الجمهوريــة الجزائريــة الديمقراطية الشعبيـة
Democratic and Popular Republic of Algeria
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
جامعة 8 ماي 1945 قالمة
University of 8 May 1945 Guelma
Faculty of Natural and Life Sciences, Earth and the Universe

A thesis submitted in partial fulfillment of the requirements for Master's degree

Domain: Natural and Life Sciences

Field: Biological Sciences

Specialty: Applied Biochemistry

Department: Biology

THEME

Identification of Hub genes and key biomarkers in patients with breast cancer by bioinformatics tools

Presented by students:

- **Miss MEGAIZI IMANE**
- **Miss DJABALI IMENE**

Before the jury composed of:

President:	**Dr. BOUMAAZA A.**	**M.C.A**	**University of Guelma**
Examiner:	**Dr. BAALI S.**	**M.A.A**	**University of Guelma**
Supervisor:	**Dr. OUMEDDOUR A.**	**M.C.A**	**University of Guelma**

June 2025

Acknowledgements

First of all, we would like to express our sincere thanks to God « الله » for giving us the strength and willpower to complete this work.

الحمد لله

*We also want to thank **Dr. BOUMAAZA A**, an associate professor at the University of Guelma, for agreeing to chair this jury.*

*Our heartfelt thanks go to **Dr. BAALI S**, an assistant professor (Grade A) at the University of Guelma, for accepting to evaluate and review our work.*

*We are very grateful to our supervisor **Dr. OUMEDDOUR A**, an associate professor at the University of Guelma, for suggesting this research topic and for dedicating his time and effort to guide us through this thesis. We truly appreciate his trust and support.*

Thank you.

Dedication

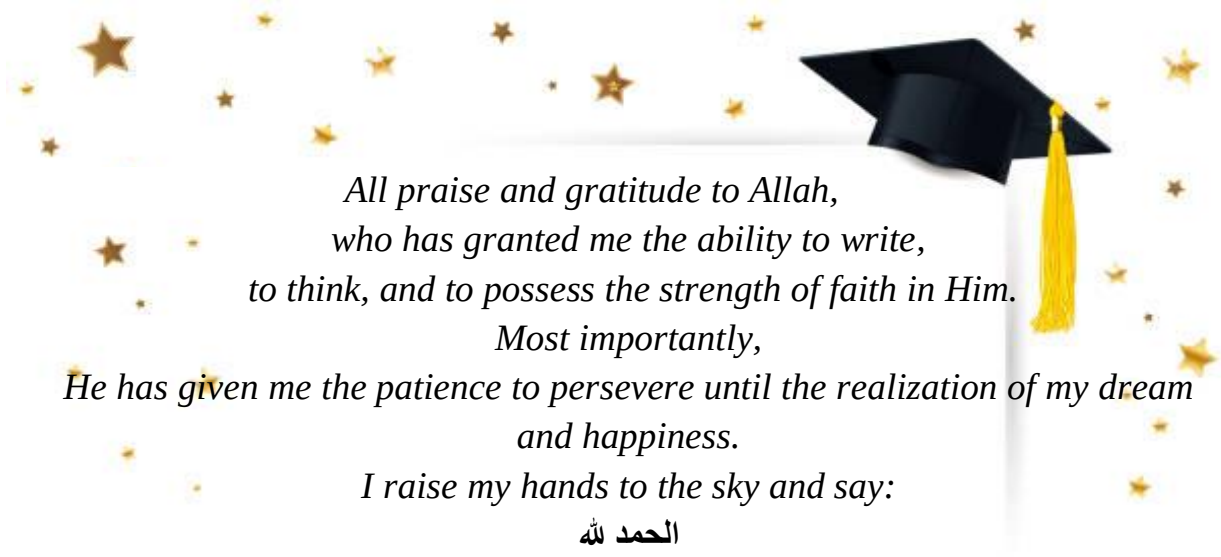
All praise and gratitude to Allah,
the source of strength and wisdom,
who guided my steps and lit my path.

*To my father **MEGAIZI M**, my rock and shield,*
whose unwavering support gave me wings to rise.
*To my dearest mother **BECHAKLIA FZ**, the heartbeat of*
my soul, whose love and prayers shaped my journey.
To my brothers and sisters, the silent warriors behind my
success, always lifting me higher.
And to every soul who lent a hand,
whispered a word of encouragement, or believed in me
this victory is not mine alone; it belongs to us all.
With gratitude,
with love,
with honor
I dedicate this work to you.

Thank you!

MEGAIZI IMANE

Dedication



All praise and gratitude to Allah,
who has granted me the ability to write,
to think, and to possess the strength of faith in Him.
Most importantly,
He has given me the patience to persevere until the realization of my dream and happiness.
I raise my hands to the sky and say:
الحمد لله

I dedicate this graduation to the one who taught me to give and whose name I carry with pride, and to the man who has always supported me:
my dear father ***DJABALI N.***

And to my angel in life, to the feeling of love, tenderness, and sincerity, to the smile of life and the secret of existence, and to the one whose prayers were the secret of my success:
my dear beloved mother ***KRIM S.***

To my steadfast support who never wavers, to the one I was blessed with as a pillar and my first and last refuge, to the one who cleared the thorns of failure from my path
– my beloved sister ***Chaima****–*

I extend my thanks and appreciation to those who encouraged and motivated me and who have had a great impact on me:
Maimouna *and my dear sister* ***Fatiha****.*

DJABALI IMENE

Table of contents

PRACTICAL PART

CHAPTER II: Methods

PRACTICAL PART

CHAPTER III: Results and discussion

List of figures

List of tables

List of abbreviation

Abbreviation	Signification
ABUS	*Automated Breast Ultrasound*
AKT	*Protein Kinase B*
AMPK	*AMP-Activated Protein Kinase*
APC	*Anaphase-Promoting Complex*
APC/C	*Anaphase-Promoting Complex/Cyclosome*
BC	*Breast Cancer*
BP	*Biological Process*
BRCA1	*Breast Cancer Gene 1*
BRCA2	*Breast Cancer Gene 2*
BSE	*Breast Self-Exams*
CC	*Cellular Component*
CEM	*Contrast-Enhanced Mammography*
CIN	*Chromosomal Instability*
Cm	*Centimeter*
DAVID	*Database For Annotation, Visualization, And Integrated Discovery*
DBT	*Digital Breast Tomosynthesis*
DCIS	*Ductal Carcinoma in Situ*
DEGs	*Differentially Expressed Genes*
DNA	*Deoxyribonucleic Acid*
E2F1	*Erythroid 2 Factor 1*
ECM	*Extracellular Matrix*
EMT	*Epithelial-Mesenchymal Transition*
ER	*Estrogen Receptors*
FC	*Fold Change*
GCO	*Global Cancer Observatory*
GEO	*Gene Expression Omnibus*
GEPIA2	*Gene Expression Profiling Interactive Analysis 2*
GO	*Gene Ontology*
GPL	*Gene Platform*

GPSM2	*G-Protein Signaling Modulator 2*
GSE	*Gene Series Expression*
HER2	*Human Epidermal Growth Factor Receptor 2*
HR	*Hormone Receptors*
HR	*Hazard Ratios*
IHC	*Immunohistochemical*
KEGG	*Kyoto Encyclopedia of Genes and Genomes*
Ki-67	*Kiel 67*
LCIS	*Lobular Carcinoma in Situ*
MAM	*Mammography*
Mad1	*Mitotic Arrest Deficient 1*
Mad2	*Mitotic Arrest Deficiency Protein 2*
MAX	*Myc-associated Factor X*
MBC	*Metastastatic Breast Cancer*
MCODE	*Molecular Complex Detection*
MF	*Molecular Function*
MiRNAs	*Micro Ribonucleic Acids*
MRI	*Magnetic Resonance Imaging*
mRNA	*Messenger Ribonucleic Acid*
MYC	*Myelocytomatosis*
OS	*Overall Survival*
P value	*Probability Value*
P21	*Protein 21*
P53	*Protein 53*
PI3K-Akt	*Phosphoinositide 3-Kinase*
PIK3CA	*Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha*
PIK3CB	*Phosphatidylinositol 4,5-Bisphosphate 3-Kinase Catalytic Subunit Beta*
PIK3R1	*Phosphoinositide-3-Kinase Regulatory Subunit 1*
PR	*Progesterone Receptors*
RNA	*Ribonucleic Acid*
SAC	*Spindle Assembly Checkpoint*
ShRNA	*Small Hairpin RNA*
STRING	*Search tool for the retrieval of interacting genes*

TFAP2C	*Transcription Factor AP-2 Gamma*
TFs	*Transcription Factors*
TNBC	*Triple-Negative Breast Cancer*
USF1	*Upstream Stimulatory Factor 1*
UTR	*Untranslated Region*

ABSTRACT

The purpose of our study was to identify key common genes associated with breast cancer and its diagnosis. The data from the three profiles were downloaded from the GEO database and analyzed using several bioinformatics tools. The common differentially expressed genes (DEGs) were determined using the *Venn diagram* visual presentation. *Gene Ontology* (GO) and *Kyoto Encyclopedia of Genes and Genomes* (KEGG) analyses were performed via the *DAVID* site, and protein-protein interaction networks (PPI) were built via the *STRING* database and visualized with *Cytoscape* software. Then, we checked the overall survival and expression of key genes using the *GEPIA2* database. 663 DEGs were obtained, of which 500 genes were up-regulated and 163 were down-regulated. According to GO and KEGG analyses these DEGs were mainly enriched in positive gene regulation, the phosphatidylinositol 3kinase/protein kinase B pathway, angiogenesis, cell division, mitotic spindle assembly checkpoint signaling, and mitotic sister chromatid segregation. In addition, eight hub genes were selected, one of which was associated with decreased overall patient survival and was significantly expressed in cancer tissue relative to normal tissue. Finally, analysis using Network Analyst revealed that UBE2C's regulation involves crucial interactions with eleven miRNAs and seven transcription factors, providing insights into its complex coregulatory network. In summary, the gene UBE2C (Ubiquitin Conjugating Enzyme E2 C) can be an excellent biomarker for breast cancer diagnosis and targeted gene therapy.

Keywords: breast cancer, bioinformatics, common differentially expressed genes, cancer development.

RESUME

Le but de notre étude était d'identifier les gènes clés communs associés au cancer du sein et à son diagnostic. Les données des trois profils ont été téléchargées à partir de la base de données GEO et analysées à l'aide de plusieurs outils bio-informatiques. Les gènes à expression différentielle communs (DEGs) ont été déterminés à l'aide de la presentation visuale *Venn diagram*. Des analyses *Gene Ontology* (GO) et *Kyoto Encyclopedia of Genes and Genomes* (KEGG) ont été effectuées via le site *DAVID*, et des réseaux d'interaction protéine-protéine (PPI) ont été construits via le site *STRING* et visualisés avec le logiciel *Cytoscape*. Ensuite, nous avons vérifié la survie globale et l'expression des gènes clés à l'aide de la base de données *GEPIA2*. On a obtenu 663 DEGs, dont 500 gènes surexprimés et 163 sous-exprimés. Les analyses GO et KEGG ont montré que ces DEGs ont été enrichis dans la régulation positive de l'expression des gènes, la régulation positive de la transduction du signal phosphatidylinositol 3 kinase/protéine kinase B, l'angiogenèse, la division cellulaire, la signalisation du point de contrôle de l'assemblage du fuseau mitotique et la ségrégation des chromatides sœurs mitotiques. De plus, huit gènes clés ont été sélectionnés, dont un était associé à une diminution de la survie globale des patients et était significativement exprimé dans les tissus cancéreux par rapport aux tissus normaux. Enfin, l'analyse utilisant la base de données NetworkAnalyst a révélé que la régulation d'UBE2C implique des interactions cruciales avec onze miRNAs et sept facteurs de transcription, fournissant des informations sur son réseau de corégulation complexe. En résumé, le gène UBE2C (Ubiquitin Conjugating Enzyme E2 C) peut être un excellent biomarqueur pour le diagnostic du cancer du sein et la thérapie génique ciblée.

Mots clés : cancer du sien, bio-informatique, les gènes à expression différentielle communs, développement du cancer.

الملخص

كان الهدف من دراستنا هو تحديد الجينات المشتركة المحورية المرتبطة بسرطان الثدي وتشخيصه. استُخرجت البيانات من ثلاث مجموعات مختلفة من قاعدة البيانات *GEO*، وخضعت لتحليل متقدم باستخدام عدد من أدوات المعلوماتية الحيوية. جرى تحديد الجينات مختلفة التعبير (*DEGs*) باستخدام المخطط المرئي *venn diagram*، فيما أُجريت تحليلات التأشير الجيني (*GO*) ومسارات *Kyoto Encyclopedia of Genes and Genomes (KEGG)*عبر منصة *DAVID* .كما أُنشئت شبكات التفاعل البروتيني (PPI) بالاعتماد على قاعدة *STRING* ، وتم تصويرها باستخدام برنامج *Cytoscape* . بعد ذلك، لتحقيق فهم أعمق لأهمية هذه الجينات، أُجري تحليل معدل البقاء الكلي والتعبير الجيني للجينات الرئيسية باستخدام قاعدة البيانات *GEPIA2* . بلغ عدد الجينات مختلفة التعبير 663 جينًا، منها 500 مفرطة التعبير، و163 منخفضة التعبير. وقد أظهرت نتائج تحليل *GO* وKEGG أن هذه الجينات تشارك في تنظيم التعبير الجيني الإيجابي، مسار فوسفاتيديل اينوزيتول 3 كينار/ بروتين كينازB، تكوّن الأوعية الدموية، انقسام الخلية، مراقبة تجمع المغزل الانقسامي، بالإضافة إلى فصل الكروماتيدات الشقيقة خلال الانقسام الخلوي. ومن بين الجينات التي تم تحديدها، برزت ثماني جينات رئيسية، أحدها ارتبط بشكل واضح بانخفاض في معدل البقاء الكلي للمرضى، كما أبدى تعبيرًا مرتفعًا في الأنسجة السرطانية مقارنة بالنسيج الطبيعي. أخيرًا، كشف تحليل Network Analyst أن تنظيم الجين UBE2C يتضمن تفاعلات مهمة مع أحد عشر miRNAs وسبعة عوامل نسخ، مما يوفر معلومات قيمة حول شبكته التنظيمية المعقدة. وبناءً على ما سبق، يُعد الجين UBE2C (Ubiquitin Conjugating Enzyme E2 C) مرشحًا واعدًا ليكون مؤشرًا حيويًا لتشخيص سرطان الثدي، بالإضافة إلى كونه هدفًا محتملاً للعلاج الجيني الموجّه.

