LEBANESE AMERICAN UNIVERSITY

A Multi-Omic Characterization of Multiple Oncogenic Interactions and its Therapeutics in Invasive Breast Carcinoma applying Bioinformatics Tools

By

Sarkis Gerges Bou Daher

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Science in Biological Sciences

School of Arts and Sciences
December 2022

LAU
الجامعة اللبنانية الأميركية
Lebanese American University

THESIS APPROVAL FORM

Student Name: Sarkis Gerges Bou Daher I.D. #: 201905728

Thesis Title: A Multi-Omic Characterization of Multiple Oncogenic Interactions and its Therapeutics in Invasive Breast Carcinoma applying Bioinformatics Tools

Program: Master of Science in Biological Sciences

Department: Natural Sciences

School: Arts and Sciences

The undersigned certify that they have examined the final electronic copy of this thesis and approved it in Partial Fulfillment of the requirements for the degree of:

Master of Science in the major of Biological Sciences

Thesis Advisor's Name: Dr. Costantine Daher Co-Advisor: Dr. Nancy Fayad

Signature: [redacted] [redacted] Date: 19 / 12 / 2022
Day Month Year

Committee Member's Name: Dr. Roy Khalaf

Signature: [redacted] Date: 19 / 12 / 2022
Day Month Year

Committee Member's Name: Dr. Jose Noel Ibrahim

Signature: [redacted] Date: 19 / 12 / 2022
Day Month Year

LAU
الجامعة اللبنانية الأميركية
Lebanese American University

THESIS COPYRIGHT RELEASE FORM

LEBANESE AMERICAN UNIVERSITY NON-EXCLUSIVE DISTRIBUTION LICENSE

Name: Sarkis Gerges Bou Daher

Signature [redacted]

Date: 29 / 11 / 2022
Day Month Year

LAU
الجامعة اللبنانية الأميركية
Lebanese American University

PLAGIARISM POLICY COMPLIANCE STATEMENT

I certify that:

1. I have read and understood LAU's Plagiarism Policy.
2. I understand that failure to comply with this Policy can lead to academic and disciplinary actions against me.
3. This work is substantially my own, and to the extent that any part of this work is not my own I have indicated that by acknowledging its sources.

Name: Sarkis Gerges Bou Daher

Signature: [redacted]

Date: 29 / 11 / 2022
Day Month Year

ACKNOWLEDGMENT

First and foremost, my gratitude is never enough to the Almighty God for his presence and guidance in my life. Without him nothing would be possible.

I want to express my deepest gratitude and sincere acknowledgement to Professor Costantine Daher who was always by my side. His presence is an example of God's blessings in life. He is not only a prominent supervisor and mentor, but also a great person that leaves a great impact in the life of everyone who encounters him. Dr. Costantine Daher guided and supported me through my journey at LAU. He was the guiding father when times were stressful. He was the Professor that guided and supported me through my academic journey. Not enough words can express my gratitude for your presence in my life. Thank you, very much, for the support. You will always be a great idol that generations will follow.

I want to thank my co-supervisor, Dr. Nancy Fayad. The soldier behind the details. This bioinformatics analysis wouldn't have been possible without your extensive expertise in the field. Thank you for dedicating time to make this mission possible. I would like to thank my committee members, Dr. Roy Khalaf and Dr. Jose Noel Ibrahim for taking time and effort to participate in my thesis and serve as committee members.

I am grateful to my parents for their support, love, prayers, and sacrifices to educate and equip me well for my future.

A Multi-Omic Characterization of Multiple Oncogenic Interactions and its Therapeutics in Invasive Breast Carcinoma applying Bioinformatics Tools

Sarkis Gerges Bou Daher

ABSTRACT

Invasive Breast Carcinoma is a complex heterogeneous disease in terms of diagnosis, clinical course, and pathology categorization. The World Health Organization (WHO) categorization does include more than a dozen variants, that are less prevalent, but are nonetheless extremely well described. In order to administer the appropriate therapy for each tumor, and to move from large, randomized research to a targeted one, it is crucial to understand the cancer type. This has now been made easier thanks to bioinformatics, the massive amounts of data produced, and the availability of tools to evaluate, understand, and then use this data in cancer therapy approaches. In this study, 12 genes associated with invasive breast carcinoma were mined in the cBioPortal for Cancer Genomics platform. Data was analyzed via bioinformatic tools such as Cytoscape, and MutationTaster. Further analysis regarding the oncogenic status of detected mutations allowed us to predict drugs that can be used in targeted treatment.*TP53* was found to have the highest alteration percentage followed by *ERBB2* and *CDH1*. Most of the gene combinations were co-occurrent with couple of mutually exclusive ones. The overall survival rate for patients harboring our 12 studied genes was lower than the unaltered group. Our study seeks to provide novel insights into invasive breast carcinoma's mutational signature, which would help improve therapeutic approaches, chemoresistance, autophagy, as well as the pathways disrupted in breast cancer.

