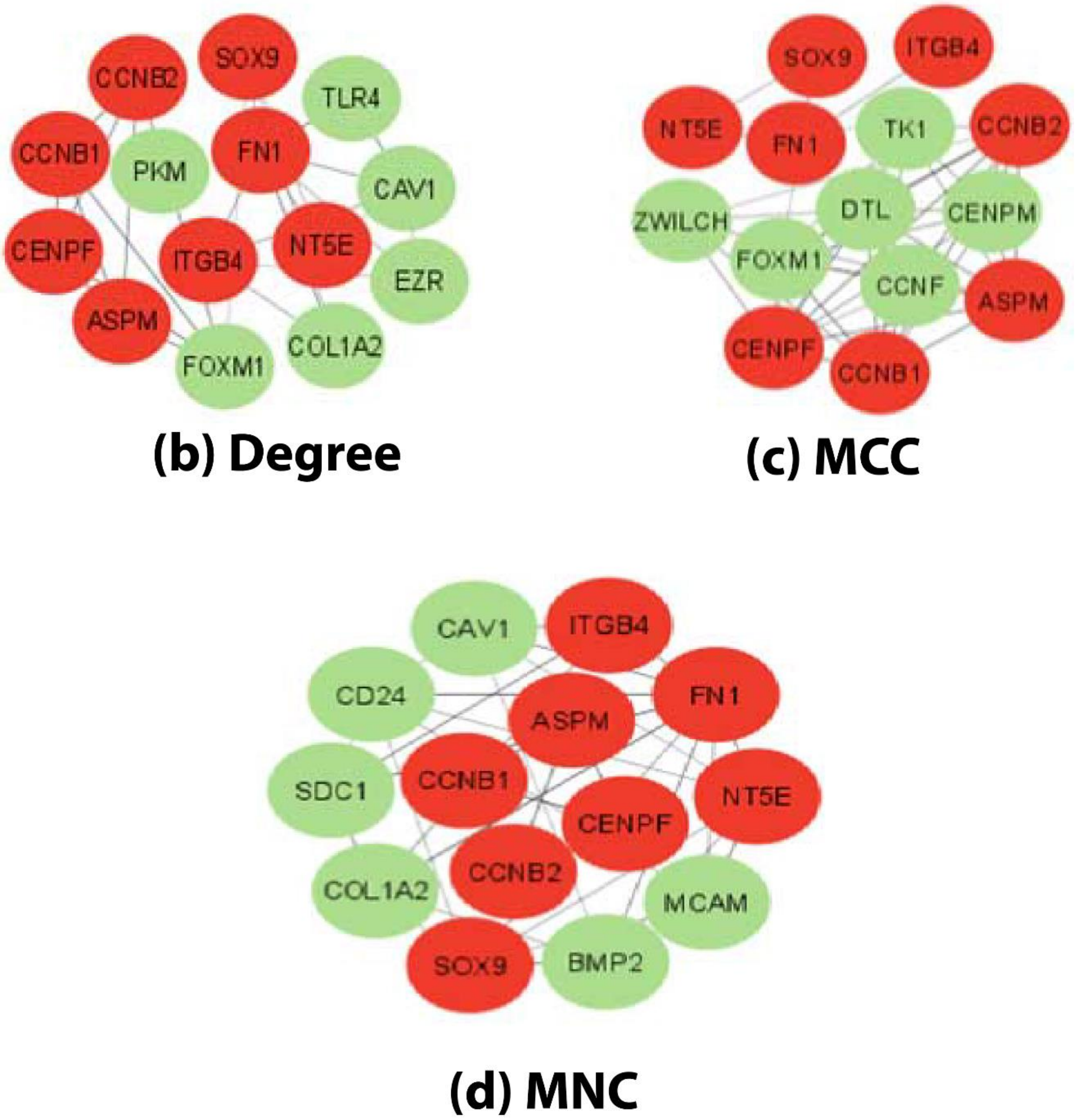
Figure 3.3: Heatmap of the DEGs**.** The heatmap shows the expression profile of the DEGs in BC compared to the adjacent normal tissues. The color scale indicates the Log2FoldChange of the expression value for each gene in cancer vs. normal tissues. Red colors indicate down-regulation and green colors indicate up-regulation.

3.2 PPI network analysis of DEGs and identification of cKGs

The PPI network of DEGs provided 360 nodes and 509 edges with an average number of neighbours are 3.992. The top 14 DEGs were selected from the PPI network based on each of three topological measures (Figure 3.4(a)), degree (Figure 3(b)), MCC (Figure 3(c)) and MNC (Figure 3(d)), by using cytoHubba plugin in Cytoscape. Finally, we selected common 8 DEGs (FN1, CCNB1, CCNB2, ASPM, CENPF, ITGB4, SOX9, and NT5E) as the KGs. We also observed that all KGs were upregulated, since log 2 FC > 1.5. In the PPI network red color indicates Hub gene and green color indicates top up regulated and down regulated DEGs, big size and octagon shape indicate Hub genes (see Figure 3.4(a))



(a) PPI Network



(b) Degree

(c) MCC

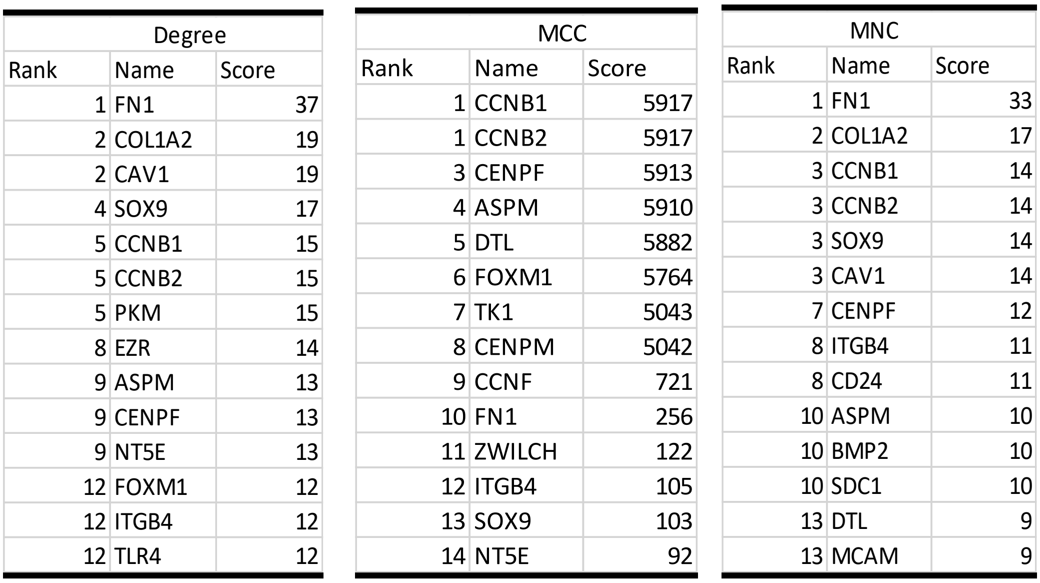
(d) MNC

Figure 3.4: PPI network analysis results of DEGs for identification cKGs by using three topological measures degree, MCC, and MNC. (a) PPI network of 400 DEGs, (b) Top-ranked 14 KGs using degree, (c) Top-ranked 14 KGs using MCC, and (d) Top-ranked 14 KGs using MNC. Then, eight common KGs are considered final KGs and indicated by red color.