الكلمات المفتاحية: سرطان الثدي، المعلوماتية الحيوية، الجينات مختلفة التعبير، تطور السرطان.

INTRODUCTION

Breast cancer is the most common cancer in women and the second-leading cause of cancer-related deaths among women worldwide. Breast cancer makes up about 24.2% of all new cancer cases in women. It is expected to represent 1.7% of all cancer cases by 2025, with around 2.3 million new cases in women, surpassing lung cancer as the leading cause of cancer worldwide. A review of the literature reports that in 2018, there were about 2 million new breast cancer cases and 626,679 deaths. This cancer can also affect younger women **[1].**

Breast cancer, a complex condition characterized by major variance at both the genetic and clinical levels, stands out in this globally difficult situation. Usually grouped according to their spread, structure, and expression of particular immunohistochemical markers, along with current genetic analysis, most breast cancers are adenocarcinomas. Research has revealed that different traits are connected to differences in the cancer's treatment response and its probable outcomes **[2]**. In situ breast cancers are also limited to the ducts or lobules of the breast **[3]**, where ductal carcinoma in situ (DCIS) is thought to be more prevalent than lobular carcinoma in situ (LCIS). Although the exact causes of the formation and advancement of in situ tumors are not completely known, lobular carcinoma in situ is not thought to have the possibility to mutate into a malignant tumor **[4-5]**; still, both types are considered risks for invasive breast cancer **[6].**

Even though breast cancer therapy has come a long way, it is still the most prevalent cancer, with the highest incidence rates among women worldwide. Thus, precise diagnostics, efficient therapy, and improved prediction of the disease's outcome would therefore depend greatly on one's knowledge of the biological pathways behind breast cancer **[7]**.

The identification of novel biomarkers for BC is crucial to its diagnosis, therapy, and prognosis. In the last decade, the rapid development of high-throughput techniques and public gene databases enables to filter out a wider range of disease-associated genes based on abundant data by utilization of microarrays, analyze them holistically, and thus identify potential new drug targets for early diagnosis and treatment **[8].** The use of bioinformatics tools to analyze data from gene expression research offers a promising field that opens new possibilities and helps to develop creative diagnostic and therapeutic approaches **[9]**.

Our study is based on the use of multiple bioinformatics tools and databases to extract and analyze a dataset in order to identify common differentially expressed genes in breast cancer tissues. Thus, our study aimed to identify new biomarkers involved in BC progression that could be used as potential targeted therapies.

This work is divided into three parts:

➤ The theoretical part is specific to breast cancer: Risk factors of breast cancer and symptoms, types, stages, diagnosis, and treatment.

➤ The first practical part (Methods) is dedicated to the bioinformatics tools used for analysis of data from patients with breast cancer.

➤ The results and discussion section which shows the different data analyses and opposes them with scientific arguments.

THEORETICAL PART

CHAPTER I: Breast cancer

1. Breast cancer

Among the most commonly diagnosed cancers is BC, which ranks second as the leading cause of cancer-related deaths among women **[10]**. Breast cancer is when the cells in the breast grow abnormally **[11].**

Apart from being the most frequently diagnosed cancer in 154 out of 185 countries, BC is the leading cause of cancer-related deaths in over 100 countries. According to the American Cancer Society, around 600 men and 40,000 women lose their lives to BC each year **[12].**

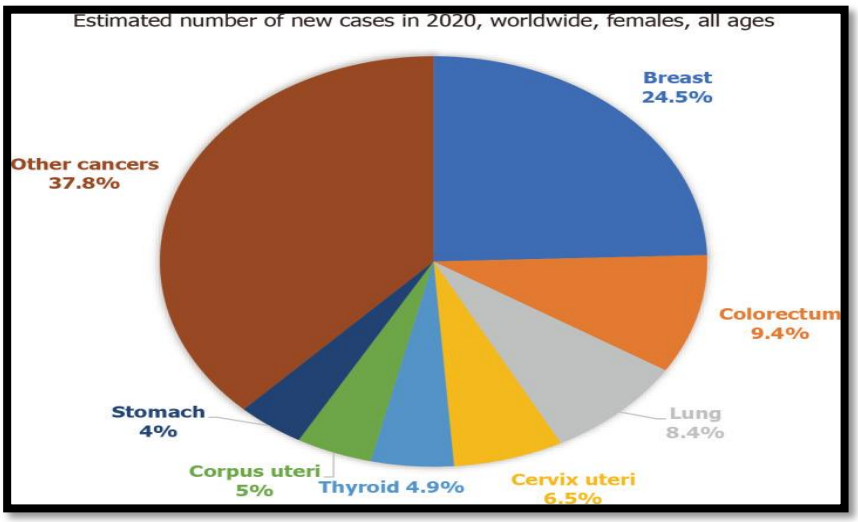


Figure 01: Estimated number of new cancer cases in 2020, worldwide **[13].**

2. Breast cancer risk factors

2.1. Demographic and genetic factors

- **Age**

Most BC patients are over 50 years old, with more than 40% being over 65 years old. The risk of getting breast cancer goes up with age: it's 1.5% at 40, 3% at 50, and over 4% at 70 **[14].**

- **Race or ethnicity and geographic location**

BC affects different races and ethnicities in different ways. The rates vary by race, and there are concerning differences in survival rates for women with breast cancer .

Black women have a higher risk of dying from breast cancer, but white women are more likely to be diagnosed with it **[15].**

- **Family history**

A family history of BC is an important factor strongly linked to a higher likelihood of developing BC. About 13-19% of individuals diagnosed with breast cancer mention having a first-degree relative who has the same condition. Additionally, the risk of BC increases significantly as the number of first-degree relatives affected by the disease grows **[14].**

- **Mutations**

Mutations in the BRCA1 and BRCA2 genes increase the risk of BC. These genes help repair DNA damage in the body, but when mutations occur, the body becomes less able to fix the damage, which increases the likelihood of cancer. Women with these mutations have a higher risk of developing breast and ovarian cancer **[16].**

2.2. Hormonal factors

Hormonal imbalance is one of the major causes of BC; hormones like estrogen and progesterone stimulate the growth of breast cells by binding with the ER and PR receptors. Once estrogen binds to ER or progesterone to PR, a cascade of signals inside the cell is triggered that can regulate cell growth and inhibit cell death. But if there's too much estrogen, or the signaling with progesterone is interrupted, these signals can become uncontrolled, triggering uncontrolled cell growth and tumor formation **[17].**

2.3. Lifestyle factors

- **Smoking**

The cancer-causing substances in tobacco smoke move through the lung alveoli into the bloodstream, where they can then be carried to the breast via lipoproteins in the plasma **[18].**

- **Obesity**

Obesity is linked to an increased risk of BC because it raises levels of certain hormones and receptors, like estrogen, which is associated with estrogen-positive BC. It can also cause genetic changes due to the large amount of fat tissue, leading to insulin imbalances. Additionally, obesity can trigger inflammation in the body, which may increase the chances of developing or worsening cancers, including BC **[19].**

- **Alcohol consumption**

Although alcohol consumption is associated with increased risk of BC, the most researched explanations include how alcohol affects estrogen levels in the body and how it interacts with receptors in breast cells, as well as the harmful effects of alcohol's breakdown products. New studies suggest that alcohol might also contribute to BC by affecting a process called epithelial-mesenchymal transition (EMT), which is involved in how cells change and spread **[20].**

2.4. Environmental factors

BC environmental risk factors involve contact with substances like pesticides, industrial pollutants, chemicals found in common consumer products, and physical agents such as radiation (often from X-rays) **[21].**

3. Symptoms

BC symptoms can vary, and some people may not have any symptoms in the early stages. Common signs include a lump or thickening in the breast or underarm; changes in the size, shape, or appearance of the breast, changes in the nipple (like it turning inward or leaking fluid), and skin changes such as, redness or the appearance of small dimples **[22].**

4. Breast cancer types

4.1. Molecular breast cancer subtypes:

BC is also classified into subtypes based on its molecular characteristics, such as the presence of certain hormone receptors. Understanding these patterns helps guide treatment more effectively.

There are four main patterns of molecular breast cancer subtypes:

* Luminal A
* Luminal B
* HER2-positive (HER2-enriched)
* Triple-negative **[23]**

4.1.1. Luminal A breast cancer

Luminal A breast cancer is characterized by the presence of hormone receptors (positive for estrogen and/or progesterone) and the absence of HER2 (negative for HER2), in addition to low levels of Ki-67, which indicates slower growth of cancer cells **[23].**

4.1.2. Luminal B breast cancer

Luminal B breast cancer is characterized by the presence of hormone receptors (positive for estrogen and/or progesterone) and may or may not have HER2 (HER2-positive or HER2-negative), in addition to high levels of Ki-67, which indicates faster growth of cancer cells **[23].**

4.1.3. HER2-enriched breast cancer

HER2-enriched breast cancer is characterized by the presence of a large amount of HER2 protein on the surface of cancer cells. This subtype is typically ER-negative and PR-negative and accounts for about 10-15% of all cases **[23].**

4.1.4. Triple negative breast cancer

Triple-negative breast cancer (TNBC) is diagnosed when the results of the immunohistochemical (IHC) test are negative for ER, PR, and HER2 receptors, which means the absence of these receptors in the cancer cells. TNBC accounts for about 10-20% of all invasive breast cancers and is characterized by its aggressive nature **[24].**

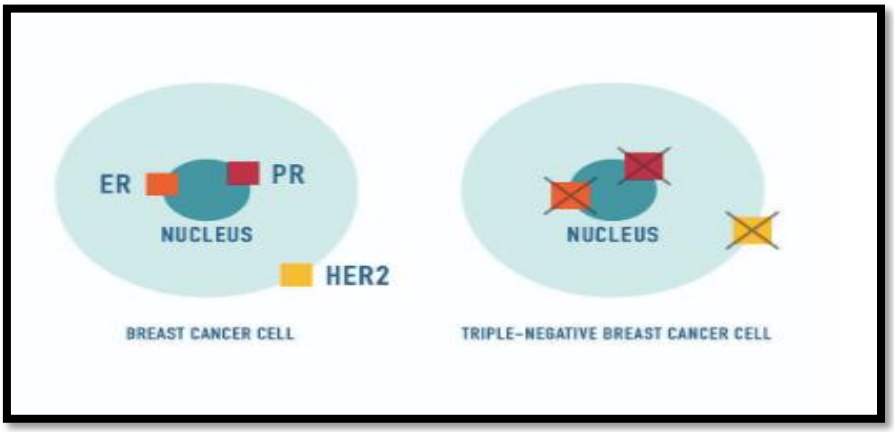


Figure 02: Illustration compares typical breast cancer cells (with ER, PR, and/or HER2) and triple-negative breast cancer cells (lacking all three) **[25].**

The Comparison between common molecular breast cancer subtypes is shown in Figure 03.

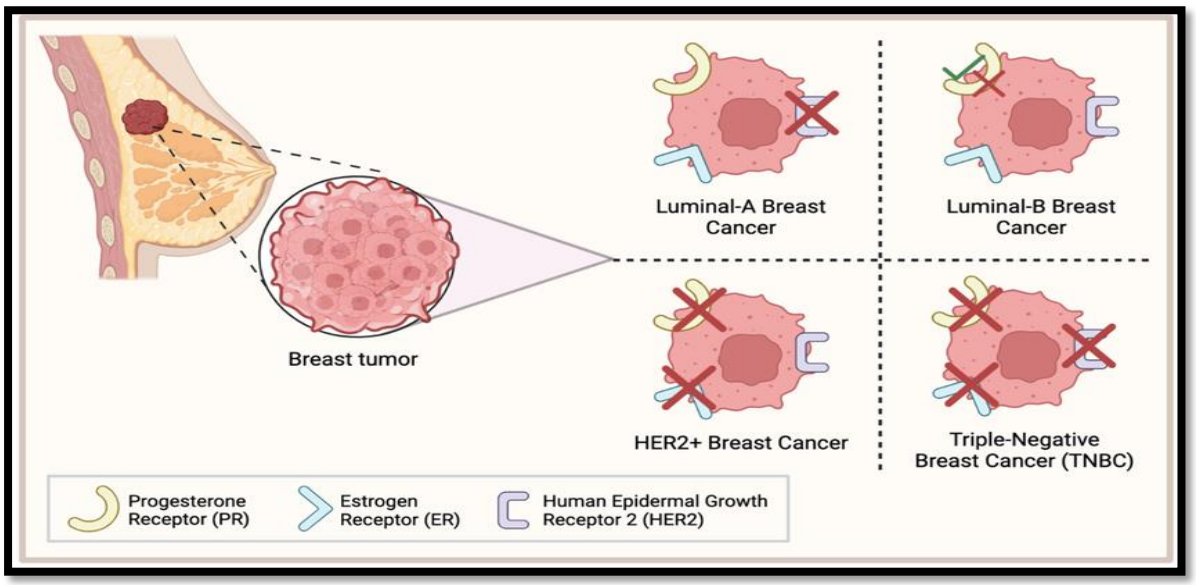


Figure 03: Classification of common molecular breast cancers subtypes **[26].**

5. Stages of breast cancer

- ✓ **Stage 0:**

Abnormal cells are found in the lining or parts of the breast ducts, which increases the risk of cancer developing in one or both breasts. This is a non-invasive tumor, smaller than two centimeters in size **[27].**

- ✓ **Stage 1 (Early stage):**

Cancer has spread to a small area of nearby tissue. This is an invasive tumor, smaller than two centimeters in size **[27].**

- ✓ **Stage 2 (Localized):**

The tumor measures between 2-5 cm, and some lymph nodes are affected. In some cases, the tumor is larger than 5 cm, with no lymph node involvement **[27].**

- ✓ **Stage 3 (Regional spread):**

The tumor exceeds 5 cm in size, with increased involvement of lymph nodes over a larger area. In some cases, no tumor may be present. The cancer could have spread to the chest wall and skin **[27].**