Keywords: Invasive Breast Carcinoma, Bioinformatics, CBioPortal, MutationTaster, Cytoscape, SNPs.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Figure 5 - The overall survival of the unaltered group in comparison to the groups of the query altered genes. 35

LIST OF ABBREVIATIONS

SNPs………………………………………………………… Single Nucleotide Polymorphisms

TCGA…………………………………………………………….... The Cancer Genome Atlas

HER…………………………………………………………… Human Epidermal Growth Factor

COSMIC…………………………………………….... Catalogue of Somatic Mutations in Cancer

CHAPTER ONE

LITERATURE REVIEW

1.1 Cancer:

1.1.1 Understanding cancer:

Cancer is at the top of morbidity and mortality causes in the Western world. It is a highly serious public health issue, and a major cause of death, second only to cardiovascular disease (Trichopoulos *et al.*,1998). Cancer includes the uncontrolled division of cells in the body and refers to a collection of over 100 diseases that develop over time (Understanding Cancer, 2007). Despite the fact that cancer can spread to almost any tissue in the body and that each type of cancer has its own unique set of traits, the fundamental mechanisms that give rise to cancer are frequently the same (Baron-Faust,1995). Uncontrolled cell growth and the development of metastatic qualities are characteristics of cancer. Oncogene activation and/or tumor suppressor gene deactivation almost always result in uncontrolled cell cycle progression and inactivation of apoptotic mechanisms (Sibaji Sarkar *et al.,* 2013). The initiation of a neoplasia is a complex scenario that involves numerous processes, the most important of which are changes in proto-oncogenes and tumor suppressor genes. Proto-oncogenes are activated in particular, while tumor suppressor genes are inactivated (Knudson, 1971). Proto-oncogenes are a set of genes that, when altered, turn normal cells into malignant cells (Adamson, 1987, Weinstein & Joe, 2006). *In situ* cancer refers to cancer cells that remain in one spot in the body, whereas invasive cancer refers to cancer cells that invade other body tissues (Understanding Cancer, 2007).

Cancer remained a mystery to scientists years ago. However, with the development of molecular biology tools over the last three decades, our understanding of cancer has transformed and evolved.

Investigation and analysis of single cell features by the scientific community were unthinkable a century earlier (Understanding Cancer, 2007). We now understand cancer, in its molecular and genetic aspects, and we know many of the molecules and genes involved. Indeed, as our knowledge about these genes increases, so does our ability able to design intriguing new ways to prevent, delay, and even reverse cancerous gene alterations.

1.1.2 Origin and types of cancer cells:

Cancer is caused by a set of genetic mutations that form a malignant cell at random (Pietras, 2011). Chemical carcinogens, chronic inflammation, radiation exposure, and genetic predisposition can all cause such malformations (Reuter *et al.*, 2010). Normal cells gain DNA mutations over time, losing control over their ability to grow and multiply, resulting in unregulated cell proliferation. Cancer cells are aberrant and unstable phenotypically and functionally, with inter- and intra-tumor heterogeneity. The former is a change in tumor composition between different persons with the same cancer type, whereas the latter is defined cellular heterogeneity within the same tumor (Blanpain, 2013). Any tissue can develop cancer cells, but the most common places are the breast, ovary, prostate, liver, stomach, pancreas, lung, brain, and bone marrow. Oncogenic signals and alterations in the genetic make-up of tumors in different organs cause them to behave differently. For instance, prostate cancer is more confined and treatable than pancreatic cancer, which is more aggressive (Klein, 2009).

Cancer can be classified into six major categories based on their tissue type. Carcinoma, which is a cancer that arises from epithelial cells that line the outside of the body or the inside of the bodily organs. Sarcoma, which is a cancer that develops in the connective and supportive tissues. Myeloma, which is a cancer that arises from the plasma cells in the bone marrow. Leukemia, which is a type of cancer that starts in the blood and is split into different types. Lymphoma, which is a cancer that starts in the lymphatic system. Finally, the mixed types of cancer, which are characterized by more than one type of cancer, merged to form a single cancer (Mandal, 2019).

1.1.3 Differences between normal cells and cancer cells

Uncontrolled proliferation is the primary distinction between cancerous and healthy cells. Normal cells divide and proliferate in a controlled manner, and when they get old or damaged, the body destroys them. This is not the case in cancer cells, though. Cancer cells, in reality, continue to grow uncontrolled and evade apoptosis in a variety of ways (National Institutes of Health, NIH-NCI, 2015). Cancer cells can also detach from one another by shedding surface proteins, allowing them to migrate to distant parts of the body, a process known by metastasis. This happens when cancer cells separate from the basement membrane and lose particular surface proteins, which is not possible in healthy cells (Guan, 2015). Other characteristics in cancer cells include aberrant cell signaling, a lack of differentiation, and the presence of multiple mutations in distinct genes such as tumor suppressors and proto-oncogenes (Cancer Research, 2014).

1.1.4 Metastasis

The multi-step intricate process of metastasis includes (i) the local invasion of cancerous cells into the surrounding tissue, (ii) the intravasation of cancer cells through the endothelium into blood vessels, (iii) the preservation of cancer cells in the bloodstream, (iv) the extravasation, and (v) the successive growth in capable organs that results in colonization (Eger & Mikulits, 2005). Metastatic disease is known to be a very exhausting process because a synchronized choreography of numerous events is required to avoid its failure, which would otherwise result in the eradication of emigrating malignant cells at any of the numerous steps along the way (Chambers *et al.*, 2002; Mehlen & Puisieux, 2006).

The debate over whether metastasis occurs early or late in the development of a tumor is still up on-going (Klein, 2009). The "parallel progression model" contends that early tumor cell dispersion occurs in tiny malignancies with a size of 1 - 4 mm, in contrast to the "linear progression model," which says that cancer cell spread occurs after considerable enlargement of original tumors.

Since 90% of all cancer-related fatalities are attributable to metastasis, cancer therapy must prioritize the prevention of cancer cell spread and the development of secondary tumors (Olivier *et al.*, 2006).

1.1.5 Cancer risk factors

Cancer risk increases with lifestyle variables, as shown by strong evidence. This risk can be considerably reduced by positive, population-wide changes (Curry *et al.*, 2003). The most prevalent cancers diagnosed in the industrialized world include colorectal, lung, prostate, and breast cancer. According to current epidemiologic research, the onset of these cancers is linked to behavioral variables (Ezzati *et al.*, 2002). Given the significant risk-modifying effects of modifiable variables, particularly for the most common malignancies, it has been predicted that 50% of cancers may be avoided (MD Anderson Cancer Center., 2020). However, extensive lifestyle modifications are required to achieve significantly lower cancer incidence rates. In the following section, the most common risk factors will be discussed.