Table 1: Selection of cKGs by taking the union of three-sets of top-ranked 8 genes produced by three topological measures with the PPI network.

DEGREE	**MCC**	**MNC**	**Common genes**
FN1	CCNB1	FN1	FN1,CCNB1,CCNB2,ASPM,CENPF, ITGB4, SOX9, NT5E.
COL1A2	CCNB2	COL1A2	
CAV1	CENPF	CCNB1	
SOX9	ASPM	SOX9	
CCNB1	DTL	CAV1	
PKM	FOXM1	CCNB2	
CCNB2	TK1	CENPF	
EZR	CENPM	ITGB4	
NT5E	CCNF	CD24	
ASPM	FN1	ASPM	
CENPF	ZWILCH	BMP2	
FOXM1	ITGB4	SDC1	
ITGB4	SOX9	NT5E	
TLR4	NT5E	MCAM	

Table 2: Top 14 in network string ranked by Degree, MCC and MNC method



Degree			MCC			MNC		
Rank	Name	Score	Rank	Name	Score	Rank	Name	Score
1	FN1	37	1	CCNB1	5917	1	FN1	33
2	COL1A2	19	1	CCNB2	5917	2	COL1A2	17
2	CAV1	19	3	CENPF	5913	3	CCNB1	14
4	SOX9	17	4	ASPM	5910	3	CCNB2	14
5	CCNB1	15	5	DTL	5882	3	SOX9	14
5	CCNB2	15	6	FOXM1	5764	3	CAV1	14
5	PKM	15	7	TK1	5043	7	CENPF	12
8	EZR	14	8	CENPM	5042	8	ITGB4	11
9	ASPM	13	9	CCNF	721	8	CD24	11
9	CENPF	13	10	FN1	256	10	ASPM	10
9	NT5E	13	11	ZWILCH	122	10	BMP2	10
12	FOXM1	12	12	ITGB4	105	10	SDC1	10
12	ITGB4	12	13	SOX9	103	13	DTL	9
12	TLR4	12	14	NT5E	92	13	MCAM	9

3.3 Expression Analyses of the Hub Genes in Breast Cancer

The expressions of the 8 cKGs in the normal tissue samples and breast tissue samples were examined by the GEPIA database. There were 1085 breast tumor samples and 291 normal samples in GEPIA obtained from TCGA database. We found that 6 genes were upregulated (FN1, CCNB1, CCNB2, ASPM, CENPF,and SOX9) and 2 genes were downregulated (ITGB4, and NT5E) in the BC tissue compared to the normal tissue (Figure. 3.5(a)–3.5(h)). Thus, our findings were validated by the TCGA data.

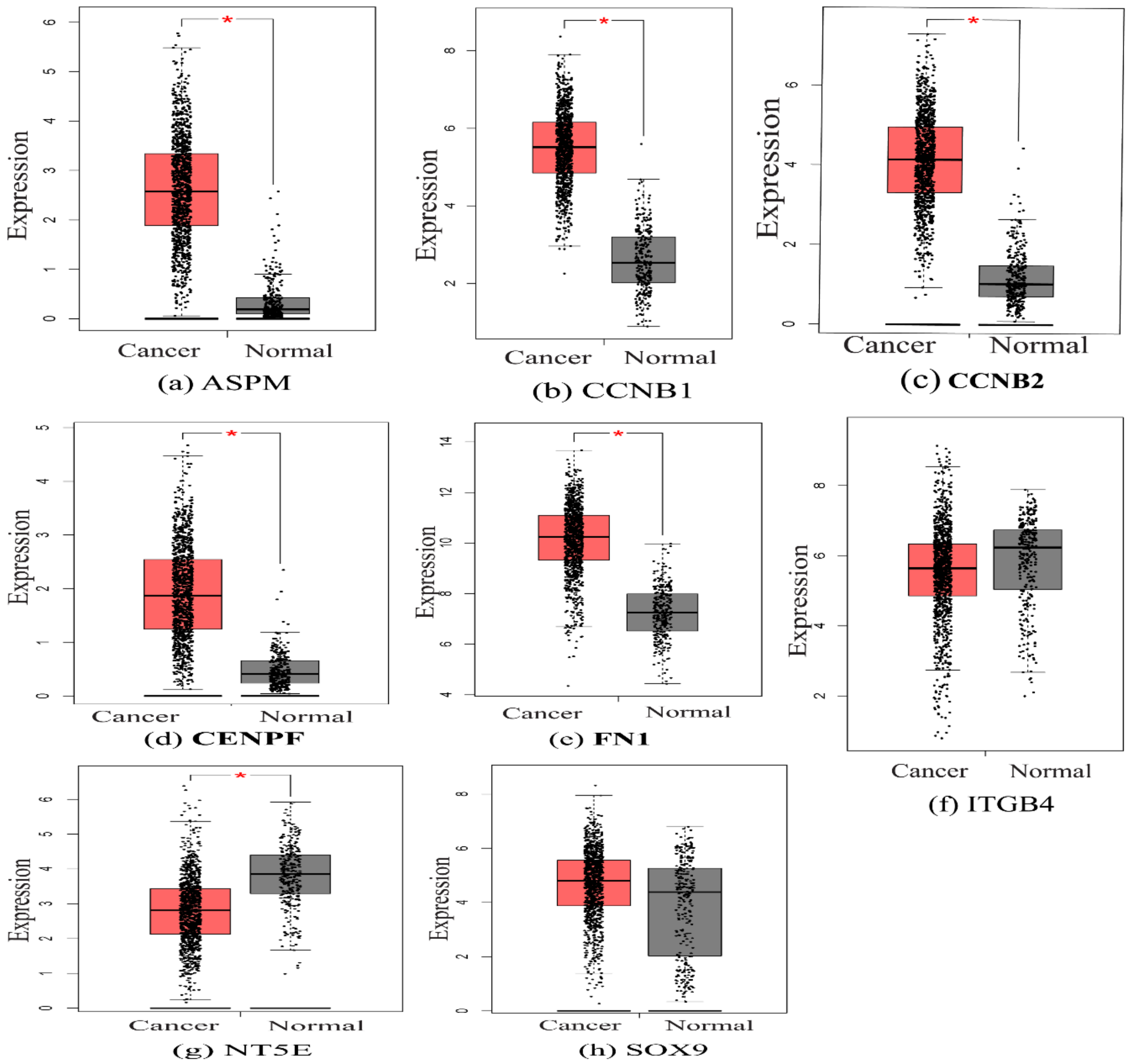


Figure 3.5: Expression and survival analyses of cKGs in BC.