✓ **Stage 4 (Distant spread):**

Cancer has moved from the breast to other parts of the body nearby **[27].**

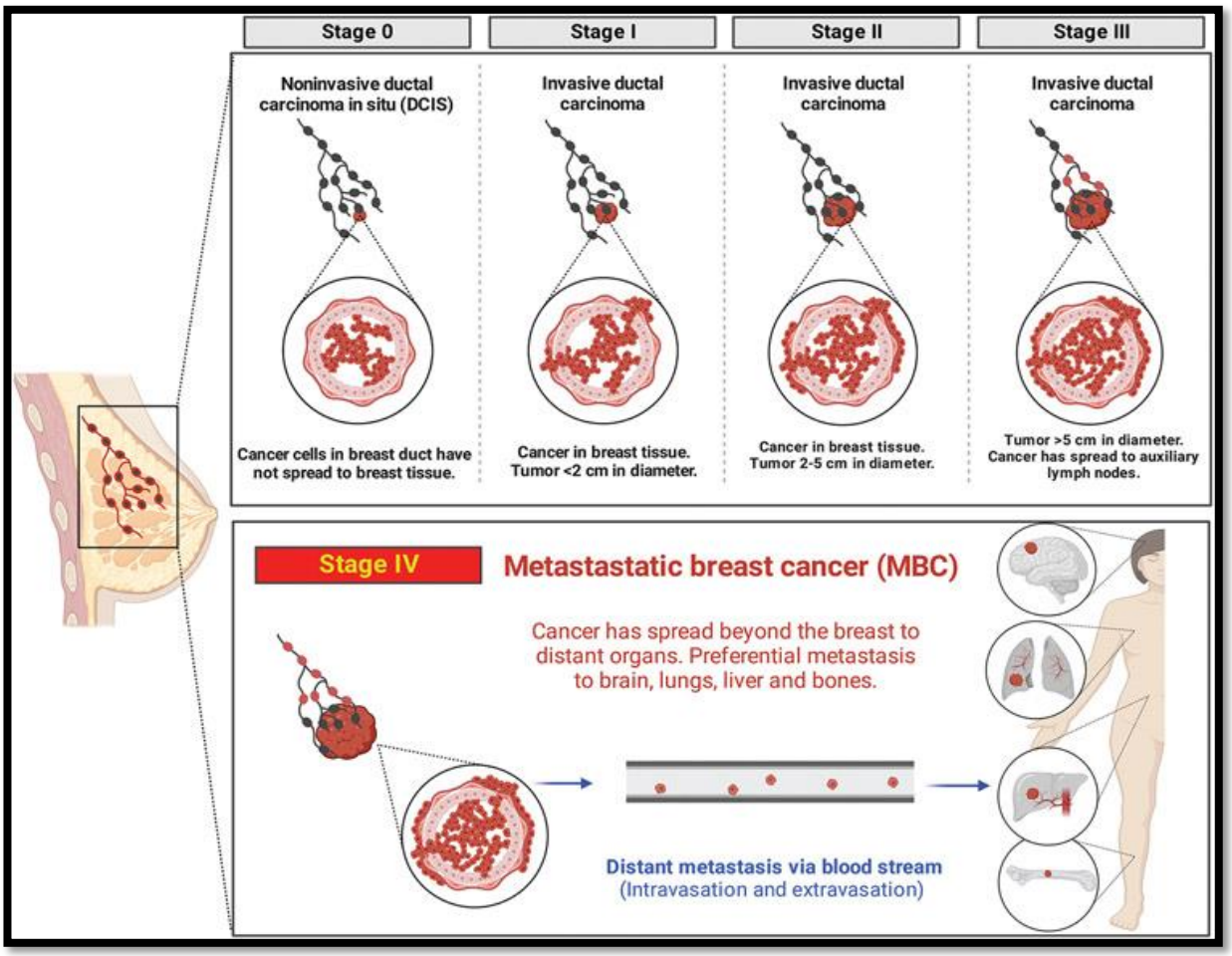


Figure 04: Illustrations showing the mechanisms of different stages of breast cancer, tumor size at each stage, 5-year survival rates, and available treatment options for breast cancer **[28].**

6. Methods of breast cancer screening

6.1. Clinical breast examination

Table 01: Comparison of common screening techniques **[29].**

Tech	Current recommendations	Advantages	Disadvantages
MAM	Recommended for women who have reached the initial age for BC screening	Convenient; economical; sensitive to microcalcification	Insensitive to high density breasts and deep lesions
DBT	Recommended, particularly for women with dense breasts, replacing DM	3D imaging, reducing tissue overlay	Greater radiation dose, examination time and cost
CEM	Recommended for women at high risk	Vascular functional imaging	Use of contrast agents; greater radiation dose
Ultra-sound	Recommended, particularly for women with dense breasts, and pregnant and lactating women	Noninvasive; real-time; no radiation; elastography; identification of cystic and solid masses	Dependent on technologist experience; insensitive to lesions without clear mass
ABUS	Recommended, particularly for women with dense breasts, and pregnant and lactating women	Less dependent on technologist skill; reproducible	Unable to assess the axillary lymph node status
MRI	Recommended for women at high risk	Most precise identification of soft tissue; reflects both the anatomical structure and lesions; displays small lesions, multifocal lesions.	Insensitive to calcification; expensive; long examination time; use of contrast agents.

6.2. Physical Examination

6.2.1. Breast self-examination (BSE)

Numerous awareness campaigns urge women to perform monthly self-examinations of the breast. However, the evidence linking breast self-exams to a reduction in mortality is scarce **[30].**

BSE can be performed in different ways:

- **Visual Inspection:** The breast should be examined in front of a mirror while standing and lying down, with the arm on the same side lifted. The goal is to look for any changes such as asymmetry, dimpling, inversion of the nipple, or redness.
- **Palpation:** Using the pads of the three middle fingers, gently press on the breast in a firm but circular motion, systematically checking the entire breast and underarm area.
- **Checking for Nipple Discharge:** Gently squeeze the nipple to check for any unusual discharge **[31].**

7. Breast cancer therapy

There are many different treatment options, and this depends on the individual, as well as the type and stage of cancer. Adjuvant therapy is when it's used after surgery and there's an absence of signs indicating cancer recurrence. Neoadjuvant therapy is when it's used before surgery to shrink the tumor and make it easier to remove, or to see how the tumor responds to a certain type of treatment. Other treatment types include:

- **Surgery:** A surgical procedure to remove the cancerous tumor.
- **Radiation therapy:** The use of high-energy X-rays to kill cancer cells.
- **Chemotherapy:** The use of powerful drugs to kill cancer cells.
- **Hormone therapy:** The use of hormones or anti-hormonal drugs to slow down or stop the growth of cancer cells, particularly in hormone-dependent cancers.
- **Targeted therapy:** The use of drugs that target specific molecules found in cancer cells, minimizing damage to healthy cells **[32].**

Table 02: breast cancer therapy **[33].**

Treatment	What Is It?	How Is It Given?	What Makes It Unique?	Common Side Effects
Surgery	Operative procedure to remove the cancer tumor from the breast	n/a	It can be performed as the only treatment, or prior to or after chemotherapy	Seroma, hematoma, lymphedema, pain, infection, cosmetic issues
Chemo-therapy	Drugs that kill cancer cells directly	Oral or intravenously	The drugs can stop or slow cancer growth and harm healthy cells	Fatigue, hair loss, bruising or bleeding, infection, anemia, nausea, vomiting, appetite changes, constipation
Hormone Therapy	Medication that is specific to types of BC that are promoted by hormones (Eg, estrogen and progesterone)	By mouth or injection in the skin	These drugs work by stopping hormones from fueling breast cancer growth	Hot flashes, vaginal dryness, night sweats, bone pain
Targeted Therapy	A drug that alters the behavior of a cancer cell	Intravenously or oral	It interferes with specific molecules involved in tumor growth and progression	Heart impairment, diarrhea, shortness of breath
Radiation	Special high-energy beams to damage cancer cells	By a machine called a linear accelerator	Targeted way to destroy cancer cells in the breast that may remain after surgery	Skin color changes such as pink or redness; itching, soreness; and possibly peeling

- Certain types, like triple-negative breast cancer, are especially hard to treat. Even though monotherapies treatments (targeted therapy, hormone therapy) can be effective for some patients, their effectiveness can decline over time, and some patients may become resistant to them. While there have been some positive outcomes with these treatments, they don't work well for everyone, especially for patients with advanced stage of BC.
- A promising new concept is combination therapy, where multiple treatments are used together to combat the disease. Different forms of combination therapy are already being applied in clinical settings and trials. Several new approaches to combining treatments have been introduced. As this treatment method becomes more common, advancements in technology are expected to bring even more creative solutions **[34].**

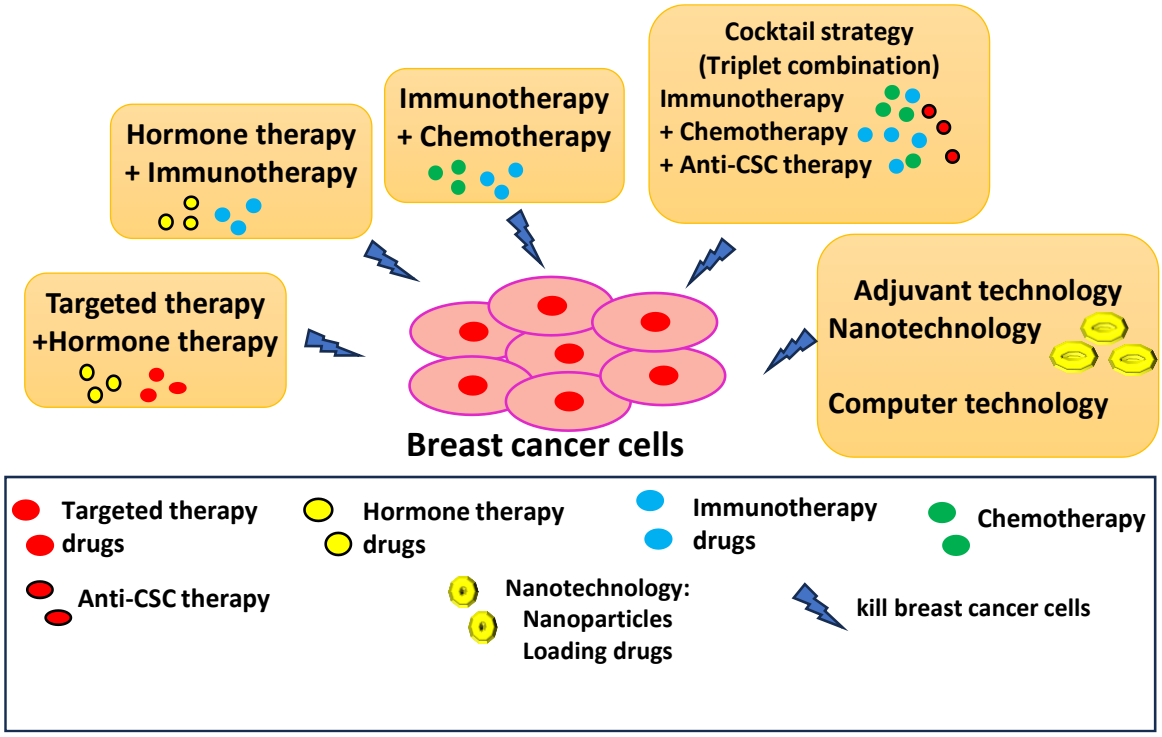


Figure 05: Schematic illustration of current treatments for breast cancer **[34].**

PRACTICAL PART

CHAPTER II: Methods

2. Methods

2.1. Study objective

The aim of this study was to analyze the differential expression of genes associated with BC using integrated bioinformatics tools.

2.2. Methods of analysis

2.2.1. Microarray data source and information

The microarray data used in this study were downloaded from the public functional genomics data repository, the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/gds/). Three gene expression profiles (**GSE29431, GSE42568,** and **GSE61304**) of BC and normal breast tissue were chosen. The array data for GSE29431 consisted of 54 primary breast carcinomas and 12 samples of normal breast tissues, the array data for GSE42568 consisted of 104 breast cancer and 17 normal breast biopsies, the array data for GSE61304 included four RNA samples obtained from normal individuals and 58 breast tumors. All the three datasets were based on the GPL570 platform ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array).

2.2.2. Data processing and identification of DEGs

The three mRNA expression microarray datasets obtained from the GEO database were analyzed using the GEO2R online tool (https://www.ncbi.nlm.nih.gov/geo/geo2r/).
The |log2-fold change (FC)| ≥ 1 and P value ≤ 0.05 were conducted as the cutoff criteria for the differentially expressed genes (DEGs). Then, the raw data in TXT format were selected using Venn diagram software (http://bioinformatics.psb.ugent.be/webtools/Venn/) to obtain the common DEGs among the three datasets. DEGs with logFC ≥ 1 were considered up-regulated genes, while DEGs with logFC ≤ -1 were considered down-regulated genes.

2.2.3. Gene ontology and pathway enrichment analysis

Gene ontology (GO) analysis is a commonly used approach for functional studies of high-throughput transcriptome or genome data. GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using the online bioinformatics tool Database for Annotation, Visualization, and Integrated Discovery (DAVID) software (version 6.8, https://david.ncifcrf.gov/). P value ≤ 0.05 was set as the cutoff criterion.

2.2.4. PPI network and module analysis

To explore the interactions of DEGs, we used the string database (version 11.5, https://string-db.org/). A high confidence level (0.700) was chosen as the minimum required interaction score. After removing the disconnected nodes, protein-protein interaction (PPI) networks were visualized using Cytoscape 3.9.0 software. In addition, the Molecular Complex Detection plug-in of Cytoscape software was used to screen modules of the PPI network (degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and max. depth = 100).

2.2.5. Selection of hub genes

Hub genes were extracted using the Cytoscape plug-in CytoHubba software. Then, the top eight essential hub genes were selected and ranked by Maximal Clique Centrality (MCC) and Degree topological algorithms.

2.2.6. Survival analysis and expression validation of hub genes

The online database Gene Expression Profiling Interactive Analysis GEPIA2 (http://gepia2.cancer-pku.cn/) was utilized to assess the effect of eight core genes on the overall survival (OS) of BC patients. To estimate the OS rate of patients with BC, samples were split into high- and low-expression groups based on the median expression. The hazard ratios (HR), corresponding 95% confidence intervals, and log-rank P values were calculated and displayed on the plot. (The GEPIA2 database tool was used to analyze the RNA sequencing expression data between BC and control samples).