1.1.5.1 Tobacco Use

Over 6 million premature deaths annually are attributed to smoking worldwide (World Health Organization, 2014). In the United States of America (USA) tobacco users lose on average over 13 years of their lives because of smoking's harmful effects (Centers for Disease Control and Prevention, 2020). In addition to delivering a variety of toxins, tobacco irritates and inflames the body's natural defenses and interferes with their function, hence affecting numerous phases of carcinogenesis.

1.1.5.2 Physical Inactivity

Colon, breast, and presumably endometrial cancer risks are increased by inactivity (Brenner, D. R. *et. al., 2017)*. Sedentary behavior has also been associated with 5% of cancer deaths worldwide (Park *et. al., 2020*).

Links between physical activity and colon and breast cancers were confirmed, highlighting the role of regular exercise in decreasing cancer risk, independently of its impacts on body weight. The dose-response link between exercise and cancer risk can be explained by various processes: (i) insulin levels, hormones, and other growth factors in the blood are lowered by exercise (Brenner, D. R. *et. al., 2017*) and (ii) prostaglandin levels and the immune response are affected by exercise as well (Scheffer, D. D. L., & Latini, A., 2020). Thankfully, it is possible to reverse the harmful consequences of a sedentary lifestyle since research indicates that increasing physical activity can reverse years of inactivity (Hamilton, M. T. *et al.*, 2008). Even 30 minutes a day of mild exercise, such as brisk walking, dramatically lowers the risk of illness (Pate & Wilmore, 1995)

1.1.5.3 Obesity and Overweight

Excess weight affects hormonal and growth factor levels, in addition to having detrimental effects on health. Numerous cancers, notably postmenopausal breast, colorectal, endometrial, renal, and esophageal cancer, are brought on by being overweight or obese. The population-attributable risk for these diseases extends from approximately 10% (postmenopausal breast cancer) to approximately 40% (endometrial cancer) (International Agency for Research on Cancer, 2002). Recent research by Calle *et al.*, (2003) reveals that obesity may also affect malignancies of the liver, prostate, pancreas, gallbladder, stomach, ovary, and cervix in addition to multiple myeloma and NHL (non-Hodgkin's lymphoma) . Generally, being obese was linked to 14 and 20% of cancer-related deaths in men and women, respectively.

1.1.5.4 Nutrition

The discovery of a diet-cancer relationship has received much scientific attention, and a wide range of variables have been examined. For instance, studies have shown that increasing folate consumption lowers the incidence of adenomatous polyps and colon cancer. Folate, which is also known as Vitamin B, is essential for DNA synthesis, DNA methylation, and plays a role in DNA repair mechanisms (Giovannucci *et al.*, 1993). The interplay of folate with various types of methylenetetrahydropholate reductase provide more evidence for this possible association (Slattery *et al.*, 1999). Additionally, mounting data shows that folate lessens the risk of alcohol consumption on breast carcinogenesis (Zhang *et al.*, 2003). Furthermore, data already in existence suggests that breast cancer risk is also slightly reduced by a high level intake of carotenoids (Zhang *et al.*, 1999).

1.1.6 Staging of Cancer – TNM

One of the three primary axes of tumor categorization, along with topographic location and histological type, is the anatomic expanse of the illness.

Staging offers a platform for the standard interchange of data among doctors on the severity of illness, a foundation for their choice of first therapeutic tactics, and a rationale for their assessment of the potential need for adjuvant therapy. Staging enables information sharing through large datasets and peer-reviewed interactions for clinical researches and permits classification of patients in exploratory and experimental therapy research. To assess nonanatomic prognostic markers at certain anatomic phases, staging is used. Staging can also be used to gauge the effectiveness of early detection measures.

The cancer staging process became common over a century ago with efforts concentrated on certain anatomical cancer locations. The tumor-node-metastasis (TNM) classification, created between 1943 and 1952 by the French surgeon Pierre Denoix, was the first attempt at developing a protocol whose concepts and codes could be applicable to all cancer locations (Denoix, 1952). It was continuously modified and evolved into the version used today (Greene & Sobin, 2008).

The TNM classification indicates the anatomic scope of cancer. It is important because it is thought that treatment options and the likelihood of survival are influenced by the tumor size at the original site (T), the existence or lack of tumor in local lymph nodes (N), and the appearance or absence of metastasis outside of the regional lymph nodes (M). Prior to therapy (clinical staging [cTNM]) and following surgical treatment (pathologic staging [pTNM]), lesions must be classified. As the main tumor grows or spreads, T is often separated into 4 major portions (T1 to T4). At least two categories exist each under N and M, which are 0 (for absence of tumor) and 1 (for tumor existence) (Greene & Sobin, 2008).

1.1.7 Cancer treatment modalities

Currently, cancer is the focus of more than 60% of all active, high-quality therapy studies globally, with the aim of finding new, effective cancer treatment modalities (Wu *et al.*, 2006). Cancer type, location and how far along it is, determine the treatment options and modalities. Some of the oldest and most popular treatments include radiotherapy, chemotherapy, radiation-based surgical instruments, and surgery. Hormone-based treatment, anti-angiogenic therapy, stem cell therapies, immunotherapy, and dendritic cell-based immunotherapy are a few of the contemporary approaches. The limitations of conventional cancer therapy procedures are brought to light by their side effects (Abbas & Rehman, 2018). Therapy against the angiogenic potential of malignancies, oncolytic virotherapy, genetic regulation of apoptotic and tumor signaling pathways, antisense, and RNAi approaches are a few of the innovative therapeutic approaches used to cure malignancy

(Dorai & Aggarwal, 2004). These therapies are mostly used to treat urinary, renal, dermatological, reproductive, and gastrointestinal cancers among others.