3.4 GO and KEGG pathway enrichment analysis of DEGs highlighting cKGs

The GO functional and KEGG pathway analysis of DEGs showed that 112 GO-BP (Biological Process) terms, 44 GO-CC (Cellular Component) terms, 23 GO-MF (Molecular Function) terms and 11 KEGG- pathway terms are enriched in DAVID by the 400 DEGs. The common enriched GO functions and KEGG pathways from the tool are 26 BPs, 3 MFs, 14 CCs, and 3 KEGGs directly associated with KGs in (Appendix A1 – A4). 26 BPs are Cell Adhesion, Cell Migration, Angiogenesis, Endocardial Cushion

Morphogenesis, Positive Regulation of Cell Proliferation, Filopodium Assembly, Cell-Matrix Adhesion, Positive Regulation Of Fibroblast Proliferation, Positive Regulation Of Transcription From Rna Polymerase Ii Promoter, Notch Signaling Pathway, Signal Transduction, Retina Development in Camera-Type Eye, Negative Regulation of Apoptotic Process, Calcium Ion Homeostasis, Response to Organic Cyclic Compound, Regulation Of Cell Cycle Process, Extracellular Matrix Assembly, Aortic Valve Morphogenesis, Negative Regulation of Transcription, Dna-Templated, Negative Regulation Of Canonical Wnt Signaling Pathway, Cellular Response to Transforming Growth Factor Beta Stimulus, Cell Differentiation, Positive Regulation of Cartilage Development, Male Gonad Development, Integrin-Mediated Signaling Pathway, Cytoskeleton Organization. 3 MFs are Protein Binding, Identical Protein Binding, and Extracellular Matrix Structural Constituent. 14 CCs are Extracellular Matrix, Focal Adhesion, Cell Surface, and Perinuclear Region of Cytoplasm, Integral Component of Membrane, Membrane, Cytoplasm, Receptor Complex, Extracellular Exosome, Cytosol, Apical Plasma Membrane, Basement Membrane, External Side of Plasma Membrane, and Chromatin. 3 KEGGs are Proteoglycans in Cancer, Ecm-Receptor Interaction, and Focal Adhesion.

Chapter 4

Discussion and Conclusion

Chapter 4
Discussion and Conclusion

4.1 Discussion

Identification of biomarkers for complex diseases such as cancer is of paramount importance in treatment, diagnosis and prognosis. Although numerous methods have been proposed to characterize biomarkers, few are from the perspective of regulatory network rewiring. Gene is one important strategy for revealing the disease mechanism from a systematic perspective. The investigation of cancer mutation and perturbation through gene rewiring is significance for addressing the underlying causal regulations responding to phenotypic transition[28]. In this study, we proposed a novel framework for identifying biomarkers based on network rewiring. Disease and normal condition-specific genes have been reconstructed from gene expression data with a priori background network respectively. The gene regulatory interactions changed between them illustrated the results of disease mutation and perturbation. DEGs is extracted and modules in it are detected sequentially.

Here, we applied the proposed framework DEGs for identifying biomarkers of breast cancer. The integrative background network based on prior knowledge and condition specific gene expression data have been used to construct normal and disease genes. We have to admit that there is limitation on missing nodes and edges, which is also expected to be as complete as possible. Totally, DEGs including 509 edges and 360 nodes have been extracted. Then we detected 8 DEGs **(FN1, CCNB1, CCNB2, ASPM, CENPF, ITGB4, SOX9, and NT5E)** as the KGs that drive the progression of breast cancer. Some literatures also suggested that these KGs are BC causing genes.

FN1 was upregulated in BRCA tissues compared with normal tissues. High FN1 mRNA expression was correlated with poor clinical outcomes and had good performance in predicting the survival status of BRCA patients. Further, Cox regression analysis showed that FN1 was an independent prognostic factor for predicting the overall survival of patients with BRCA[29].

CCNB1 (also known as CyclinB1) belongs to the highly conserved cyclin family and is significantly overexpressed in various cancer types. In this study, we demonstrated that CCNB1 had significant predictive power in distant metastasis free survival, disease free survival, recurrence free survival and overall survival of ER+ breast cancer patients. We also found that CCNB1 was closely associated with hormone therapy resistance. In addition, gene set enrichment analysis (GSEA) revealed that its expression was positively associated with genes overexpressed in endocrine therapy resistant samples. The overexpression of the CCNB1 KG suppresses NSCLC progression through inhibition of MEOX1 overexpression [30].

High CCNB2 protein expression was independently associated with LVI positivity in addition to other features of aggressive behaviour, including larger tumour size, higher histological grade, hormonal receptor-negativity, and HER2-positivity, and with shorter survival[31].

The current study indicates that high NT5E expression would be a potential prognostic factor to human solid tumors, especially the lung, breast, gastric and ovarian cancer. High NT5E expression was correlated with distant/local lymph node metastases. NT5E is also a promising target in future cancer immunotherapy and has the potential significance as a biomarker for anti-PD-1/PD-L1 treatment.[32].

High expression of KG ASPM is correlated with worse relapse-free survival in BC, and plays significant role in spindle micro- tubule organization in cell division [33]. The ASPM gene has been conducted with other types of cancer such as bladder cancer, pancreatic cancer, hepatocellular cancer, etc.

The CENPF gene has been proposed as an innovative therapeutic target by activating the PI3K–AKT–mTORC1 signaling pathway for the treatment of breast cancer patients [34].

We find that a basal epithelial marker, integrin-β4 (ITGB4), can be used to enable stratification of mesenchymal-like triple-negative breast cancer (TNBC) cells that differ from one another in their relative tumorigenic abilities.

SOX9 plays a crucial role in the regulation of several different processes during cancer pathogenesis, such as cell growth, apoptosis, migration, invasion, stemness, drug resistance, and immune escape. The expression of transcriptional factor SOX9 is significantly induced in BC patient samples[35].