2.2.7. Transcription factor-miRNA target gene

To explore the human transcription factors (TFs)-miRNA coregulatory interaction of the related key gene UBE2C, the Network Analyst database (https://www.networkanalyst.ca/) was used. The RegNetwork repository provided data on TF-miRNA coregulatory interactions, which facilitated the identification of regulatory TFs and miRNAs that regulate the gene UBE2C.

PRACTICAL PART

CHAPTER III: Results and discussion

3. Results

3.1. DEGs identification

DEGs were screened among each microarray dataset using GEO2R online tool with P value ≤ 0.05 and |logFC| ≥ 1.

The GSE61304 dataset, included 1881 up-regulated genes and 1035 down-regulated genes.

The GSE42568 dataset, contained 2991 up-regulated genes and 2787 down-regulated genes.

The GSE29431 dataset, revealed 3206 up-regulated genes and 1067 down-regulated genes.

The Volcano Plots of the three GEO datasets are presented in Figure 06.

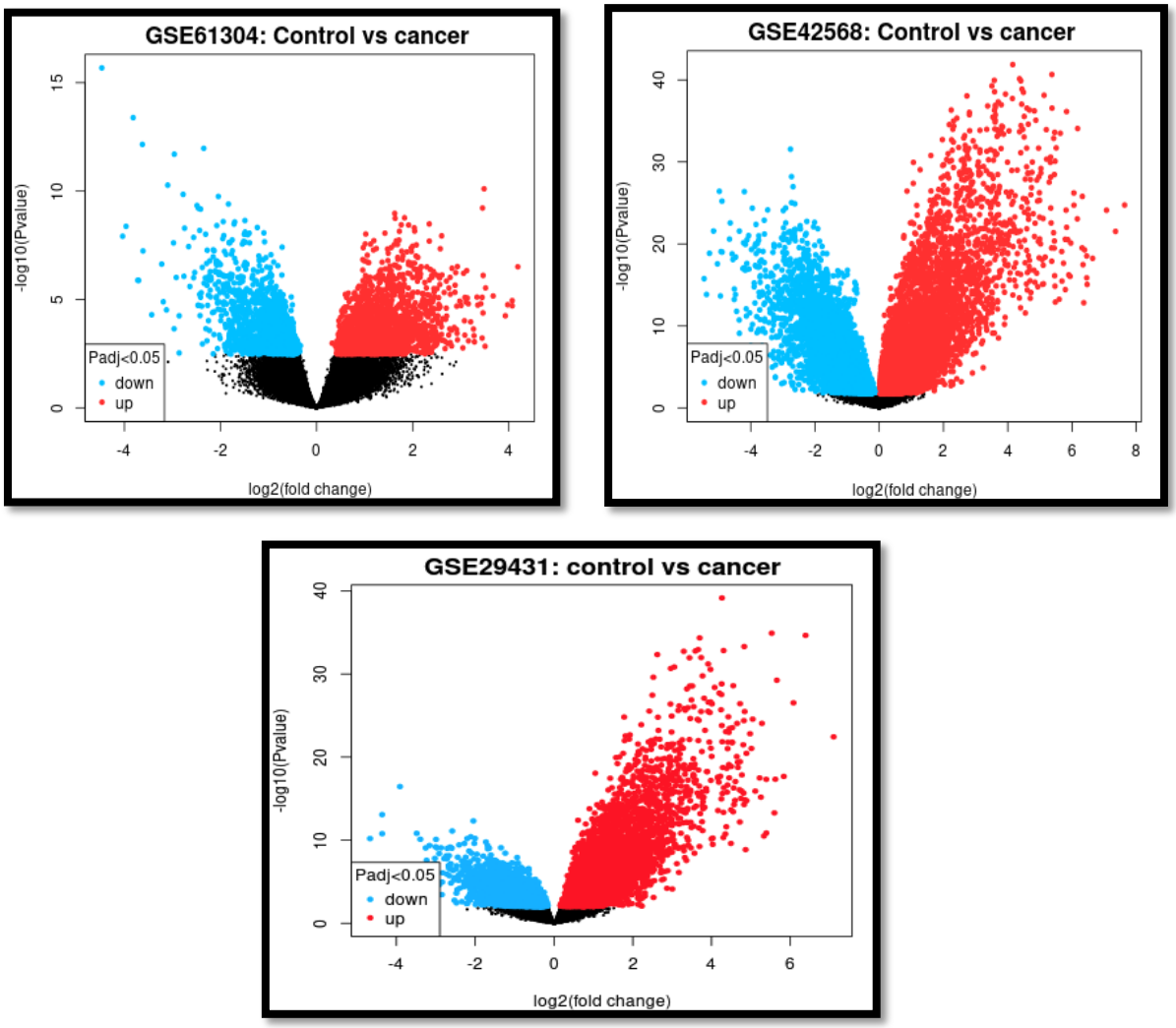
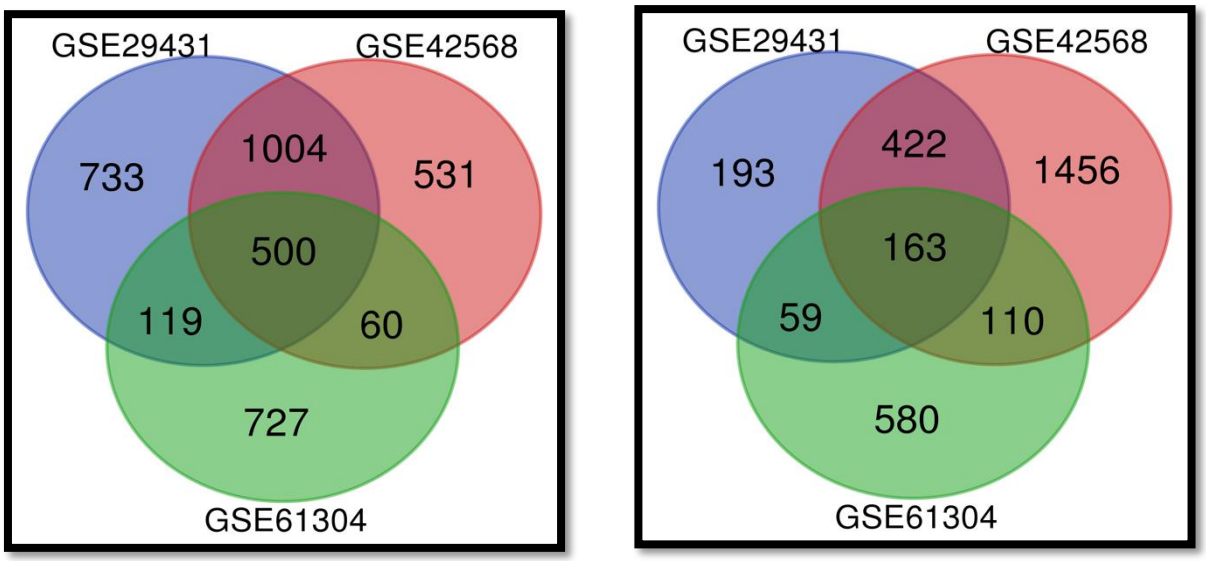


Figure 06: Volcano plots of the three GEO datasets, blue represents down-regulated genes, red represents up-regulated genes (P ≤ 0.05 and |logFC| ≥ 1).

Genes were determined using a Venn diagram software of common differential expression (DEGs) between three profiles. A total of 663 common DEGs were found, including 500 up-regulated genes and 163 down-regulated genes. They were identified in all three GEO datasets, as shown in Figure 07.



A- Up-regulated ***B- Down-regulated***

Figure 07: Identification of common DEGs in the three datasets.

A, 500 DEGs were up-regulated. B, 163 were down-regulated.

The list of common up-regulated and down-regulated DEGs from the three datasets are shown in the Table in the Annexes.

3.2. Gene ontology and pathway enrichment analysis

An enrichment analysis of GO was performed for common DEGs. The selection criterion for the analysis was P value ≤ 0.05. The analysis includes BP, CC and MF.

- In BP, up-regulated genes were primarily enriched in the following pathways: positive regulation of gene expression, positive regulation of phosphatidylinositol 3-kinase/protein kinase B signal transduction, angiogenesis, negative regulation of angiogenesis and response to glucose. The down-regulated genes were enriched in the following pathways: cell division, mitotic spindle assembly checkpoint signaling, mitotic sister chromatid segregation, chromosome segregation and mitotic cell cycle.
- In CC, up-regulated genes were more enriched in the following pathways: the collagen-containing extracellular matrix, caveola, extracellular region, cell surface and extracellular space. The down-regulated genes were more enriched in spindle, kinetochore, midbody, mitotic spindle and centrosome.
- In MF, up-regulated genes were extensively enriched in the following pathways: integrin binding, DNA-binding transcription factor activity, heparin binding, DNA-binding transcription activator activity, RNA polymerase II-specific and extracellular matrix structural constituent. The down-regulated genes were primarily enriched in the following pathways: microtubule binding, microtubule motor activity, protein binding, extracellular matrix structural constituent and extracellular matrix structural constituent conferring tensile strength.

The analysis of the GOs of the common DEGs is presented in Table 03.

Table 03: analysis of the GO of the common DEGs expressed in the three datasets.

Category	Term	Count	%	P-Value
	Up-regulated genes			
GOTERM_BP_DIRECT	positive regulation of gene expression	36	8.0	7.5E-10
	positive regulation of phosphatidylinositol 3-kinase/protein kinase B signal transduction	19	4.2	1.9E-7
	angiogenesis	21	4.7	9.0E-7
	negative regulation of angiogenesis	14	3.1	5.2E-6
	response to glucose	10	2.2	6.5E-6
GOTERM_CC_DIRECT	collagen-containing extracellular matrix	33	7.3	4.3E-11
	caveola	14	3.1	4.2E-9
	extracellular region	87	19.3	1.2E-8
	cell surface	37	8.2	1.7E-7
	extracellular space	76	16.9	2.1E-7
GOTERM_MF_DIRECT	integrin binding	16	3.5	2.3E-6
	DNA-binding transcription factor activity	31	6.9	1.7E-5
	heparin binding	15	3.3	3.4E-5
	DNA-binding transcription activator activity, RNA polymerase II-specific	26	5.8	7.6E-5
	extracellular matrix structural constituent	11	2.4	2.7E-4
	Down-regulated genes			
GOTERM_BP_DIRECT	cell division	34	22,8	5,1E-26
	mitotic spindle assembly checkpoint signaling	10	6,7	8,7E-13
	mitotic sister chromatid segregation	10	6,7	3,9E-12
	chromosome segregation	13	8,7	8,5E-12
	mitotic cell cycle	14	9,4	2,6E-11
GOTERM_CC_DIRECT	spindle	18	12,1	6,8E-16
	kinetochore	16	10,7	4,9E-13
	midbody	15	10,1	9,8E-11
	mitotic spindle	13	8,7	3,3E-10
	centrosome	23	15,4	1,5E-9
GOTERM_MF_DIRECT	microtubule binding	19	12,8	3,8E-12
	microtubule motor activity	9	6,0	1,5E-8
	protein binding	129	86,6	1,6E-6
	extracellular matrix structural constituent	9	6,0	3,8E-6
	extracellular matrix structural constituent conferring tensile strength	6	4,0	2,5E-5

3.3. Analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG)

The selection criterion for the analysis of KEGG was P value ≤ 0.05.

Up-regulated genes were primarily enriched in the following pathways: Regulation of lipolysis in adipocytes, Cytoskeleton in muscle cells, AMPK signaling pathway, PPAR signaling pathway, PI3K-Akt signaling pathway. The down-regulated genes were enriched in the following pathways: Cell cycle, Oocyte meiosis, Motor proteins, ECM-receptor interaction, Cytoskeleton in muscle cells.

Table 04: KEGG analysis of common DEGs in breast cancer.

Category	Term	Count	%	P-Value
Up-regulated genes				
KEGG_PATHWAY	Regulation of lipolysis in adipocytes	12	2.7	2.0E-7
	Cytoskeleton in muscle cells	22	4.9	3.7E-7
	AMPK signaling pathway	15	3.3	2.2E-6
	PPAR signaling pathway	11	2.4	1.9E-5
	PI3K-Akt signaling pathway	21	4.7	8.2E-4
Down-regulated genes				
KEGG_PATHWAY	Cell cycle	18	12,1	9,4E-15
	Oocyte meiosis	9	6,0	2,1E-5
	Motor proteins	10	6,7	3,9E-5
	ECM-receptor interaction	7	4,7	1,0E-4
	Cytoskeleton in muscle cells	10	6,7	1,4E-4

3.4. Protein-protein interaction analysis

To better assess the interactions between the identified common DEGs, we performed the following: Create a PPI network using the *STRING* database. A high confidence level (0.700) was chosen as the minimum required interaction score. After removing the disconnected nodes, protein-protein interaction (PPI) networks were visualized using Cytoscape 3.9.0 software.

The PPI network of common DEGs is shown in Figure 08.

The plotted PPI network is composed of 65 up-regulated genes and 43 down-regulated genes.