1.1.7.1 Surgical Intervention

Surgical intervention, unlike chemotherapy and radiation, guarantees the least amount of tissue injury to the structures surrounding the tumor. Surgery is believed to be one of the most traditional and potential therapies for many malignant and benign tumors (Abbas & Rehman, 2018). The ability to surgically remove the growth without undue danger of cellular injury makes surgery a viable therapy option in some cases. Various surgical procedures, either open or minimally invasive, *i.e.* laproscopic, can be carried out based on several variables, including the size and location of the tumor as well as the patient's preferences and financial capabilities (Abbas & Rehman, 2018).

1.1.7.2 Radiation-Based Surgical Intervention

Another method to eradicate a tumor locally is by surgical procedures based on radiation. There are several varieties. Ionizing radiation is used in stereotactic radiosurgery (SRS), a type of interventional radiology, to destroy specific regions within an organ or tissue. With this method, a very tiny portion of the body absorbs a very high dosage of radiation (Kano, 2017). To treat small to medium-sized lesions and cancers, the Gamma Knife procedure uses light emissions from focused gamma rays (Hassen-Khodja, 2004). High-energy X-rays are used by linear accelerator (LINAC) devices to treat tumors and other ailments (Fogh, 2010). The molecular procedure known as proton beam therapy uses particles such as protons or neutrons rather than radiation beams (Okolsky, 2010).

1.1.7.3 Radiation Therapy

Radiation therapy may be thought of as a mechanical agent used to eliminate cancer cells. It is now a prestigious area of expertise in medicine with branches like radiation oncology. Ionizing radiation is the type of radiation utilized in treatment. The word "ionizing" refers to the radiation's ability to electrically excite particles in biological bodies upon impact. This process transfers charge from the rays to the body's cells as they pass through, resulting in direct cellular destruction, or the induction of apoptosis (Abbas & Rehman, 2018). Radiation treatment has the negative side effect of additionally damaging nearby normal cells that are not part of the primary tumor mass. However, advances in imaging technology, attempts to precisely target the cancer mass, as well as normal cells' capacity to return to normal function thanks to effective repair mechanisms, limit the overall harm caused by radiation (Goldblum *et al.*, 2013).

1.1.7.4 Chemotherapy

Chemotherapy stops the growth of tumors by (i) impairing their capacity to divide and (ii) inducing apoptosis. In normal physiological conditions, cell viability and cell death are balanced and controlled by apoptosis for eliminating old and damaged cells, and cell division for the emergence of new ones. This is not true for cancer cells that have longevity and a greater potential for replication (Abbas & Rehman, 2018), especially in malignant masses where the ratio of cell growth to cell death is high. Chemotherapy alters cancerous cells to either stop their division or destroy them. Unfortunately, because chemotherapy also affects healthy tissue, there may be side effects such alopecia, vomiting, diarrhea, and exhaustion, depending on the dosage. Patients who get intensive chemotherapy treatments lose their immune systems, which can make them more susceptible to serious infections and eventually death. (Bhosle & Hall, 2009).

1.2 Breast Cancer

The most prevalent cancer in women and one of the leading causes of their mortality is breast cancer (Ferley *et al.,* 2013). Breast cancer is a complex illness, involving several underlying factors. Even though this cancer is not specific to a certain region, significant regional variations are seen with breast cancer incidence, as well as mortality/survival rates. Population structure, way of life, genetics, and environment are all factors that could be linked to regional variability of breast cancer (Hortobagyi *et al.,* 2005). The prevalence of breast cancer has increased and is continuing to rise as a result of modifications in related risk variables (Parkin & Fernandez, 2006).

1.2.1 Epidemiology of breast cancer

The most frequent cancer in women and the second most prevalent malignancy overall is breast cancer (Ferley *et al.*, 2013). In the United States of America (USA), there is an estimated 12.4% risk of a woman developing breast cancer during her life (Desantis *et al.,* 2017). An estimated 264,000 cases of breast cancer are diagnosed in women and about 2,400 in men, with 42,000 women and 500 men dying each year from breast cancer in the USA (Center for Disease Control; last updated September 26, 2022). A study done in 2017 differentiating between in situ and invasive breast cancer in the USA show 6,341 new instances of breast cancer in situ and 252,710 new cases of invasive breast cancer, respectively (Desantis *et al.*, 2017).

Globally, the world health organization estimated around 2.3 million women diagnosed with breast cancer and 685 000 deaths globally in the year 2020 (World Health Organization, 2021). Although cancer can occur everywhere in the globe, incidence rates are greater in first world developed countries, and racial and ethnic differences have a significant impact on breast cancer incidence rates (DeSantis *et al.*, 2014). It is estimated that breast cancer incidence rates fluctuate over the world (Ghoncheh *et al.,* 2016), ranging from 27 per 100,000 in Middle Africa and East

Asia to 92 per 100,000 in Northern America. By 2050, it is anticipated that 3.2 million new cases of breast cancer will be diagnosed (Hortobagyi *et al.,* 2005). The incidence rate of breast cancer in the geriatric population increases in number in developed nations (Hortobagyi *et al.,* 2005). The Asia-Pacific area accounts for over 24% of all breast cancer cases. Studies have shown that the highest incidence rates are occurring in China, Japan, and Indonesia (Ghoncheh *et al.,* 2016; Youlden *et al.,* 2014).