The GO functional and KEGG pathway enrichment analyses of DEGs revealed that some of GO terms and KEGG pathways are involved in our identified KGs which are directly associated with BC (Appendix). The BP (Table A1) term Cell Adhesion (involved in KG**: ITGB4,FN1**), Cell Migration (**ITGB4**), Angiogenesis **(FN1)**, Endocardial Cushion Morphogenesis **(SOX9)**, Positive Regulation Of Cell Proliferation **(FN1,SOX9)**, Filopodium Assembly **(ITGB4)**, Cell-Matrix Adhesion **(ITGB4,FN1)**, Positive Regulation Of Fibroblast Proliferation **(CCNB1,FN1)**, Positive Regulation Of Transcription From Rna Polymerase Ii Promoter **(SOX9)**, Notch Signaling Pathway **(SOX9)**, Signal Transduction **(SOX9)**, Retina Development In Camera-Type Eye **(SOX9)**, Negative Regulation Of Apoptotic Process **(SOX9)**, Calcium Ion Homeostasis **(NT5E)**, Response To Organic Cyclic Compound **(SOX9)**, Regulation Of Cell Cycle Process **(SOX9)**, Extracellular Matrix Assembly **(SOX9)**, Aortic Valve Morphogenesis **(SOX9)**, Negative Regulation Of Transcription, Dna-Templated **(CENPF,SOX9)**, Negative Regulation Of Canonical Wnt Signaling Pathway **(SOX9)**, Cellular Response To Transforming Growth Factor Beta Stimulus **(SOX9)**, Cell Differentiation **(CENPF)**, Positive Regulation Of Cartilage Development(SOX9), Male Gonad Development **(ASPM, SOX9),** Integrin-Mediated Signaling Pathway(**ITGB4**, **FN1)**, Cytoskeleton Organization(**SOX9)** were reported those are responsible for BC.

The CC (Table A2) term are Extracellular Matrix **(FN1)**, Focal Adhesion **(ITGB4)**, Cell Surface (**ITGB4,NT5E**), Perinuclear Region Of Cytoplasm **(CENPF)**, Integral Component Of Membrane **(NT5E)**, Membrane **(CCNB2,CCNB1,NT5E)**, Cytoplasm **(CCNB2,CCNB1, ASPM,CENPF)**, Receptor Complex **(ITGB4)**, Extracellular Exosome **(ITGB4, NT5E,FN1)**, Cytosol **(CCNB2,CCNB1,NT5E,CENPF)**, Apical Plasma Membrane **(ASPM,FN1)**, Basement Membrane **(ITGB4,FN1)**, External Side Of Plasma Membrane **(NT5E)**, Chromatin **(SOX9,CENPF)** can process and present mutation-induced BC-relevant nascent antigens that contribute to tumor resistance.

The MF (Table A3) term are Protein Binding **(FN1, CCNB1, CCNB2, CENPF, ITGB4, SOX9, and NT5E)**, Identical Protein Binding **(NT5E, and FN1)**, Extracellular Matrix Structural Constituent **(FN1**) were reported those are responsible for BC.

The KEGG (Table A4) term are Proteoglycans in Cancer **(FN1)**, Ecm-Receptor Interaction **(ITGB4**, **FN1)**, Focal Adhesion (**ITGB4**, **FN1)** were reported those are responsible for BC.

4.2 Conclusion

We identified a total of 3277 DEGs with 1528 upregulated and 1749 downregulated genes. Through the PPI network analysis, we screened 8 cKGs. The GO term analysis showed that some DEGs were enriched in several significant biological processes, molecular functions, and cellular components. Again, the KEGG pathway analysis showed that some DEGs were associated with several BC related pathways. The expressions of the hub genes were validated by the TCGA data. These eight genes **FN1, CCNB1, CCNB2, ASPM, CENPF, ITGB4, SOX9**, and **NT5E** might be considered as potential biomarkers for BC diagnosis and treatment. Therefore, our findings might be effective for identified potential biomarkers for BC patients.

4.3 Areas of Further Research

In this project, we use the GEO2R web tool however other methods can be applied. Further analysis such as expression analysis, drug discovery can be applied for better validation.

References

[1] D. Surya Gowri and T. Amudha, “A review on mammogram image enhancement techniques for breast cancer detection,” in *Proceedings - 2014 International Conference on Intelligent Computing Applications, ICICA 2014*, 2014. doi: 10.1109/ICICA.2014.19.

[2] M. Akram, M. Iqbal, M. Daniyal, and A. U. Khan, “Awareness and current knowledge of breast cancer,” *Biological Research*. 2017. doi: 10.1186/s40659-017-0140-9.

[3] S. Guiu *et al.*, “Invasive lobular breast cancer and its variants: How special are they for systemic therapy decisions?,” *Critical Reviews in Oncology/Hematology*. 2014. doi: 10.1016/j.critrevonc.2014.07.003.

[4] V. Pleasant, “Benign Breast Disease,” *Clin. Obstet. Gynecol.*, 2022, doi: 10.1097/GRF.0000000000000719.

[5] R. H. Johnson, C. K. Anders, J. K. Litton, K. J. Ruddy, and A. Bleyer, “Breast cancer in adolescents and young adults,” *Pediatric Blood and Cancer*. 2018. doi: 10.1002/pbc.27397.

[6] S. C. Jones, “Coverage of breast cancer in the Australian print media - Does advertising and editorial coverage reflect correct social marketing messages?,” *J. Health Commun.*, 2004, doi: 10.1080/10810730490468441.

[7] K. A. Delman, “ Introducing the ‘Virtual Tumor Board’ series in CA: A Cancer Journal for Clinicians ,” *CA. Cancer J. Clin.*, 2020, doi: 10.3322/caac.21598.

[8] R. Hou *et al.*, “Prediction of upstaged ductal carcinoma in situ using forced labeling and domain adaptation,” *IEEE Trans. Biomed. Eng.*, 2020, doi: 10.1109/TBME.2019.2940195.

[9] L. Giordano, A. Oliviero, G. M. Peretti, and N. Maffulli, “Erratum: The presence of residents during orthopedic operation exerts no negative influence on outcome (British Medical Bulletin (2019) 130 (65-80) DOI: 10.1093/bmb/ldz009),” *British Medical Bulletin*. 2019. doi: 10.1093/bmb/ldz020.

[10] R. L. Mulder *et al.*, “Recommendations for breast cancer surveillance for female survivors of childhood, adolescent, and young adult cancer given chest radiation: A report from the International Late Effects of Childhood Cancer Guideline Harmonization Group,” *The Lancet Oncology*. 2013. doi: 10.1016/S1470-2045(13)70303-6.

[11] A. R. Carmichael, “Obesity and prognosis of breast cancer,” *Obesity Reviews*. 2006. doi: 10.1111/j.1467-789X.2006.00261.x.

[12] M. C. Pike, M. D. Krailo, B. E. Henderson, J. T. Casagrande, and D. G. Hoel, “‘Hormonal’ risk factors, ‘Breast tissue age’ and the age-incidence of breast cancer,” *Nature*, vol. 303, no. 5920, pp. 767–770, 1983, doi: 10.1038/303767a0.

[13] W. C. Willett, M. J. Stampfer, G. A. Colditz, B. A. Rosner, C. H. Hennekens, and F. E. Speizer, “Moderate Alcohol Consumption and the Risk of Breast Cancer,” *N. Engl. J. Med.*, 1987, doi: 10.1056/nejm198705073161902.