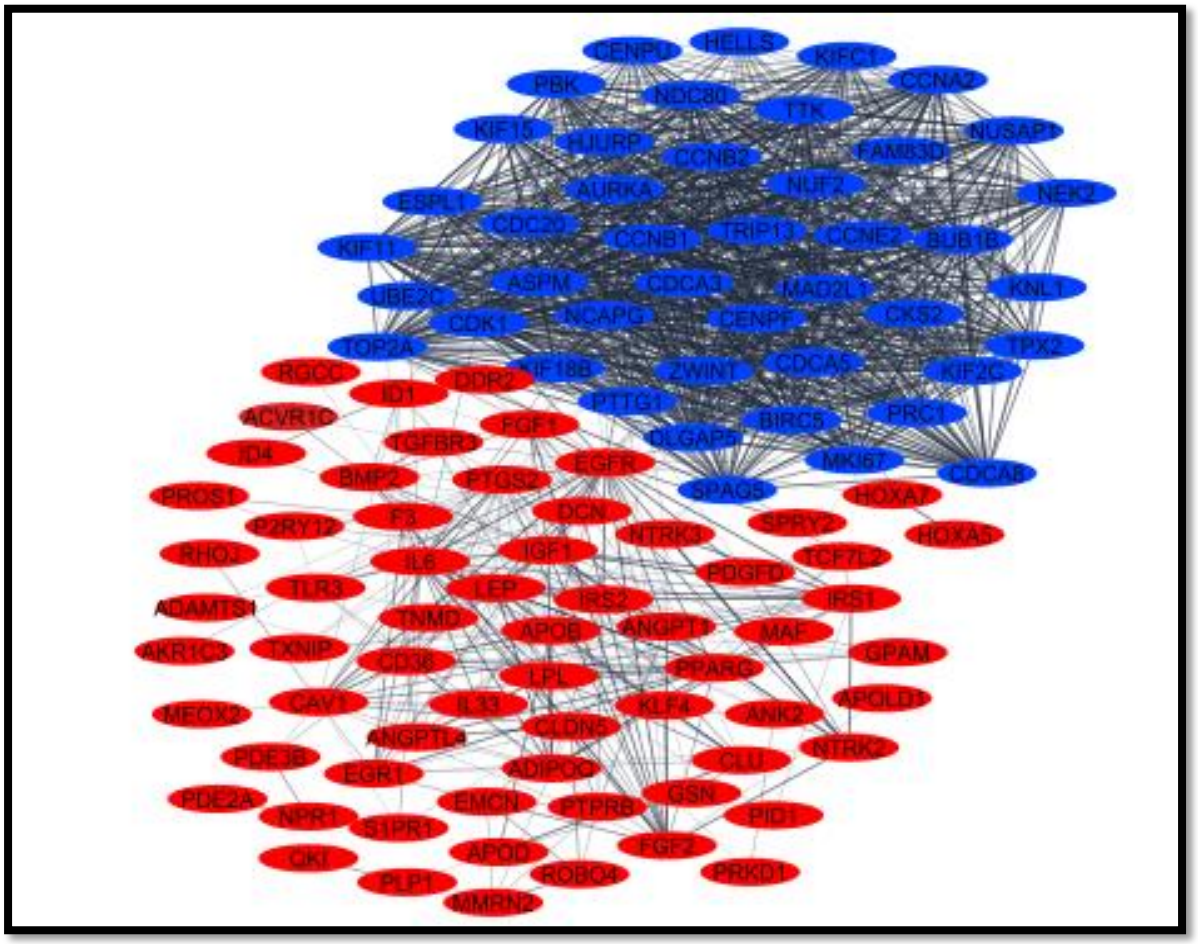


Figure 08: The PPI network created by the STRING tool and visualized by Cytoscape software (degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and max. depth = 100). Red color indicates up-regulated genes; blue color shows the down-regulated genes.

3.5. Selection of hub genes

From CytoHubba, we found a functional unit of PPI network. Contained eight genes: **BUB1B, CDC20, CDK1, MAD2L1, UBE2C, TPX2, PBK, TOP2A**. The 8 hub genes were selected based on their high connectivity and centrality in the PPI network as determined by topological algorithms and Maximal Clique Centrality. These eight genes are considered as the hub genes associated with pathogenesis of BC (**Figure 9**).

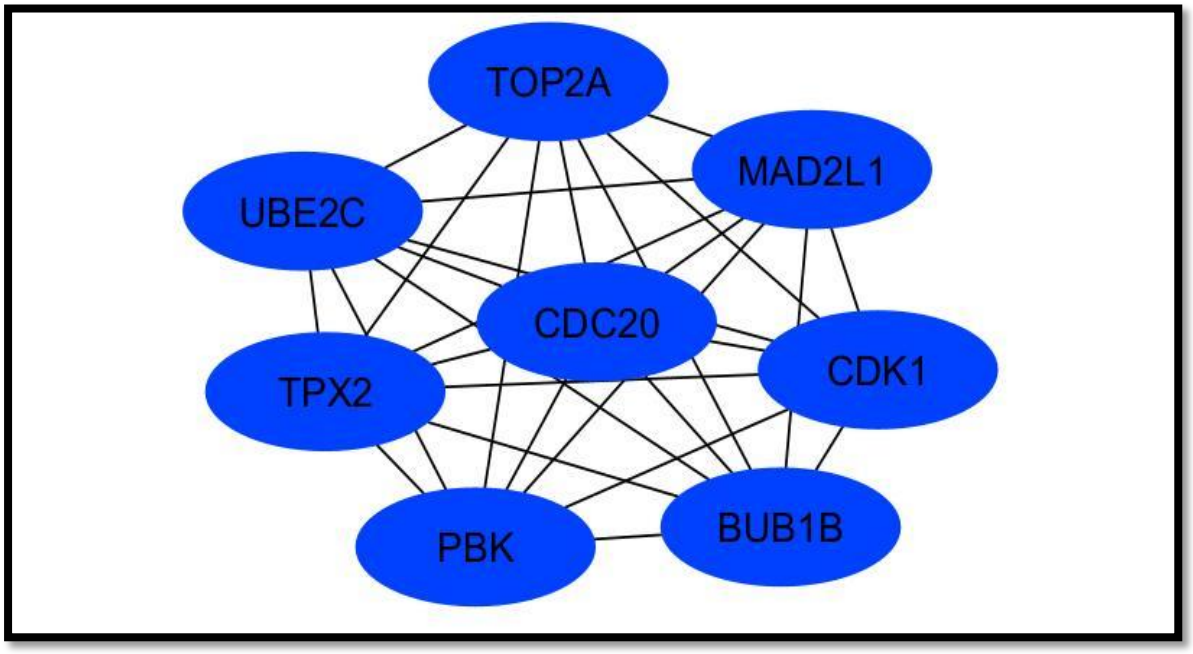


Figure 09: Top eight hub genes ranked by MCC and degree topological algorithms**.** All of them were down-regulated genes.

3.6. Survival analysis of hub genes

The survival analysis is used to determine the relationship between the expression profile of one or more genes and patient survival time. In this study, the overall survival of hub genes was analyzed using the GEPIA2 database to explore their prognostic value in breast cancer.

To perform the analysis, patients were divided into two groups based on the median expression value of each hub gene: a high-expression group and a low-expression group.

The log-rank test was applied as a standard method to evaluate whether these differences were statistically significant.

Each survival plot demonstrates how the expression level of a specific gene affects patient survival over time (measured in months). In these plots:

- The red line represents patients with high gene expression.
- The blue line represents patients with low gene expression.

The log-rank p-value displayed on each plot indicates whether the observed survival difference between the two groups is statistically significant:

- A p-value ≤ 0.05 suggests that the gene's expression has a significant effect on survival outcomes.

Based on the results:

- BUB1B ($p = 0.26$), CDK1 ($p = 0.061$), MAD2L ($p = 0.26$), CDC20 ($p = 0.51$), TPX2 ($p = 0.11$), PBK ($p = 0.3$), and TOP2A ($p = 0.89$) did not show statistically significant differences.
- However, UBE2C, with a p-value of 0.029, was the only gene among the selected hub genes that showed a significant association with poorer survival.

Therefore, these findings suggest that only the key gene UBE2C had a significantly worse survival rate

The results are illustrated in **Figure 10**.

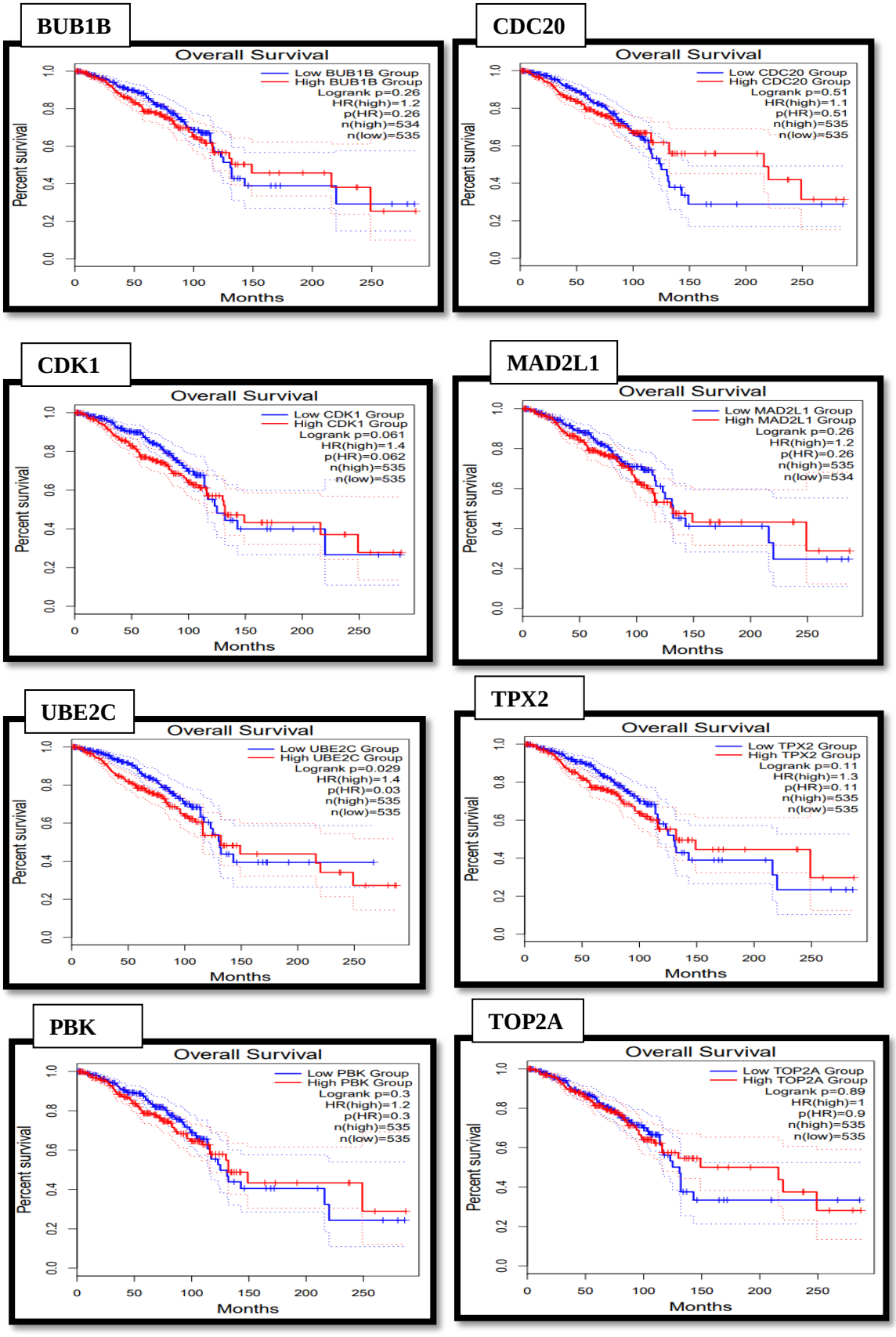


Figure 10: Analysis of overall survival of the eight hub genes by GPAIA2 database.

3.7. Validation of expression of hub genes

Then using the GEPIA2 dataset to analyze the RNA sequencing expression data between BC and control samples, each graph compares the expression of a gene between:

Tumor samples (T) shown in red and Normal samples (N) shown in gray

The red boxes are consistently higher than the gray ones, suggesting that these genes are upregulated in breast cancer tissues. The asterisk (*) above each plot usually indicates statistical significance (p value ≤ 0.05), meaning the difference in expression between tumor and normal tissues is significant. We found that the mRNA expression of the eight hub genes was higher in BC tissues than in normal breast tissues (**Figure 11**).

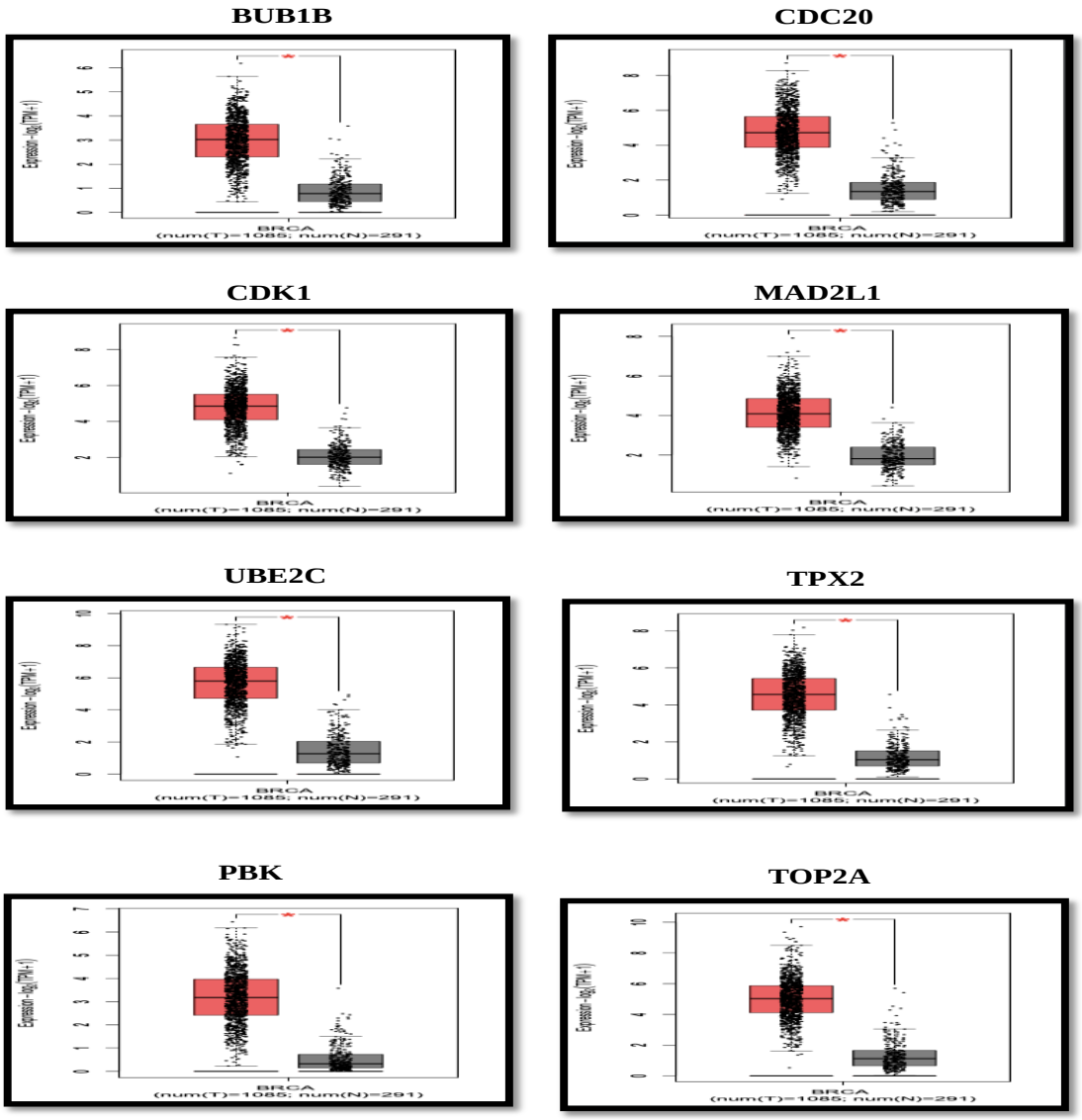


Figure 11: Expression profiles of hub genes in BC patients (GEPIA2). The red box represents BC samples; the gray box represents normal samples.