Nonetheless, the five-year survival rate for breast cancer was 89% between 2005 and 2011 thanks to the increased access to treatment and screening and surveillance regimens (Rojas & Stuckey, 2016). The survival rate amongst African women is low, which might be attributed to the long delays in detection and treatment of breast cancer (Abdulrahman & Rahman, 2012). Research conducted by Momenimovahed & Salehiniya (2019) documented the following incidence rates of breast cancer in various parts of the world: Western Europe scored 96.0 out of a possible 100, Northern America scored 91.6, Northern Europe scored 89.4, Australia/New Zealand scored 85.8, South-Central Asia scored 28.2, and Eastern Asia scored 27.0.

1.2.2 Mortality

In 2012, 324,000 deaths were confirmed to be from breast cancer, making it the fifth-highest cancer-related deaths globally and the top cause of mortality in developing nations. In industrialized nations, breast cancer was the second leading cause of mortality, second only to lung cancer with 197,000 fatalities, accounting for 15.4% of all fatalities (Ferley *et al.*, 2013). However, the mortality rate of breast cancer has significantly decreased in high-income nations thanks to (i) improved treatment, (ii) diagnostic techniques, and (iii) the advocacy and development in breast cancer management (Carioli *et al.*, 2018). Age is also a factor in the diagnosis and prognosis of breast cancer, with women 50 years of age or older accounting for 89% of breast cancer fatalities in the USA in 2017 (Desantis *et al.*, 2017).

Breast cancer has an age-standardized mortality rate of 12.9 per 100,000 (Kim *et al.*, 2015) worldwide, with Africa having the highest age-standardized mortality rate (Azubuike *et al.*, 2018). In East Asia, there are six cases of death for every 100,000 people, but in Western Africa, there are 20 cases for every 100,000 people (Ferley *et al.*, 2013). In North America, the mortality-to-incidence rate ratio is 0.16, which denotes a greater survival rate, but in Asia, it ranges from 0.23 to 0.48 (Kim *et al.*, 2015). While developed nations have greater rates of breast cancer prevalence, which might be due to a higher rate of diagnosis, the majority of Asian nations have higher fatality rates, which might be caused by a lack of or delayed screening, as well as the poor to moderate economic status, making breast cancer among the leading killers in these nations (Ghoncheh *et al.*, 2015).

1.2.3 Risk Factors

1.2.3.1 Demographic Factors

Gender

Less than 1% of all cancer cases in men are breast cancer, which is a condition mainly affecting women (Giordano *et al.*, 2002). Aging men with either (i) a family history of breast cancer, (ii) hormonal imbalance experienced in previous years, or (iii) were exposed to radiations, were found to be more prone to develop this disease. Genetically, the most prevalent risk factor for breast cancer in men is found to be mutations in the BRCA2 gene (Abdelwahad, 2017).

Age

Age is the second-most significant documented risk factor for breast cancer after gender (Thakur *et al.*, 2017). The incidence of breast cancer notably rises with aging, peaks at menopausal ages, and then progressively decline or remains stable (Kim *et al.*, 2015). In a case-control study, being older than 50 years old was linked to a higher risk of breast cancer (Mahouri *et al.*, 2007). However, in younger women, breast tumors tend to be bigger, at more advanced stages, have positive lymph nodes and a worse prognosis (Assi *et al.*, 2013).

Reproductive Factors

The influence of ovarian hormones, which start around pubescent age and continue during monthly cycles, is the main association factor between reproductive variables and breast cancer. These hormonal changes are evidently influenced by the number of pregnancies and the onset of menopause (Thakur *et al.*, 2017).

Menarche and Menopause

An earlier age of menarche, *i.e.* the first occurrence of menstruation, increases risk of breast cancer by twofolds according to the results of a case-control research (Thakur *et al.*, 2017). Numerous investigations have supported this conclusion (Bhadoria & Kapil, 2013; Clavel-Chapelon & Launoy, 2005; Kim *et al.*, 2015), but others disagreed, stating that there’s no link between sexual maturity at a younger age and a higher risk of breast cancer (Nguyen *et al.,* 2016; Tamakoshi *et al.,* 2005). The latter was also linked to menopause occurring in women older than 50 years, showing a link between breast cancer risk and advanced age at menopause (Kim *et al.*, 2015; Laamiri *et al.*, 2015; Thakur *et al.*, 2017).

Pregnancy

Breast cancer risk declines with parity in women who are of reproductive age (Clavel-Chapelon & Launoy, 1995; Kim *et al.*, 2015). A full-term pregnancy is regarded as a significant predictor (Dai & Liu, 2009; Ma *et al.*, 2010), with women who are nulliparous having a higher breast cancer risk than those with more than three children (Balekouzou & Yin, 2017). Moreover, a full term pregnancy at a younger age can lower the risk of the illness by up to 23% (Laamiri *et al.*, 2015).

However, having more than five full-term pregnancies increased the risk of breast cancer (Mahouri *et al.*, 2007).

A noteworthy remark is a link between women who deliver their first child before thirty-three gestational weeks and a virtually quadrupled chance of developing breast cancer (Willett WC *et al.*, 2014). Furthermore, this study's findings indicated a link between a first pregnancy's placental rupture and a higher risk of breast cancer. Preeclampsia may be protective against breast cancer, according to the findings of several studies (Brasky & Li *et al.*, 2013). This may be because the condition increases levels of IGF-1-binding protein, AFP, and hCG while decreasing levels of the hormones AFP and estrogen (Innes & Byers, 2004).

1.2.3.2 Hormonal Factors

Contraception

Oral hormonal contraceptives pills contribute to the onset of breast cancer, according to several studies (Beaber *et al.*, 2014; Marchbanks & Macdonald, 2002; Bhadoria & Kapil, 2013; Dai & Liu, 2009; Laamiri *et al.*, 2015). This conclusion was however disputed by other studies, mainly involving 35- to 64-year-old women, that suggested that present or past usage of contraceptive tablets was not linked to a higher risk of breast cancer (McDonald and colleagues, 2012; Dai & Liu, 2009; Laamiri *et al.*, 2015).