[14] A. Cheepsattayakorn and R. Cheepsattayakorn, “Andrographis paniculata(Green chiretta) may combat COVID-19,” *J. Lung, Pulm. Respir. Res.*, 2020, doi: 10.15406/jlprr.2020.07.00224.

[15] R. R. Wing *et al.*, “Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes,” *Diabetes Care*, 2011, doi: 10.2337/dc10-2415.

[16] A. W. Kurian *et al.*, “Gaps in incorporating germline genetic testing into treatment decision-making for early-stage breast cancer,” in *Journal of Clinical Oncology*, 2017. doi: 10.1200/JCO.2016.71.6480.

[17] S. McCormick, P. Brown, and S. Zavestoski, “The personal is scientific, the scientific is political: The public paradigm of the environmental breast cancer movement,” *Sociological Forum*. 2003. doi: 10.1023/B:SOFO.0000003003.00251.2f.

[18] Z. Antysheva *et al.*, “Abstract 1227: Molecular-based tumor grade predictor for breast cancer, clear cell renal cell carcinoma, and lung adenocarcinoma,” *Cancer Res.*, 2022, doi: 10.1158/1538-7445.am2022-1227.

[19] P. Shannon *et al.*, “Cytoscape: A Software Environment for Integrated Models,” *Genome Res.*, 2003.

[20] P. Shannon *et al.*, “Cytoscape: A software Environment for integrated models of biomolecular interaction networks,” *Genome Res.*, 2003, doi: 10.1101/gr.1239303.

[21] C. H. Chin, S. H. Chen, H. H. Wu, C. W. Ho, M. T. Ko, and C. Y. Lin, “cytoHubba: Identifying hub objects and sub-networks from complex interactome,” *BMC Syst. Biol.*, 2014, doi: 10.1186/1752-0509-8-S4-S11.

[22] F. Liu, J. Dong, D. Zhou, and Q. Zhang, “Identification of key candidate genes related to inflammatory osteolysis associated with Vitamin E-blended UHMWPE debris of

orthopedic implants by integrated bioinformatics analysis and experimental confirmation," *J. Inflamm. Res.*, 2021, doi: 10.2147/JIR.S320839.

[23] Y. Zhang, X. Zheng, M. Chen, Y. Li, Y. Yan, and P. Wang, "Urban fine-grained spatial structure detection based on a new traffic flow interaction analysis framework," *ISPRS Int. J. Geo-Information*, 2021, doi: 10.3390/ijgi10040227.

[24] C. Lu *et al.*, "Why do essential proteins tend to be clustered in the yeast interactome network?," *Mol. Biosyst.*, 2010, doi: 10.1039/b921069e.

[25] C. Y. Lin, C. H. Chin, H. H. Wu, S. H. Chen, C. W. Ho, and M. T. Ko, "Hubba: hub objects analyzer--a framework of interactome hubs identification for network biology.," *Nucleic Acids Res.*, 2008, doi: 10.1093/nar/gkn257.

[26] Z. Tang, B. Kang, C. Li, T. Chen, and Z. Zhang, "GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis," *Nucleic Acids Res.*, 2019, doi: 10.1093/nar/gkz430.

[27] M. S. Reza *et al.*, "Bioinformatics Screening of Potential Biomarkers from mRNA Expression Profiles to Discover Drug Targets and Agents for Cervical Cancer," *Int. J. Mol. Sci.*, 2022, doi: 10.3390/ijms23073968.

[28] G. Novelli, C. Ciccacci, P. Borgiani, M. P. Amati, and E. Abadie, "Genetic tests and genomic biomarkers: Regulation, qualification and validation," *Clinical Cases in Mineral and Bone Metabolism*. 2008.

[29] X. X. Zhang, J. H. Luo, and L. Q. Wu, "FN1 overexpression is correlated with unfavorable prognosis and immune infiltrates in breast cancer," *Front. Genet.*, 2022, doi: 10.3389/fgene.2022.913659.

[30] K. Ding, W. Li, Z. Zou, X. Zou, and C. Wang, "CCNB1 is a prognostic biomarker for ER+ breast cancer," *Med. Hypotheses*, 2014, doi: 10.1016/j.mehy.2014.06.013.

[31] A. I. Aljohani *et al.*, "Upregulation of Cyclin B2 (CCNB2) in breast cancer contributes to the development of lymphovascular invasion," *Am. J. Cancer Res.*, 2022.

[32] T. Jiang *et al.*, "Comprehensive evaluation of NT5E/CD73 expression and its prognostic significance in distinct types of cancers," *BMC Cancer*, 2018, doi: 10.1186/s12885-018-4073-7.

[33] G. B., L. A., E. A.C., D. C., B. J., and L. Q., "An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients," *Breast Cancer Res. Treat.*, 2010.

[34] J. Sun *et al.*, "Overexpression of CENPF correlates with poor prognosis and tumor bone metastasis in breast cancer," *Cancer Cell Int.*, 2019, doi: 10.1186/s12935-019-0986-8.

[35] S. Jana *et al.*, "SOX9: The master regulator of cell fate in breast cancer," *Biochemical Pharmacology*. 2020. doi: 10.1016/j.bcp.2019.113789.

APPENDIX

Table A1: Biological Process (BP)

Term	**Count**	**P Value**	**Genes**
Cell Adhesion	24	0.0000934	AOC3, SRPX, COL15A1, TNFRSF12A, **ITGB4**, MCAM, **FN1**, NEXN, RND3, PCDH18, PTPRF, ALCAM, CHL1, SSPN, ACKR3, ITGB8, SYMPK, ITGA7, COL6A6, SVEP1, PGM5, CD24, NECTIN3, SSX2IP
Cell Migration	15	0.0000303	**ITGB4**, SH3KBP1, RHOH, RND3, PRKCZ, PTPRF, MYO18A, SDC1, ITGB8, CD248, ITGB1BP1, IGFBP6, CD24, EPHB3, CTHRC1
Angiogenesis	14	0.0000360	ROBO4, COL15A1, CALCRL, TNFRSF12A, MCAM, CAV1, NPR3, **FN1**, ARHGAP24, TMEM100, PTPRB, ACKR3, HOXA7, EPHB3
Endocardial Cushion Morphogenesis	4	0.00528804	ADAMTS5, BMP2, TWIST1, **SOX9**
Positive Regulation Of Cell Proliferation	19	0.00591442	KLB, SPHK1, **FN1**, FOXM1, PRLR, PRKCZ, HOXC10, BMP2, SFRP1, EDNRB, PDGFD, NAMPT, ID4, CD248,

to be Continued

Table A1: Biological Process (BP)