3.8. Network of transcription factors- miRNA coregulatory of key genes

Interactions between miRNAs, genes, and transcription factors (TFs) were collected using Network Analyst. The key gene UBE2C was analyzed to identify its associated miRNAs and TFs (Figure 12). UBE2C is regulated by 11 miRNAs and 07 TFs.

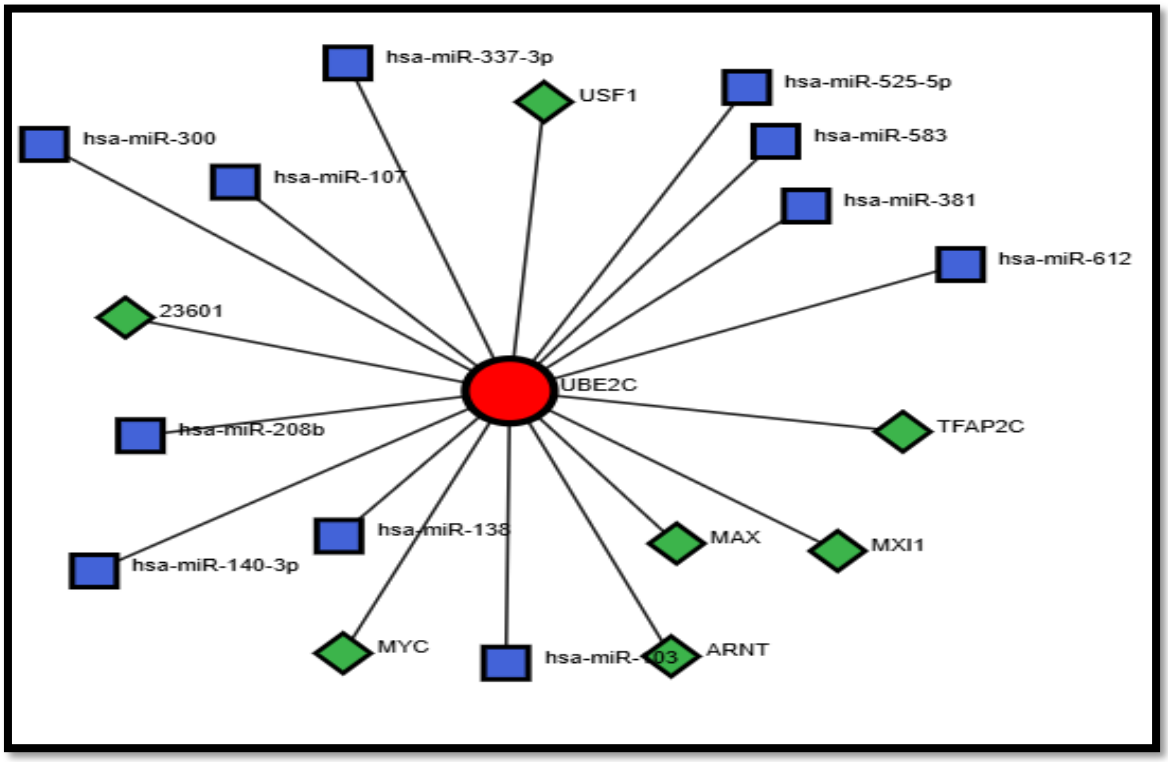


Figure 12: Network of Transcription factors- miRNA coregulatory and key gene. The red node represents the key gene (UBE2C), The blue nodes represent microRNAs (miRNAs) that interact with the key gene. The green nodes represent transcription factors.

4. Discussion

This study aims to use bioinformatics tools to analyze BC microarray data and perform comparisons to identify differently expressed genes (DEGs), we analyzed three GEO datasets (**GSE29431**, **GSE42568** and **GSE61304**) and discovered 663 overlapping differentially expressed genes (DEGs), which included 500 up-regulated genes and 163 down-regulated genes **[35].**

To better understand the function of these genes, the GO analysis on the common DEGs according to BP,CC and MF showed that the up-regulated genes were mainly enriched in *positive regulation of gene expression, positive regulation of phosphatidylinositol 3-kinase/protein kinase B signal transduction, angiogenesis, negative regulation of angiogenesis, response to glucose*, another level of complexity in the control of gene expression in BC is provided by microRNAs (miRNAs), small non-coding RNA molecules that bind to the 3' untranslated region (UTR) of targeted mRNAs and hence influence gene expression. Although miRNAs mostly act as negative regulators of gene expression, in particular situations, some miRNAs can also act as positive regulators. For instance, some miRNAs can prevent the translation of repressor proteins that would otherwise block their expression and therefore boost the expression of target genes. Furthermore, by indirectly affecting the expression of their target genes, miRNAs could control the activity of transcription factors. The interaction of miRNAs and transcription factors in breast cancer is a sophisticated and dynamic mechanism that helps to finely adjust gene expression patterns **[36].** In addition, the phosphatidylinositol 3-kinase (PI3K)/protein kinase B (AKT) is a signaling pathway involved in cell proliferation, survival, invasion, migration, apoptosis, glucose metabolism and DNA repair **[37]**. The PI3K pathway is often altered in BC due to mutations or gene amplifications affecting its key components particularly the catalytic subunits p110α (encoded by *PIK3CA*) and p110β (*PIK3CB*), as well as the regulatory subunit p85α (*PIK3R1*). Among these, *PIK3CA* is the most frequently mutated gene in various human cancers. It encodes the p110α subunit and is commonly amplified in cancers of the head and neck, cervix, stomach, lungs, and breast. Notably, the highest mutation rates of *PIK3CA* have been observed in prostate, breast, endometrial, and colon cancers **[38].**

The link between angiogenesis and cancer is attributed to the visionary pioneer Judah Folkman (1933–2008), who was the first to propose that tumor growth is directly dependent on the development of blood vessel networks. In the case of malignant tumors, it is now clear that they have a very limited ability to grow without the support of blood vessels. Therefore,

the formation of blood vessels is a crucial step to ensure a continuous supply of essential nutrients to the tumor **[39].** Studies indicate that tumors, especially BC, depend on angiogenesis, the formation of new blood vessels, for growth beyond a few centimeters. This process not only supplies essential nutrients but also facilitates tumor spread and metastasis. Tumors respond to low oxygen levels by activating angiogenesis, leading to the development of new blood vessels. This crucial process is also involved in normal functions like development and healing. Once activated, tumors enter a persistent state of angiogenesis, allowing them to enhance their blood supply and promote rapid growth **[40].**

On the other hand, the down-regulated genes were more enriched in *cell division, mitotic spindle assembly checkpoint signaling, mitotic sister chromatid segregation, chromosome segregation, mitotic cell cycle.* A series of distinct genetic alterations accumulates in a cell before it becomes malignant. These changes involve the activation of cellular oncogenes and the inactivation of tumor suppressor genes. Several of these various genetic changes can occur only during cell division **[41].** The mitotic spindle assembly checkpoint is like a careful supervisor during cell division. It makes absolutely sure that all the chromosomes are lined up correctly and firmly attached to the spindle fibers before giving the all clear; for the cell to split. If something is wrong, this checkpoint can hit the brakes on the cell cycle. This pause is really important because it stops the cell from dividing with messed-up chromosomes, which can lead to genomic instability, a major factor in the development of cancers, including BC **[42].** The SAC's command center relies on key proteins like Mad1, Mad2, BubR1, Bub3, and Mps1.

These proteins team up to form a complex that puts the brakes on the anaphase-promoting complex/cyclosome (APC/C), a vital enzyme that kicks off the chromosome separation process. When Mad2 encounters a chromosome that's not properly attached (an unattached kinetochore), it changes its shape and becomes an active blocker of APC/C. Another important player, BubR1, directly latches onto and inhibits APC/C, effectively preventing the premature pulling apart of sister chromatids **[43].** These findings are in line with previous reports showing that in BC, malignant cells often exhibit defects in the spindle assembly checkpoint during mitosis, leading to inaccurate segregation of sister chromatids. This missegregation results in chromosomal instability (CIN), a hallmark of many solid tumors, including BC **[44]**. BRCA1, a well-known gene associated with BC susceptibility, plays a critical role in maintaining proper mitotic regulation. Loss of BRCA1 function leads to premature inactivation of the spindle checkpoint, resulting in mis segregation of sister chromatids. This contributes to the development of cells with severe genomic instability,

promoting tumor progression **[45].**

Then, in this study also we performed a module analysis on the PPI network constructed the eight hub genes. Through the GEPIA2 database we obtain that the genes (BUB1B, CDC20, CDK1, MAD2L1, PBK, TOP2A, TPX2, UBE2C) were significantly differentially expressed between tumor tissues and normal tissues. The OS analysis showed that the expressions of the key genes (UBE2C) may indicate that the survival chances of BC patients are low. Mitotic Arrest Deficient 2 Like 1 (MAD2L1) is a crucial member of the MAD2 protein family. Located on chromosome 4 in humans, MAD2L1 is a key component of the mitotic checkpoint complex. This complex ensures proper cell division by monitoring the process. Disrupting MAD2L1's function in mammalian cells can interfere with this checkpoint, leading to errors in cell division. Research has shown that mutations in MAD2L1 can contribute to tumor formation by causing chromosomal instability and aneuploidy (an abnormal number of chromosomes). MAD2L1 has been linked to several types of cancer, including lung adenocarcinoma, colorectal cancer, cervical cancer, hepatocellular carcinoma (HCC), acute T-cell lymphoma, BC, and stomach cancer **[46].**

BUB1B encodes a kinase participating in the spindle checkpoint; this protein plays a critical role in the inhibition of the anaphase-promoting complex/cyclosome (APC/C) and functions in delaying the onset of anaphase and ensuring proper chromosome segregation. The oncogene role of BUB1B has been observed in cancers such as prostate cancer, glioblastoma and gastric cancer. BUB1B is overexpressed in BC, and the level of BUB1B mRNA is significantly correlated with intrachromosomal instability. In addition, BUB1B is preferentially expressed in high-grade breast cancer, and its expression level exhibits significant associations with long-term survival **[47].** BUB1B causes higher chromosomal instability in BC cells **[48].**

The TOP2A gene has been highlighted in eight studies as playing a significant prognostic role in BC patients. Its overexpression is associated with a poor prognosis, particularly in luminal or hormone receptor-positive BC. As a result, TOP2A could serve as both a prognostic and predictive biomarker in BC **[49].** These findings align with previous studies that have shown UBE2C to be a crucial factor in cancer progression and prognosis. For example, Chao-hua Mo and colleagues examined the prognostic value of UBE2C expression at both the transcriptomic level (in 1006 cases) and the protein level (in 209 breast cancer tissue samples). Their results revealed that high UBE2C expression is linked to poorer outcomes and more aggressive tumor characteristics in breast cancer **[50].**

PBK has been recognized as a kinase that regulates mitosis by phosphorylating GPSM2 (G-protein signaling modulator 2) in breast cancer cells. It also interacts with p53 to control the expression of cell cycle genes. Numerous studies have found that PBK is overexpressed in various cancers, including breast, prostate, colon, bladder, and lung cancer, and serves as a prognostic biomarker for poor outcomes. Additionally, previous research has shown that PBK expression is controlled by the transcription factors Myc and E2F1 **[51].** In the early stages of mitosis, CDC20 activates the anaphase-promoting complex (APC), leading to the formation of the E3 ubiquitin ligase complex APCCdc20. This complex functions to destroy key regulatory proteins of the cell cycle, thereby facilitating the process of mitosis. Additionally, CDC20 acts as a spindle checkpoint, ensuring the proper separation of chromosomes between daughter cells **[52].** When CDC20 transforms into an oncoprotein, the high expression of the oncogene CDC20 has been demonstrated in various types of human malignancies **[53].** Indeed, previous studies have shown that its expression significantly increases in various types of cancerous tumors, including breast, pancreatic, prostate, colon, bladder, and lung cancers. This overexpression of CDC20, whether at the gene or protein level, is associated with a poor prognosis for these tumors **[54].**

Cyclin-dependent kinase 1 (CDK1) (CDK1 gene) is a key enzyme that regulates the cell division cycle. It plays a vital role in the transition from the G2 phase to the M phase, where the cell prepares to divide, and also regulates the progression of the G1 phase and the transition to the S phase, where the cell grows and replicates its DNA. When there is a disruption in CDK1 activity, it can lead to uncontrolled and excessive cell growth, which is a hallmark of cancerous tumors **[55].** A previous study showed that dysregulation of CDK enzymes, particularly CDK1, is associated with decreased survival rates in breast cancer patients. This study, along with our research findings, confirms that targeting these enzymes, especially CDK1, represents a promising strategy for breast cancer treatment, particularly in cases where the cancer has spread to other parts of the body (metastatic tumors) **[56].**