Post-menopausal hormone therapy

On one hand, hormone therapy was found to have an effect on breast cancer development, as was shown by statistical analyses based on 51 studies. On the other hand, this effect declines once usage is stopped, and disappears after five years (American Cancer Society, 2015). According to the findings of a study including 1 million women, using hormone replacement therapy (HRT) is linked to a higher breast cancer development risk and fatality rate, with the risk being more important for those who use estrogen-progesterone mixture strategies than for those who use alternative HRT techniques (Beral, 2003).

1.2.3.3 Breast-related factors

Lactation

Research has highlighted noted the importance of lactation in preventing breast cancer; it is thought to be a preventive factor against malignancy (Freund *et al.*, 2005; Kim *et al.*, 2015). A relationship between the occurrence of breast cancer and breastfeeding duration was established, with longer nursing periods having an increased protective impact (Bhadoria & Kapil, 2013; Dai & Liu, 2009; Laamiri *et al.*, 2015). According to the findings of a case-control research, the interaction of two preventive factors—two or more pregnancies and nursing for more than 13 months—can lower the risk of breast cancer by up to 50% (Jeong *et al.*, 2017). Additionally, nursing may help breast cancer patients have a favorable prognosis, have a lower recurrence rate, and higher survival rate.

Benign breast disorders

Benign breast illness is a highly significant breast cancer risk factor (Fioretti *et al.*, 1999; Roman & Quintana, 2017; Zendehdel *et al.*, 2018). Histological categorization and a family history may play a crucial role in determining the risk. Postmenopausal women with benign breast conditions, such as fibrosis, have a lower chance of developing breast cancer (Arthur *et al.*, 2017).

1.2.3.4 Genetic Factors

Twelve Key Genes

Thanks to the advances in biotechnological and DNA analysis tools, the scientific community was able to highlight genes that, if modified, would be involved in the onset of tumor development. In the following section, we enumerate and delve in 12 genes that were identified as key in the onset of invasive breast carcinoma. Up to 85% of breast cancer cases in women by the age of 60 result from *TP53* gene mutations. The median age at diagnosis for most of these breast cancers is 34 years. A germline *TP53* gene mutation is prevalent in 5-8% of breast cancer patients under the age of 30. Her2 and/or hormone receptor positivity are more prevalent in breast malignancies in women who have *TP53* mutations (Schon, K., & Tischkowitz, M., 2018). *ERBB2* is crucial for the development of human cancers. About 30% of human breast cancers and many other cancer types, such as ovarian, stomach, bladder, salivary, and lung carcinomas, have the *ERBB2* gene amplified or overexpressed (Slamon DJ *et al.,* 1989). Amplification or overexpression of *ERBB2* impairs normal cell-control processes and results in the development of aggressive tumor cells, according to a large body of research. In comparison to individuals whose cancer does not overexpress *ERBB2*, patients with breast cancer resulting from *ERBB2*-overexpressing had significantly worse overall survival rates and shorter disease-free intervals (Ming Tan *et al.,*2013). Moreover, increased breast cancer metastasis is caused by overexpression of *ERBB2* (Holbro T. *et al.,* 2003). Although several genetic variables influence the risk of breast cancer development, dominant autosomal mutations in the *BRCA1* and *BRCA2* genes account for around 40% of inherited breast cancer cases (Cobain & Milliron, 2016). According to a study done in 2017, 55 to 65% of *BRCA1* mutation holders and 45% of *BRCA2* mutation holders would acquire malignancy by the time they are seventy years old (Godet & Gilkes, 2017). Another gene associated with lobular breast cancer if mutated is *CDH1*, which, in its human form encodes for E-cadherin protein (Corso *et al.,* 2016). *CHEK2* is also a crucial gene in tumorigenesis of breast cancer once mutated. In fact *CDH1* and *CHEK2* are even more common than *BRCA1* or *BRCA2* in tumorigenesis (Apostolou & Papasotiriou, 2017). *PTEN, STK11, BRIP1, ATM, PALB2, and EGFR,* once mutated, also contribute to breast cancer to a certain degree (Lynce & Isaacs, 2016).

Family History

A previous familial history of breast cancer incidences is one of the most important risk factors. Researchers found that women without *BRCA* gene defects who have a family history of breast cancer (two or more instances in women under 50 years old or three or more cases at any age) are around eleven times more prone to illness (Metcalfe *et al.*, 2009).

1.2.4 Types of Breast Cancer

1.2.4.1 Histological subtypes

Adenocarcinomas are malignant tumors that develop from mammary epithelial cells. In breast adenocarcinomas, clinicians have discovered a few distinctive patterns, architectural and cytological, that are invariably linked to various clinical presentations and/or prognosis (Weigelt & Reis-Filho, 2009). Up to 25% of all breast tumors exhibit these characteristics, which are known as "histological special kinds." There are currently at least 17 different acknowledged histological special kinds (Ellis, 2003). Invasive ductal carcinomas not otherwise defined or of no special type refer to adenocarcinomas that lack the necessary criteria to qualify as one of the special kinds. This term refers to 50–80% of breast carcinomas. It is important to highlight that histopathological grade and type offer complimentary knowledge regarding the clinical and pathological characteristics of the tumor (Rakha, 2008).