Term	Count	P Value	Genes
			ITGB1BP1, **SOX9**, SINHCAF, SOX4, SHANK2
Filopodium Assembly	4	0.00774828	FGD3, **ITGB4**, ARHGEF4, EZR
Cell-Matrix Adhesion	7	0.01127108	SGCE, ECM2, **ITGB4**, **FN1**, ITGB8, ITGA7, ITGB1BP1
Positive Regulation Of Fibroblast Proliferation	5	0.01459415	**CCNB1**, PDGFD, SPHK1, **FN1**, CD248
Positive Regulation Of Transcription From Rna Polymerase Ii Promoter	32	0.01558244	CASZ1, EHF, ZNF493, CEBPD, CITED2, TWIST1, RORB, FOXM1, MEOX1, HOXC10, ZNF827, NAMPT, ZNF107, **SOX9**, HOXA7, SOX4, SPDEF, DDX17, ZNF782, DMRT2, VDR, LMO3, MICAL2, INHBA, POU2F3, DAB2, BMP2, ID4, HNRNPD, ITGB1BP1, BMPR1B, TLR4
Notch Signaling Pathway	7	0.01616113	TMEM100, PTP4A3, BMP2, PLN, PERP, ITGB1BP1, **SOX9**
Signal Transduction	33	0.01694412	CDS1, RET, COL15A1, BEX2, TENM3, NPR1, NPR3, ADRA1A, PRKCZ, AKAP12, ARHGAP21, GRK3, ARHGAP20, ALCAM, CHL1, NAMPT, FLNB, SFN, **SOX9**, IGFBP6, RHPN2, GABRE, PDE8B, MX1, ARHGAP24,

to be Continued

Table A1: Biological Process (BP)

Term	Count	P Value	Genes
			ARHGAP32, INPP4B, MRAS, CTTN, PPP1R1A, DLC1, OGN, SKAP2
Retina Development In Camera-Type Eye	5	0.0174481	RET, SERPINF1, RORB, **SOX9**, BMPR1B
Negative Regulation Of Apoptotic Process	17	0.01795648	BNIP3L, TNFAIP8, RPL10, CITED2, SPHK1, IFI6, ASNS, TWIST1, PRLR, PRKCZ, DAB2, SFRP1, EDNRB, MYO18A, YME1L1, **SOX9**, CHML
Calcium Ion Homeostasis	4	0.01865481	**NT5E**, ASPH, CAV1, ABCC6
Response To Organic Cyclic Compound	5	0.01953081	SFRP1, EDNRB, EPHX1, **SOX9**, ACACB
Regulation Of Cell Cycle Process	3	0.02329389	SFRP1, DCUN1D3, **SOX9**
Extracellular Matrix Assembly	3	0.02329389	COL1A2, QSOX1, **SOX9**
Aortic Valve Morphogenesis	4	0.02351501	ADAMTS5, ROCK2, TWIST1, **SOX9**
Negative Regulation Of Transcription, Dna-Templated	18	0.02452054	KDM5B, CEBPD, CITED2, CBX3, VDR, TWIST1, RORB, FOXM1, **CENPF**, BMP2, SFRP1, ZNF827, MDFIC, ID4, TRIB3, **SOX9**, HOXA7, TRIM11
Negative Regulation Of Canonical Wnt Signaling Pathway	8	0.02685465	DAB2, SFRP1, BMP2, CAV1, JADE1, IGFBP6, **SOX9**, CTHRC1

to be Continued

Table A1: Biological Process (BP)

Term	Count	P Value	Genes
Cellular Response To Transforming Growth Factor Beta Stimulus	5	0.02799396	DAB2, SFRP1, PDGFD, CAV1, **SOX9**
Cell Differentiation	20	0.0280397	EHF, TNFRSF12A, CAV2, VDR, STRBP, CAV1, RARRES2, INHBA, ABHD5, ARHGAP24, UGCG, **CENPF**, SFRP1, ID4, FLNB, ITGB1BP1, JHY, CAMK2G, SOX4, SPDEF
Positive Regulation Of Cartilage Development	3	0.0336606	BMP2, **SOX9**, BMPR1B
Male Gonad Development	6	0.03590002	**ASPM**, SFRP1, CITED2, BIK, **SOX9**, INHBA
Integrin-Mediated Signaling Pathway	6	0.04242945	ADAMTS1, **ITGB4**, **FN1**, ITGB8, ITGA7, ITGB1BP1
Cytoskeleton Organization	7	0.04610064	FGD3, ARHGAP21, SH3KBP1, TPM1, MICAL2, **SOX9**, LMNB2

Table A2: Cellular Component (CC)

Term	Count	P Value	Genes
Extracellular Matrix	16	0.0000461	COLEC12, COL15A1, ECM2, LRRN3, RARRES2, **FN1**, PRELP, ADAMTS5,

to be Continued

Table A2: Cellular Component (CC)

Term	Count	P Value	Genes
			ADAMTS2, COL1A2, ADAMTS1, OGN, MMP28, TIMP3, COL6A6, CD248
Focal Adhesion	20	0.0000161	TPM4, **ITGB4**, CAV2, SH3KBP1, MCAM, CAV1, NEXN, RND3, ARHGAP24, AKAP12, DAB2, ALCAM, CTTN, DLC1, ITGB8, EVL, FLNB, PGM5, EZR, TNS1
Cell Surface	23	0.00162793	AOC3, SRPX, IGSF3, **ITGB4**, GP2, CLU, PRLR, PDIA4, GHR, BST2, BMP2, **NT5E**, SFRP1, CA4, MYO18A, SLC39A6, SDC1, ACKR3, ITGB8, ITGA7, CD24, TLR4, TNS1
Perinuclear Region Of Cytoplasm	25	0.00211787	DCUN1D3, CCNF, PRKCZ, CLU, TMEM100, PLN, CA4, EPG5, RHPN2, TPD52, GALNT3, CAV2, CAV1, SERPINF1, MX1, SYNJ2, INHBA, DAB2, GJB2, **CENPF**, MAP1B, ACKR3, ITGB1BP1, EZR, TLR4
Integral Component Of Membrane	118	0.00230432	KLB, GALNT16, PGAP3, ZDHHC2, CELSR2, CDC14B, TMEM267, EDNRB, STS, ALCAM, FAM174B, SLC39A6, VSTM4, ENPP5, SLC12A8, BNIP3L, SLC38A1, TPM1, SHISA2, PRLR, ERN1, INPP4B, YME1L1, CDS1, DIPK1A, CALCRL, IGSF3, NPR1, NPR3, PCDH18, UGCG, RDH10, BTNL9, B3GALNT1, LRRN3, CAV2, MCAM, CAV1, EPHX1, GP2, SUCO, LRP1B, BST2, GJB2, PTPRB, FXYD1, ECHDC1, SDC1, CYSTM1, RETREG2, BMPR1B, RET, ROBO4, TENM3, SH3KBP1, C2CD2, HBB, ADRA1A, PTPRF, GHR, TMEM47, SGCE, TMEM100, CHL1, NNT, CA4,

to be Continued

Table A2: Cellular Component (CC)