TPX2 gene, also known as XKIP2 targeting protein, plays a vital role in the regulation of microtubules associated with the cell's kinetochore. Recent studies have shown a close association between the TPX2 gene and the development of malignant tumors, including breast cancer **[57].** Laboratory experiments demonstrated that inhibiting the TPX2 gene with small hairpin RNA (shRNA) hinders breast cancer cell growth and proliferation, while promoting apoptosis. This inhibition resulted in decreased PI3K expression, reduced AKT phosphorylation, and increased p53 and p21 expression, indicating that TPX2 could be a viable target for anticancer therapy **[58].**

Then, the interaction network highlights the role of UBE2C as a key gene regulated by 11 miRNAs and 07 transcription factors (TFs). The red node represents UBE2C, while the blue nodes correspond to miRNAs interacting with this gene. The green nodes represent transcription factors (TFs) that also regulate UBE2C expression. MicroRNAs (miRNAs) are small RNA molecules that play a role in regulating biological processes. They are processed from stem-loop regions in longer RNA transcripts. MicroRNAs are categorized into families based on their targeting characteristics, which are mainly determined by the sequence of their extended seed region. In this network, miRNAs such as hsa-miR-583, hsa-miR-337-3p, and hsa-miR-140-3p interact with UBE2C **[59].**

In tumor samples, UBE2C was highly expressed. When UBE2C expression was reduced, tumor cell proliferation and migration were inhibited, and the activity of the PI3K/AKT signaling pathway was decreased. Additionally, reducing UBE2C expression slowed tumor growth in animal models. The expression of UBE2C was lowered by miR-140-3p, as it directly targets UBE2C. UBE2C is overexpressed, which promotes tumor growth and migration while enhancing the activity of the PI3K/AKT pathway. However, when miR-140-3p expression was forced, these effects were partially reversed **[60].**

Transcription factors (TFs) play a key role in regulating various cellular processes. These processes include internal functions, like development and differentiation, as well as external ones, such as responding to signals from the environment **[61].** In this network, transcription factors such as MYC, MAX, USF1 and TFAP2C regulate the UBE2C expression. TFAP2C has been implicated in the progression of breast cancer by controlling genes that are involved in regulating the cell cycle **[62].** MYC is a well-known oncogene that drives cell growth and metabolism, and its interaction with UBE2C may play a key role in the progression of cancer **[63].** The transcription factor MYC (MYC proto-oncogene) plays a key role as a signaling hub in various cellular processes that support the growth of many types of cancer. It regulates the expression of both protein-coding and non-coding RNAs, which in turn control essential metabolic pathways, cell death, cell proliferation, differentiation, stress responses, and drug resistance mechanisms. Activation of MYC has been commonly observed in the progression of breast cancer **[64].**

CONCLUSION

This study used bioinformatics tools to analyze datasets of patients with breast cancer to assess the expression levels of functional genes, key pathways, common key genes, and, subsequently, provide new insights into the prediction of the prognosis and the discovery of targeted anti-breast cancer therapies. As such, this study has allowed us to gain access to the following results:

• Identification of 663 common genes to the three expression profiles of patients, including 500 up-regulated genes and 163 down-regulated genes.

• Different biological processes are associated with breast cancer, including positive regulation of gene expression, positive regulation of phosphatidylinositol 3kinase/protein kinase B signal transduction, angiogenesis, cell division, and mitotic spindle assembly checkpoint signaling, mitotic sister chromatid segregation, chromosome segregation, and mitotic cell cycle.

• Eight genes were identified as hub genes including: BUB1B, CDC20, CDK1, MAD2L1, UBE2C, TPX2, PBK, TOP2A.

• The key gene UBE2C was identified as being significantly associated with poor survival of patients with breast cancer patients, and was confirmed to be expressed in breast cancer tissues. These findings offer key insights into breast cancer progression and the development of new anticancer therapies.

BIBLIOGRAPHIC REFERENCES

1. Gerami R, Sadeghi Joni S, Akhondi N, et al. A literature review on the imaging methods for breast cancer. J Cancer Imaging. 2022;14(3):171–176.
2. Prat A, Pineda E, Adamo B, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast Edinb Scotl. 2015;24 Suppl 2:S26–35.
3. Malhotra GK, Zhao X, Band H, et al. Histological, molecular and functional subtypes of breast cancers. Cancer Biol Ther. 2010;10:955–60.
4. Makki J. Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. Clin Med Insights Pathol. 2015;8:23–31
5. Giuliano AE, Connolly JL, Edge SB, et al. Breast Cancer—Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA Cancer J Clin. 2017;67:290-303.
6. Houghton SC, Hankinson SE. Cancer Progress and Priorities: Breast Cancer. Cancer Epidemiol Biomarkers Prev. 2021;30(5):822–844.
7. Jin H, Huang X, Shao K, et al. Integrated bioinformatics analysis to identify 15 hub genes in breast cancer. Oncol Lett. 2019 May 30;18(5):1023-1034.
8. Wang N, Zhang H, Li D, et al. Identification of novel biomarkers in breast cancer via integrated bioinformatics analysis and experimental validation. Cancer Biology & Therapy. 2021;22(1):12431–12446.
9. Margaret, Ay Ly. Identification of potential breast cancer stem cell biomarkers in the secretome using a network interaction approach analysis. Asian Pac J Cancer Prev. 2024;25(5):1803.
10. Tong CWS, Wu M, Cho WCS, et al. Recent Advances in the Treatment of Breast Cancer. Front Oncol. 2018;8:227.
11. Gurmessa DK, Jimma W. Explainable machine learning for breast cancer diagnosis from mammography and ultrasound images: a systematic review. BMJ Health Care Inform. 2024;31(1).
12. Al-Karawi D, Al-Zaidi S, Helael KA, et al. A Review of Artificial Intelligence in Breast Imaging. Tomography. 2024;10(5):705–726.
13. Mehrotra R, Yadav K. Breast cancer in India: Present scenario and the challenges ahead. World J Clin Oncol. 2022;13(3):209–18.
14. Łukasiewicz S, Czeczelewski M, Forma A, et al. Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review. Cancers (Basel). 2021;13(17):4287.

15. Yedjou CG, Sims JN, Miele L, et al. Health and racial disparity in breast cancer. Adv Exp Med Biol. 2019;1152:31-49.

16. Nicolis O, De Los Angeles D, Taramasco C. A contemporary review of breast cancer risk factors and the role of artificial intelligence. Front Oncol. 2024;14:1356014.

17. Satpathi S, Gaurkar SS, Potdukhe A, et al. Unveiling the Role of Hormonal Imbalance in Breast Cancer Development: A Comprehensive Review. Cureus. 2023;15(7).

18. Terry PD, Rohan TE. Cigarette smoking and the risk of breast cancer in women: a review of the literature. Cancer Epidemiol Biomarkers Prev. 2002;11(10):953-971.

19. Ajabnoor GMA. The molecular and genetic interactions between obesity and breast cancer risk. Medicina (Kaunas). 2023;59(7):1338.

20. Liu Y, Nguyen N, Colditz GA. Links between alcohol consumption and breast cancer: a look at the evidence. Womens Health (Lond). 2015;11(1):65-77.

21. Kulkarni S, Lewis K, Arp Adams S, et al. A comprehensive analysis of how environmental risks of breast cancer are portrayed on the internet. Am J Health Educ. 2018;49(4):222-233.

22. Iacob R, Iacob ER, Stoicescu ER, et al. Evaluating the role of breast ultrasound in early detection of breast cancer in low- and middle-income countries: A comprehensive narrative review. Bioengineering. 2024;11(3):262.

23. Orrantina-Borunda E, Anchondo-Nuñez P, Acuña-Aguilar LE, et al. Subtypes of breast cancer. Breast Cancer. 2022 Aug 6.

24. Aysola K, Desai A, Welch C, et al. Triple negative breast cancer - an overview. Hereditary Genet. 2013;2013(Suppl 2):001.

25. Marimon RP. Development of an immunotoxin derived from Pseudomonas aeruginosa for the treatment of triple-negative breast cancer [Master's thesis]. Lisboa: Universidade de Lisboa; 2023.

26. Kirkby M, Popatia AM, Lavoie JR, et al. The Potential of Hormonal Therapies for Treatment of Triple-Negative Breast Cancer. Cancers. 2023;15(19):4702.

27. Bhat SA, Kumar V, Dhanjal DS, et al. Biogenic nanoparticles: pioneering a new era in breast cancer therapeutics—a comprehensive review. Nanoscale Res Lett. 2024;19(1):121.

28. Ashkarran AA, Lin Z, Rana J, et al. Impact of Nanomedicine in Women's Metastatic Breast Cancer. Small. 2023;20(41).

29. Ding R, Xiao Y, Mo M, et al. Breast cancer screening and early diagnosis in Chinese women. J Cancer. 2022;19(4):450–467.

30. Fuller MS, Lee CI, Elmore JG. Breast cancer screening: An evidence-based update. Med Clin North Am. 2015;99(3):451–468.

31. Dietze E, Jones V, Seewaldt V. Breast Self-Examination – the case for a second look. Curr Breast Cancer Rep. 2020;12(2):118–124.

32. Mounce R. Breast cancer treatment by stage. Cooperhealth. Camden (NJ): Cooper University Health Care; 2015 Aug 31.

33. Gentry SS. Breast cancer treatments. Conquer: the journey informed. 2020 Oct [cited 2021 Jan 24].

34. Wang Y, Minden A. Current Molecular Combination Therapies Used for the Treatment of Breast Cancer. Int J Mol Sci. 2022;23(19):11046.

35. Chen G, Yu M, Cao J, et al. Identification of candidate biomarkers correlated with poor prognosis of breast cancer based on bioinformatics analysis. Bioengineered. 2021 Aug 12;12(1):5149-5161.

36. Muñoz JP, Pérez-Moreno P, Pérez Y, et al. The Role of MicroRNAs in Breast Cancer and the Challenges of Their Clinical Application. Diagnostics1 (Basel). 2023 Sep 28;13(19):3072.

37. Miricescu D, Totan A, Stanescu-Spinu II, et al. PI3K/AKT/mTOR signaling pathway in breast cancer: From molecular landscape to clinical aspects. Int J Mol Sci. 2020 Dec 26;22(1):173.

38. Paplomata E, O'Regan R. The PI3K/AKT/mTOR pathway in breast cancer: targets, trials and biomarkers. Ther Adv Med Oncol. 2014 Jul;6(4):154–66.

39. Longatto Filho A, Lopes JM, Schmitt FC. Angiogenesis and breast cancer. J Oncol. 2010 Oct 7;2010:576384.

40. Gamde MS, Ogenyi OD. Angiogenesis in breast cancer: a review. Asian Pac J Cancer Biol. 2024;9(1):97-103.

41. Preston-Martin S, Pike MC, Ross RK, et al. Increased cell division as a cause of human cancer. Cancer Res. 1990 Dec 1;50(23):7415–7421.

42. Gorbunova V, Gordenin DA. The spindle checkpoint and chromosome segregation in meiosis. The FEBS journal. 2015 Jul;282(13):2471-87.

43. Bharadwaj R, Yu H. The spindle checkpoint, aneuploidy, and cancer. Oncogene. 2004 Mar 15;23(11):2016-27.

44. Bakhoum SF, Genovese G, Compton DA. Deviant kinetochore microtubule dynamics underlie chromosomal instability in human cells. Nat Cell Biol. 2009;11(7):856–63.

45. Chabalier C, Lamare C, Racca C, et al. BRCA1 downregulation leads to premature inactivation of spindle checkpoint and chromosome missegregation in breast cancer cells. Cell Cycle. 2006;5(9):1001–7.

46. Liu S, Miao M, Kang L. Upregulation of MAD2L1 mediated by ncRNA axis is associated with poor prognosis and tumor immune infiltration in hepatocellular carcinoma: A review. Medicine (Baltimore). 2023 Jan 13;102(2):e32625.

47. Cai Y, Mei J, Tay G, et al. Identification of five hub genes as monitoring biomarkers for breast cancer metastasis in silico. Hereditas. 2019 Dec;156(1).

48. Fu Y, Zhou QZ, Zhang XL, et al. Identification of hub genes using coexpression network analysis in breast cancer as a tool to predict different stages. Med Sci Monit. 2019 Nov 23;25:8873–8890.

49. Alam MS, Sultana A, Mollah MNH, et al. Identification of hub of the hub-genes from different individual studies for early diagnosis, prognosis, and therapies of breast cancer. Bioinform Biol Insights. 2024 Sep 4; 18:11779322241272386.

50. Kariri Y, Toss MS, Alsaleem M, et al. Ubiquitin-conjugating enzyme 2C (UBE2C) is a poor prognostic biomarker in invasive breast cancer. Breast Cancer Res Treat. 2022;192:529-39.

51. Dou X, Wei J, Sun A, et al. PBK/TOPK mediates geranylgeranylation signaling for breast cancer cell proliferation. Cancer Cell Int. 2015;15:27.

52. He W, Meng J. CDC20: A novel therapeutic target in cancer. Am J Transl Res. 2023;15(2):678.

53. Zhou Z, He M, Shah AA, et al. Insights into APC/C: From cellular function to diseases and therapeutics. Cell Div. 2016;11:9.

54. Messeha SS. Prognostic and therapeutic implications of cell division cycle 20 homolog in breast cancer. Cancers. 2024;16(14):2546.

55. Xing Z, Wang X, Liu J, et al. Expression and prognostic value of CDK1, CCNA2, and CCNB1 gene clusters in human breast cancer. J Int Med Res. 2021;49(4):0300060520980647.

56. Mehraj U, Sofi S, Alshehri B, et al. Expression pattern and prognostic significance of CDKs in breast cancer: An integrated bioinformatic study. Cancer Biomarkers. 2022;34:505-19.

57. Huang C, Han Z, Wu D. Effects of TPX2 gene on radiotherapy sensitization in breast cancer stem cells. Oncol Lett. 2017;14(2):1531-5.

58. Chen M, Zhang H, Zhang G, et al. Targeting TPX2 suppresses proliferation and

promotes apoptosis via repression of the PI3k/AKT/P21 signaling pathway and activation of p53 pathway in breast cancer. Biochem Biophys Res Commun. 2018;507(1-4):74-82.

59. Bartel DP. Metazoan microRNAs. Cell. 2018 Mar 22;173(1):20-51.

60. Huang W, Zhou Y, Ren J, et al. UBE2C, targeted by miR-140-3p, promotes the progression of osteosarcoma via PI3K/AKT signaling pathway. J Recept Signal Transduct Res. 2024 Jun;44(3):107-114.