1.2.4.2 Architectural subtypes

The lobules, which oversee the production of milk, the ducts, which carry milk to the nipple, and the connective tissues, which keep the entire structure together, make up the three sections of the breast. When speaking of noninvasive breast cancer, two types are discussed: lobular carcinoma in situ (LCIS) and ductal carcinoma in situ (DCIS). The configuration of the typical lobules' contour is a feature of LCIS. Malignant cells are known to fill the spaces in DCIS. When speaking of invasive breast cancer, however, different types can be differentiated: Infiltrating ductal carcinoma (IDC), one of the most prevalent forms, is responsible for roughly 80% of breast cancer diagnoses. IDC originates in the ducts that generate breast milk and then spreads through the duct walls to the structures enclosing the breast. Infiltrating lobular carcinoma (ILC), a kind of malignancy that originates in the milk lobules and travels to other parts of the body, has been linked to 10-15% of all breast cancer cases. There are less prevalent forms of breast cancer that do occur. Medullary carcinoma, mutinous carcinoma, and tubular carcinoma are the less frequent (Sharma *et al.,* 2010).

1.2.5 Staging of Breast Cancer

The tumor-node-metastasis (TNM) approach for cancer staging includes seven versions that have been released by the American Joint Committee on Cancer (AJCC) since 1959.

The foundation of the AJCC staging system's seventh edition was topographical staging, which included the size of the original tumor (T), the condition of the local lymph nodes (N), and the presence or absence of metastases (M). The eighth edition has mostly not modified this. The T stage, which ranges from T1 to T4, is determined by the size and extent of loco-regional invasion by the main tumor. The degree of nodal involvement determines the N stage. To identify the M stage, distant metastases are assessed. Based on various combinations of T, N, and M status, the seventh edition employed nine phases (0, IA, IB, IIA, IIB, IIIA, IIIB, IIIC, and IV) (Jieun & Min Jung, 2019). The eighth edition, whose improvements were founded on the strongest degree of proof from recently collected clinical and pathological data, was announced in 2017 *(Gabriel et al.,* 2017). A fundamental shift has occurred, and breast cancer is now viewed as a collection of illnesses with various molecular traits that signal various prognoses, patterns of recurrence, disseminations, and sensitivity to existing therapy (Prat & Pineda, 2015). As a result, the committee modified the conventional anatomic TNM staging to include biomarkers (histologic grade, hormone receptor, HER2 expression, and multigene panels) (Gabriel *et al.*, 2017).

1.2.6 Principles of Therapy

The major objectives of therapy for nonmetastatic breast cancer include removing the malignancy from the breast and local lymph nodes and avoiding metastatic spread and reoccurrence. Local treatment for nonmetastatic breast cancer entails surgical excision of the tumor, biopsy or resection of the axillary lymph nodes; additionally, postoperative radiotherapy might be considered. Neoadjuvant and adjuvant systemic treatment are both possible. The conventional systemic therapy for breast cancer is determined by the subtype and includes chemotherapy alone for triple-negative breast cancer, trastuzumab-based *ERBB2*-directed immunotherapy plus chemotherapy for all HR+ tumors (with endocrine therapy given in addition, if concurrent HR positivity), and chemotherapy for all *ERBB2+* tumors (Waks & Winer, 2019).

The treatment targets for metastatic breast cancer include symptom relief and life extension. Almost of individuals with metastatic breast cancer are still incurable currently. Metastatic breast cancer is treated with systemic treatment using the same fundamental categories as the neoadjuvant/adjuvant strategies described below. Only in cases of metastatic illness are local therapeutic techniques (surgery and radiation) often employed for symptomatic relief (Waks & Winer, 2019).

1.3 Bioinformatics

The study of bioinformatics involves the use of computation to derive information from biological data. Through the creation of algorithms and software, it encompasses the gathering, storing, retrieving, manipulation, and modeling of data for analysis, visualization, or prediction (nature.com). A logical development of the Human Genome Project, the Cancer Genome Atlas

(TCGA) project aims to create a thorough atlas of genetic alterations related to cancer (Hanauer,David A. *et al.*, Current Molecular Medicine, Volume 7, Number 1, 2007).

1.3.1 cBioPortal

The collective TCGA project results and studies from various literatures, cBioPortal now has more than 200 studies. Numerous gene datasets can be studied and integrated using this gateway. The cBioportal facilitates the research of various gene mutation types in cancers along with other functions by analyzing and mapping the gene expression, copy number, and mutations. Due to the complex nature or invasive breast carcinoma and the genes implicated in carcinogenesis and tumor formation, a software like cBioPortal is crucial to be utilized to research and contrast all the elements in this particular cancer type. The cBioPortal ((http://cbioportal.org) for Cancer Genomics was created to make it easy for scientists to acquire complex data and to translate genomic information into biological understanding, clinical studies, and treatments. (Gao *et al.*, 2014).

The simplicity of usage is one of cBioPortal's key characteristics. A simplified 4-step online interface makes all of the portal's functionality accessible. Users may be directed to decide:

1) An interesting cancer study, such as invasive breast carcinoma

2) Genomic profiles, such as copy number variations and mutations.

3) A case study involving all TCGA patients who have an invasive breast carcinoma

4) A set of genes of interest. Users can insert any arbitrary gene sets or pathways of interest by writing the gene symbols or gene identities.

Additionally, users could compute the co-occurrence and mutual exclusivity of all genes. The instrument allows users to run cross-cancer queries. (Cerami *et al.*, 2012)

1.3.2 Cytoscape

Based on experimental data, Cytoscape enables biologists to investigate and analyze biological networks like pathways and gene regulatory systems. When information on gene expression and mutual exclusivity score, for instance, was gathered and then submitted to Cystoscape, a network of nodes was created. A network like this can be imported.

Since a file doesn't have a layout when it is submitted to Cytoscape, Cytoscape will automatically create one.

1) Select "Layout Apply Preferred Layout" from the menu.

2) To examine the styles, select the Styles tab.

3) To view the edge attribute mappings, click the "Edge" tab at the bottom.