Term	Count	P Value	Genes
			ARFGEF3, ATP6AP1, TNFRSF12A, ELOVL5, STRADB, ABCC6, EMP1, TMC5, IL17RB, TRDN, SLC25A16, MRAP, ASPH, TLCD4, ITGA7, PLIN1, SEC22B, SMIM20, TLR4, SHANK2, PTGER4, COLEC12, COL15A1, PTGER3, IFI6, SLC7A1, **NT5E**, SLC25A29, ERMP1, PLN, MARVELD3, FLNB, CHML, AOC3, CYB5A, PEX19, GALNT3, KCNJ8, GOLM1, BIK, DMRT2, CYP4B1, CLCN4, TSPAN13, AADAC, SCD, C4ORF3, ACKR3, CD248, CYB561, F2RL2, NPIPB5
Membrane	86	0.00321709	RET, DHRS13, SRPX, ROBO4, TENM3, SH3KBP1, C2CD2, CELSR2, CLU, IPO4, GHR, TMEM267, STS, CHL1, NNT, CA4, MYO18A, SVEP1, VSTM4, SLC12A8, DDX17, BNIP3L, SLC38A1, TPM4, ATP6AP1, ELOVL5, SPHK1, ABCC6, SORD, EMP1, CTPS1, SYTL4, PRLR, IL17RB, TRDN, MRAS, TLCD4, YME1L1, EVL, ITGB1BP1, EZR, SEC22B, PAFAH1B3, PRR11, COLEC12, CDS1, PTGER4, COL15A1, CALCRL, RPL10, IGSF3, PTGER3, SLC7A1, PRKCZ, **CCNB2**, CYTH2, UGCG, **NT5E**, **CCNB1**, ERMP1, PLN, RDH10, MARVELD3, FLNB, CHML, CAMK2G, AOC3, B3GALNT1, CYB5A, LRRN3, GALNT3, CAV2, CAV1, MX1, SUCO, LRP1B, BST2, CLCN4, TSPAN13, SCD, DLC1, BMPR1B, RETREG2, CD24, CYB561, NPIPB5

to be Continued

Table A2: Cellular Component (CC)

Term	Count	P Value	Genes
Cytoplasm	121	0.00378062	CCNF, JADE1, CELSR2, CLU, IPO4, DCAF7, CDC14B, KLHL31, RGS4, CAPN8, LGALS1, FNTA, WDR90, NAMPT, HABP4, SVEP1, AZI2, HADH, SOX4, TNS1, TPD52, ARRDC1, LMO3, KRT7, ERN1, INPP4B, RRAGC, EZR, PRR11, TRIM11, SKAP2, IQCA1, BEX2, CALCRL, DCUN1D3, PRKCZ, SNX3, **CCNB2**, **CCNB1**, RDH10, SYMPK, SFN, SPTBN1, NEBL, FUS, VDR, MICAL2, SUCO, HSPA12A, ARHGAP24, LARP6, QKI, BST2, DAB2, PPP1R1A, HOOK1, RCAN3, ID4, HNRNPD, SPATS2L, APOBEC3B, CLIC5, TNFAIP8, SH3KBP1, CITED2, HMGB3, BICDL1, MEOX1, RND3, ADRA1A, AKAP12, EPB41L4B, CA3, MMP28, DBT, MYO18A, ME1, EPG5, STRBP, STRADB, SPHK1, RHOH, ELMOD3, CTPS1, ECRG4, **ASPM**, PTP4A3, PKM, MDFIC, CTTN, EVL, ITGB1BP1, SEC22B, PAFAH1B3, SSX2IP, TLR4, DCTN6, ROCK2, CAPG, CCNDBP1, HSD17B11, FGD3, CYTH2, MARVELD3, EPS8L1, SH3BP4, FLNB, CHML, CAMK2G, PAIP1, CTHRC1, AOC3, PEX19, TAF15, TXNL1, MX1, XPO7, **CENPF**, DLC1, CD248
Receptor Complex	11	0.00478186	GHR, RET, PTPRB, NPR1**, ITGB4**, VDR, BMPR1B, PRLR, TLR4, LRP1B, EPHB3
Extracellular Exosome	55	0.00541579	CLIC5, ROBO4, **ITGB4**, HP, HBB, CLU, PTPRF, ALCAM, LGALS1, CHL1, NAMPT, CA4, ITGB8, QSOX1, SLC38A1, TPM4,

to be Continued

Table A2: Cellular Component (CC)

Term	Count	P Value	Genes
			ATP6AP1, ARRDC1, GPX3, SERPINF1, SORD, KRT7, TMC5, SFRP1, PKM, DAAM2, EZR, CREB5, COL15A1, NPR3, PRELP, CAPG, CST6, PRKCZ, GNS, THSD4, SNX3, **NT5E**, EPS8L1, SH3BP4, FLNB, SFN, SPTBN1, GALNT3, NEBL, **FN1**, GP2, HSPA12A, BST2, COL1A2, KRT15, OGN, SDC1, CD248, CYSTM1
Cytosol	115	0.01341579	ZWILCH, CCNF, JADE1, CLU, DCAF7, RGS4, PREX2, PREX1, FNTA, NAMPT, HABP4, PDE8B, EPHB3, DDX17, TPM4, TPM1, MCCC1, KRT7, INPP4B, SFRP1, RRAGC, MAP1B, TRIB3, EZR, TRIM11, SKAP2, PFKFB1, ZCCHC7, CASZ1, ABHD5, ACACB, PRKCZ, FBXO41, SNX3, **CCNB2**, ARHGAP21, **CCNB1**, GRK3, ARHGAP20, SYMPK, STAP2, SFN, RHPN2, RAB11FIP3, SPTBN1, VDR, ASNS, SYNJ2, POU2F3, BST2, ARHGAP32, DAB2, GJB2, HOOK1, RCAN3, ECHDC1, HNRNPD, SPATS2L, KDM5B, DENND1B, SH3KBP1, C2CD2, HBB, ADRA1A, AKAP12, GHR, EPB41L4B, CA3, DBT, ME1, PGM5, TK1, STRADB, SPHK1, SORD, CTPS1, TRDN, H2BC12, PKM, CTTN, ARHGEF4, EVL, ITGB1BP1, PLIN1, DTL, SFI1, SGK2, PAFAH1B3, SHANK2, DCTN6, RPL10, ROCK2, HSD17B11, FGD3, CYTH2, **NT5E**, DNAJB4, POLR2E, MICALL2, EPS8L1, FLNB, CHML, CAMK2G, PAIP1, CYB5A, KDM4B,

to be Continued

Table A2: Cellular Component (CC)