61. Weidemüller P, Kholmatov M, Petsalaki E, et al. Transcription factors: Bridge between cell signaling and gene regulation. Proteomics. 2021 Dec;21(23-24)

62. McPherson LA, Rothrock CR, Barczak AJ, et al. TFAP2C regulates proliferation and luminal identity in breast cancer. Oncogene. 2020;39(7):1330-1344.

63. Dang CV. MYC on the path to cancer. Cell. 2016;149(1):22-35.

64. Fallah Y, Brundage J, Allegakoen P, et al. MYC-Driven Pathways in Breast Cancer Subtypes. Biomolecules. 2017 Jul 11;7(3):53.

- **Cytoscape:** is an open-source bioinformatics software platform that is used for visualizing molecular interaction networks and biological pathways: It provides a visual representation of how molecules interact with each other and the pathways they are involved in.
- **CytoHubba:** is a tool used in bioinformatics to identify hub genes or hub nodes within protein-protein interaction (PPI) networks. It is a plugin for the popular bioinformatics platform Cytoscape, which is used for visualizing molecular interaction networks.
- **DAVID (Database for Annotation, Visualization and Integrated Discovery)**: is a bioinformatics resource that provides tools for understanding the biological meaning behind large lists of genes or proteins.
- **FC (Fold Change):** is a way to describe how much a quantity changes between two measurements. It is often used in biology to compare things like gene expression levels under different conditions.
- **GEO (Gene Expression Omnibus):** It is a large, public database of gene expression data housed at the National Center for Biotechnology Information (NCBI) in the United States.
- **GO (Gene Ontology):** is a major bioinformatics initiative that aims to standardize the representation of gene and protein functions across different databases and organisms. It provides a structured, controlled vocabulary of terms to describe: Molecular Function (MF), Biological Process (BP), Cellular Component (CC).
- **KEGG (Kyoto Encyclopedia of Genes and Genomes):** is a bioinformatics resource that provides a collection of databases and tools for understanding the functions of biological systems, especially at the molecular level. It focuses on pathways, genomes, diseases, and drug interactions, and is widely used in genomics, systems biology, and bioinformatics research.
- **NCBI (National Center for Biotechnology Information)**: is a branch of the U.S. National Library of Medicine (NLM), part of the National Institutes of Health (NIH). It provides a wide array of databases, tools, and resources for biotechnology and biomedicine research, with a major focus on molecular biology, genetics, and health-related data.
- **OS (Overall Survival):** is a term commonly used in medical research and clinical trials, especially in oncology and other fields involving serious diseases, to measure the length

of time from either the diagnosis of a disease or the start of treatment until death from any cause. It's a key outcome used to assess the effectiveness of a treatment or intervention.

- **The p-value (probability value)**: is a fundamental concept in statistical hypothesis testing, and it helps determine the significance of your results. It essentially measures the strength of evidence against the null hypothesis.
- **Venn diagram:** is a graphical representation used to show the relationships between different sets or groups. It consists of overlapping circles (or other shapes) that visually demonstrate how the sets intersect or how they are distinct from one another. Venn diagrams are often used in mathematics, logic, statistics, and even in everyday scenarios to illustrate logical relationships and set operations.
- **Volcano plot:** is a type of graphical representation used in genomics, transcriptomics, and proteomics research to visualize the results of differential expression analyses, such as gene expression studies. It combines two important statistical measures—magnitude of change (often represented by fold-change) and statistical significance (often represented by p-value)—into a single plot.
- **GEPIA2 (Gene Expression Profiling Interactive Analysis, version 2):** A web-based resource that provides interactive analysis of gene expression data derived from large-scale cancer and normal tissue datasets, such as TCGA and GTEx. GEPIA2 enables researchers to perform differential gene expression analysis, survival analysis, correlation analysis, and download gene expression data, facilitating the identification of potential biomarkers and the exploration of gene functions in cancer biology.
- **Network Analyst:** Network Analyst is a web-based platform designed for the comprehensive analysis and visualization of biological networks. It enables researchers to explore complex interactions between genes, proteins, microRNAs, and other biomolecule

Table: List of common DEGs expressed in the three datasets.

Gene symbol (up-regulated genes, 500 genes)
CLMP SAMD4A IGF1 NAALAD2 PELI2 PPP1R14A LINC01279 EPB41L4B GNAL RGN LOC101930168///LOC100509620///AQP7 ERG LEP HOXA10-HOXA9///MIR196B///HOXA9 VGLL3 LOC100653057///CES1 LTBP4 GHR NAT8L ISM1 MAOA MRAP EIF1 LOC101930400///AKR1C2 GDF10 BCHE MEIS2 KL FAM162B ID1 GPD1 NIPSNAP3B TF THRB NPR1 SPTBN1 SPX IRS1 TSPAN7 MRGPRF PENK SMAD9 PCDH18 QKI ACADL SAA2///SAA1 SMIM10L2B///SMIM10L2A FGF1 PIR-FIGF///FIGF SPG20 ZBTB20 HLF SPRY2 MYH1 ADRA1A FHL1 RBPMS SORBS2 KLF4 SH3D19 PLSCR4 CEBPD ADRB1 ALDH6A1 ITSN1 PFKFB1 ZFP36 NDRG2 SVEP1 SEMA3G ATP8B4 GLDN PDGFD PREX2 HEPN1///HEPACAM FGF14-AS2 SCN4B KANK3 SYNE3///LINC00341 ANKDD1A PLAC9 SH3BGRL2 PTGS2 C1QTNF7 SPIDR NEGR1 EHD2 HOXA5 ADH1C RNF150 RUNX1T1 SGK2 MIR6883///PER1 S100B BMP2 LGALS12 SEL1L2 IRS2 DENND2A SIK2 PTPN14 KCNAB1 CLIC5 RBP7 TRHDE DMGDH PROS1 ENPP2 MME LOC101930114 LOC284825 SOCS3 ZC3H12C KLF9 FIGN GYG2 AMOTL2 PDK4 FILIP1 MAGI2-AS3 ST6GALNAC3 LOC101929500///CRIM1 RERGL ANGPTL1 SLC22A3 FAM13A C14orf180 EFEMP1 LINC01140 MOCS1 PLXNA4 SYN2 ADGRF5 HBA2///HBA1 ARHGAP20 NR3C2 EMCN LOC101926960 OGFRL1 GPIHBP1 FOXP2 NRN1 SHE OGN GIPC2 ANGPTL4 TLN2 INS-IGF2///IGF2 CCDC3 NTRK2 CA4 AASS TACC1 SPRY1 TCEAL7 LVRN ABCA6 CXCL2 GNAI1 EGFR MYEOV LYVE1 IGFBP6 ARHGEF28 TNXB///TNXA LEPROT///LEPR MIR99AHG TENM3 PALMD AKAP12 MEOX2 HBG2///HBG1 APOD APOLD1 LRRC70///IPO11 WNT11 ROBO4 KLB PID1 EBF2 ZNF423 SART1 SEMA6D SYNE1 DMRT2 PDZD2 KLHL31 LHFP HBB GNG11 GPAM RHOJ SRPX CLIP4 TFPI ABCA9 CA3 TWIST2 SLC35G2 MIR548F5///MAB21L1 MYZAP TSLP PRRG3 PLPP3 LOXL4 TGFBR3 MYL9 LRRN4CL AKR1C1 LPL LINC00968 FAXDC2 GPC3 ACACB MAF DMD DLC1 PPP1R1A LINC00312 NOVA1 CCBE1 BOC PGM5 FXYD1 CAV2 FOSB TPPP GSTM5 SOBP ITGA7 RSPO3 KCNIP2 DCN CKMT2 TIMP4 SLIT2 C2orf88 HSPB7 PLP1 TMTC1 PCOLCE2 RGS7BP SYNM PHLDB2 CCDC178 SLC4A4 C2orf40 AOX1 GULP1 CFD CH25H RFX2 PDE3B MOB3B WLS COBLL1 EIF4EBP2 SOX7 ADAMTS18 MICU3 CHRDL1 LAMA2 ANXA1 NID1 ITIH5 TMEM110-MUSTN1///MUSTN1///TMEM110 MMRN2 SYNPO2 PLIN4 TUSC5 CCDC50 BTNL9 MYOM1 CLDN11 MAML2 AADAC MEOX1 TRDN FAM150B SLC16A7 ANKRD35 CACHD1 DNASE1L3 CRYAB CD36 RNF180 DCLK1 ADAMTS9 GRK3 LIPE FGD4 SGCB APOB CFH PTPRB ITGA1 ARID5B PPARG PQLC2L PPP2R1B CORO2B MT1M HCAR3 MID1 LOC401317///CREB5 TSHZ2 ZNF366 EGR1 PLAGL1 NR3C1 EMP1 PGM5-AS1 FAM149A PKD1L2 ABCA8 SLC19A3 LEPROT AOC3 BHMT2 TMEM132C TMEM47 CHL1 ADGRL4 HSPB2 GSN ANK2 PLA2G4A LMOD1 ZBTB16 ASPA CCL15-CCL14///CCL14 EDNRB SLIT3 MYCT1 SAA2-SAA4///SAA4 TNMD LOC105379426 EBF3 EMX2OS PCK1 CIDEA P2RY12 RGCC LRRC34 VWF ACVR1C CIDEC TM4SF18 CRIM1 EBF1 CD300LG ABI3BP RUNDC3B CRYBG3 CYYR1 TNS1 ADRB2 C8orf88 G0S2 PLEKHH2 SLC7A10 NTRK3 LOC100506558///MATN2 LOC101930400///AKR1C2///AKR1C1 SGCG WASF3 ARRDC4 HOXA7 EGR3 TBX15 BMPER TXNIP ZNF521 PDE1A IL6 TLR3 CPED1 ANGPT1 EEF1A1 TEF TMEM100 ARHGEF40 CREB5 SAA2-SAA4///SAA2///SAA1 ACSM5 PLIN1 COPG2IT1 DPT MMP28 FAM13C MFAP4 ECM2 LATS2 PDE2A LHCGR SCARA5 CLDN5 ATP1A2 DUSP6 IGSF10 RBP4 CDO1 SDPR ID4 SOX5 EXOSC7///CLEC3B ADAMTS1 NFIB ACKR1 SORBS1 LIFR STXBP6 DEFB132 S1PR1 TTC28 MTURN LOC654342///LOC645166 NECTIN3 FOS FGF2 CLU PAMR1 CXCL12 PTH2R LOC101929583///LOC101928195///LOC100996643///LOC100133920///LOC286297 LDB2 AVPR1A C16orf89 PRKD1 CAV1 ALDH1A1 JAM2 TNN PPM1L ANKRD29 GPR146 WDFY3-AS2 GLYAT LIPE-AS1 BLOC1S1-RDH5///RDH5 INMT ZEB2 ITM2A COX7A1 S100A10 PDE11A KLHL29 GPRASP1 CDKN1C AQP1 LOC101928635///ALDH1A2 APCDD1 PCDH9 CEP126 FMO2 LMO3 SOCS2 RBMS3 CSN1S1 TMOD1 FGFBP2 TGFBR2 IL33 MIR6756///MCAM ADH1B TCF7L2 MYRIP MIR22///MIR22HG FABP4 KCNB1 FAM107A CCDC85A GPX3 LOC102723493 WDR86 ABLIM3 LOC101929726 AKR1C3 DDR2 FREM1 F3 CASQ2 GGTA1P MAMDC2 EOGT CEP112 ADAMTS5 TRHDE-AS1 SFRP1 GALNT16 DAAM1 COL6A6 TNS2 ZAK C6 LOC100506718///FLRT2 IL11RA GPLD1 ADIPOQ
Gene symbol (down-regulated genes, 163 genes)
HIST1H3F///HIST1H3B///HIST1H3H///HIST1H3J///HIST1H3G///HIST1H3I///HIST1H3E///HIST1H3C///HIST1H3D///HIST1H3A TESMIN TPX2 COL4A5 CCNB1 RAB3IP CAPS GINS1 KPNA2 COL1A1 ANLN BIRC5 MIR6787///SLC16A3 FOXM1 CDK1 ABHD2 RGS4 CENPU APOBEC3B SLC9A3R1 AURKA FN1 MAD2L1 KIF4A CYB561 KRT8 SPP1 SQLE FAM72A///FAM72D///FAM72B///FAM72C MELK HN1 ZWINT CDH11 NDC80 CCNA2 E2F7 BGN NUF2 MMP1 PTTG1 CDCA5 UBE2T CKS2 ECT2 KIF23 DEPDC1 MMP11 CCDC167 PLPP4 IQGAP3 MTFR2 LMNB1 CRABP2 SPAG5 POSTN CLDN3 CCNB2 PRC1 CDT1 CEP55 COMP CCNE2 RALGPS2 MCM4 WISP1 SBK1 RUNX2 DLGAP5 RAD51AP1 ESRP1 CXCL10 VCAN MKI67 SULF1 STRA13 HIST1H2BD CERS6 COL11A1 HIST2H2AA4///HIST2H2AA3 KIAA0101 NEK2 GINS2 COL12A1 SLC44A4 ESPL1 E2F8 KIF26B S100P MMP13 HMGB3 SAMD12 KIF11 KIF18B KLHDC7B IFI30///PIK3R2 KIFC1 CXCL11 SYNGR2 LEF1 BICDL1 S100A14 SLAMF8 KIF2C GPRC5A LRRC15 CDC20 MIAT PLEKHF2 PBK PRR11 TRIP13 ISG15 NUP210 BCAS4 TPD52 TK1 ASPM INHBA CDCA3 IFI6 ATAD2 KNL1 STARD10 BRIP1 UBE2C SDC1 RRM2 SLC35F6///CENPA TOP2A MICAL2 GJB2 RBM47 HELLS FANCI MSI2 CNTNAP2 KIF15 HIST1H2BC///HIST1H2BI///HIST1H2BE///HIST1H2BF///HIST1H2BG BUB1B HJURP DTL FAM83D HMMR AP1M2 CDCP1 HS6ST3 KIF20A ORC6 TLCD1 COL10A1 TDO2 COL5A1 TRIM59 CTHRC1 UHRF1 TTK CDKN3 NCAPG RMI2 CENPF NUSAP1 CDCA8