Term	**Count**	**P Value**	**Genes**
			PEX19, TXNL1, MX1, FBXL16, **CENPF**, KRT15, DLC1, KLF9, CENPM
Apical Plasma Membrane	14	0.01432917	CLIC5, CAV1, GP2, **FN1**, ECRG4, SLC7A1, PRKCZ, BST2, **ASPM**, FXYD1, CA4, EZR, F2RL2, SHANK2
Basement Membrane	6	0.0240888	COL15A1, ADAMTS1**, ITGB4**, SERPINF1, **FN1**, TIMP3
External Side Of Plasma Membrane	15	0.03921375	BTNL9, ST14, MCAM, GP2, PRLR, GHR**, NT5E,** ALCAM, CA4, SDC1, ACKR3, ITGA7, CD248, CD24, TLR4
Chromatin	27	0.0408277	CASZ1, EHF, CEBPD, CITED2, SHOX2, TSHZ2, TWIST1, RORB, FOXM1, MEOX1, HOXC10, IPO4, ZNF827, **SOX9**, HOXA7, SOX4, SPDEF, CBX3, DMRT2, VDR, PCGF2, POU2F3, TBX15, **CENPF**, KLF9, MKX, CREB5

Table A3: Molecular Function (MF)

Term	**Count**	**P Value**	**Genes**
Protein Binding	257	0.0000628	PGAP3, CCNF, DCAF7, PREX1, ALCAM, LGALS1, EDNRB, WDR90, NAMPT, HABP4, **SOX9**, HOXA7, VSTM4, SOX4, SLC12A8, EPHB3, TNS1, DDX17, ARRDC1, SERPINF1, MCCC1, KRT7, SYTL4, INPP4B, SFRP1, RRAGC, DAAM2, MAP1B, YME1L1, TRIB3,

to be Continued

Table A3: Molecular Function (MF)

Term	Count	P Value	Genes
			EZR, TRIM11, SKAP2, CDS1, BEX2, CSTF3, DCUN1D3, TSHZ2, TWIST1, PCDH18, GNS, PRKCZ, SNX3, UGCG, ZNF827, RDH10, LDHD, SYMPK, SFN, RAB11FIP3, SPTBN1, NEBL, KNOP1, **FN1**, MICAL2, HSPA12A, LARP6, BST2, GJB2, COL1A2, PPP1R1A, ECHDC1, HNRNPD, ID4, SDC1, CYSTM1, RETREG2, CD24, APOBEC3B, RET, CLIC5, SRPX, ROBO4, CITED2, SH3KBP1, SETD7, HBB, MEOX1, HOXC10, TMEM47, AKAP12, MMP28, EPG5, TIMP3, TK1, TNFRSF12A, ELOVL5, STRBP, STRADB, SPHK1, ELMOD3, RHOH, EMP1, ECRG4, IL17RB, MRAP, MRAS, MDFIC, CTTN, TLCD4, ARHGEF4, SINHCAF, PLIN1, SFI1, DLD, SGK2, PAFAH1B3, TLR4, PTGER4, RPL10, IFI6, CAPG, CCNDBP1, SLC7A1, CST6, **NT5E**, ERMP1, SERTAD4, PERP, MICALL2, ZNF107, FLNB, CHML, SPDEF, PAIP1, AOC3, PEX19, GOLM1, BIK, DMRT2, PCGF2, XPO7, SEPTIN8, **CENPF**, AADAC, DLC1, KRT15, CYB561, F2RL2, KLB, HDDC2, ZWILCH, JADE1, HP, RORB, CLU, IPO4, CDC14B, TMEM267, EME1, FNTA, AZI2, TPD52, BNIP3L, SLC38A1, TPM4, LMO3, TPM1,

to be Continued

Table A3: Molecular Function (MF)

Term	Count	P Value	Genes
			SHISA2, PRLR, DHDDS, PDIA4, ERN1, GPATCH11, ZNF92, PFKFB1, ZCCHC7, CALCRL, NPR3, ABHD5, ACACB, ARHGAP21, **CCNB2**, GRK3, **CCNB1**, STAP2, IGFBP6, CBX3, VDR, CAV2, FUS, RARRES2, CAV1, EPHX1, ASNS, INHBA, SYNJ2, POU2F3, ARHGAP24, LRP1B, QKI, ARHGAP32, DAB2, BMP2, PTPRB, HOOK1, OGN, RCAN3, BMPR1B, KDM5B, TNFAIP8, **ITGB4**, HMGB3, FOXM1, ADRA1A, RND3, LMNB2, GHR, TMEM100, ADAMTS5, CA3, ADAMTS1, SNRNP70, CA4, ME1, MYO18A, ATP6AP1, GPX3, CTPS1, TRDN, H2BC12, PTP4A3, PKM, ASPH, EFHC1, ITGA7, ITGB1BP1, EVL, SEC22B, SMIM20, DTL, SSX2IP, SHANK2, CREB5, COLEC12, CEBPD, ROCK2, HSD17B11, CYTH2, PLN, DNAJB4, TRA2A, MARVELD3, POLR2E, EPS8L1, SH3BP4, CAMK2G, CYB5A, TAF15, KCNJ8, MX1, TBX15, CLCN4, SCD, ACKR3, CD248, NECTIN3

to be Continued

Identical Protein Binding	46	0.004103	PFKFB1, CEBPD, ACACB, LMNB2, GHR, ADAMTS5, **NT5E**, ALCAM, PLN, NAMPT, ME1, SH3BP4, FLNB, SFN, IGFBP6, TK1, HADH, CAMK2G, AOC3, BNIP3L, TPM4, CBX3, FUS, DMRT2, ARRDC1, GPX3, CAV1, TPM1, MX1, **FN1**, SORD, CTPS1, INHBA, ERN1, BST2, MRAP, GJB2, SFRP1, COL1A2, HOOK1, SDC1, EZR, NECTIN3, PAFAH1B3, TLR4, CREB5
Extracellular Matrix Structural Constituent	8	0.01201	SRPX, COL15A1, COL1A2, GP2, **FN1**, PRELP, THSD4, CTHRC1

Table A4: KEGG Pathways

Term	**Count**	**P Value**	**Genes**
Proteoglycans In Cancer	14	0.0000585	ROCK2, CAV2, CAV1, **FN1**, TWIST1, MRAS, COL1A2, CTTN, TIMP3, SDC1, FLNB, EZR, CAMK2G, TLR4
Ecm-Receptor Interaction	7	0.012918	COL1A2, **ITGB4**, **FN1**, ITGB8, SDC1, ITGA7, COL6A6
Focal Adhesion	11	0.013388	COL1A2, ROCK2, CAV2, **ITGB4**, PDGFD, CAV1, **FN1**, ITGB8, ITGA7, COL6A6, FLNB