4) To see "Stroke Color," scroll down.

5) Click and choose "interaction" in the Column area.

Cytoscape allows users to conceal labels, node images, and other data. These are viewable by selecting "View Show Graphics Details" Tab. (Su *et al.*, 2015).

1.3.3 MutationTaster

DNA sequence changes' pathogenicity is assessed by MutationTaster. It foresees the effects of amino acid substitutions on the protein's ability to function. It also demonstrates the impacts of deletion, synonymous and intronic modifications. Now, the average time for one analysis is less than 0.10 seconds. The specific mutation is automatically mapped to all the genes and transcripts by MutationTaster, which also provides information about the mutation's chromosomal location and expected effects (Schwarz *et al.*, 2014).

1.4 Mutual Exclusivity

In cancer, relationships between genetic changes, such co-occurrence or mutual exclusivity, are frequently seen. Understanding these relationships may shed light on the disease's origin and clinical therapy (Elisabeth Remy *et al.*, 2015). Mutual exclusivity suggests synthetic lethality, a phenomenon of interest in the study of cancer treatments, while co-occurrence suggests a cooperative involvement in the development of cancer (Daub, J.T., *et al.*, 2021). Studying which gene combination is co-occurrent or mutually exclusive is used to identify a set of related driver genes that are of great importance and predictive of tumorigenesis and patient survival.

STUDY OBJECTIVES

Bioinformatics have evolved and shaped our understanding by incorporating huge databases for the purpose of detection and application of targeted therapies for numerous diseases. This is happening by the gathering of data on all human genes and proteins, then analyzing the sequence and molecular data for genome-based therapeutics.

In this study, we aim to uncover the genetic risk factors associated with breast cancer, via a data mining bioinformatic approach. Following extensive literature review, 12 key role genes were identified in Invasive Breast Carcinoma and will be the subject of analysis and comparison within our study.

Using several bioinformatic databases and tools, detailed below, we aimed to:

- Generate a network of mutual exclusivity of genes of interest.
- Conduct a mutational analysis to identify single nucleotide polymorphisms, and estimate the impact of the variant on the gene product or protein, to predict oncogenicity.
- Search for direct competitive inhibitors (drugs) of the 12 studied genes.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Key role genes in Invasive Breast Carcinoma

Following extensive literature review, data was collected for 12 key role genes in Invasive Breast Carcinoma - *BRCA1, BRCA2, PTEN, TP53, CDH1, STK11, CHEK2, BRIP1, ATM, PALB2, ERBB2, EGFR* - from 3,617 samples (3,610 patients).

2.2. cBioPortal

2.2.1. Introduction

cBioPortal was used to integrate our 12 query genes. The portal enables the analysis across genes, samples, and data formats for the study of multidimensional cancer genomics data. Users can examine the genomic modification in a tumor sample and compare the frequency of gene alteration. Users can also examine all biological pathways, survival analysis, and mutual exclusivity analysis. The data and a publication-quality summary graphic are available for download by users. Access to cancer genetic data is made easier through cBioportal. This portal offers a straightforward method for combining data, interpreting graphs, and visualizing outcomes, all of which assist scientists and researchers in using the cancer genomic sets for gaining fresh understanding of cancer biology (Cermani *et al.*, 2012; Gao *et al.*, 2014).

2.2.2 Oncoprint

The genomic changes in the sample across the query genes are summarized and represented graphically in an oncoprint. The columns stand in for the samples, while the rows stand in for the genes. Gene expression alterations such as mutations, amplifications, deletions, and color coding are displayed using glyphs. By sliding the mouse over the column, you may see more details, such as the patient ID or cell line, and see the clinical features. Additionally, heatmaps can be added to the Oncoprint for a better perspective of the data. By selecting the SVG button, you can download the Oncoprint as an XML file in the scalable vector graphic (SVG) format, which can then be exported (Gao *et al.*, 2014; Wu *et al.*, 2019). On Prism 9, the Oncoprint was graphically represented.

2.2.3 Mutual Exclusivity

In order to determine whether two genes co-occur or are mutually exclusive, cBioPortal can recognize patterns between them. The portal analyzes these relationships using Fisher's exact test (Gao *et al.*, 2014; Wu *et al.*, 2019).

2.2.4 Mutations

A visual representation of every non-synonymous mutation discovered in the target gene is provided on this page. All of the mutation positions, sample identification number, protein change, annotations, copy number, type (nonsense, missense, frameshift insertion or deletion, splice site, in-frame insertion or deletion), mutation’s number in the sample, mutation’s number at this position in COSMIC (Catalogue of Somatic Mutations in Cancer), and frequency are all included in the summary. The RefSeq isoform has become the standard for all DNA canonical mutations. (Gao *et al.*,2014).

2.2.5 Comparison/Survival

The differences including the altered group and the unmodified group in terms of overall survival and disease-free survival might be examined. Additionally, charts to contrast tumor samples with and without a particular mutation in a query gene are readily available. Kaplan Meier charts are used to display the results. (Gao *et al.*,2014).

2.3 Gene Expression Profiling and Analysis

A total of 3,617 breast tissue samples from 3610 patients were analyzed for their mutations in the 12 genes: *BRCA1, BRCA2, PTEN, TP53, CDH1, STK11, CHEK2, BRIP1, ATM, PALB2, ERBB2 and EGFR* using The Cancer Genome Atlas (TCGA) which is a landmark cancer genomics program that has characterized and sequenced thousands of cancer samples (cancer.gov). Data analysis was performed using cBioPortal. The data from each cohort was quantile normalized separately after applying strict quality control criteria using the program's algorithm. Venn diagrams and tables were generated using GraphPad Prism to plot the data accordingly.

2.4 Targeted therapeutic approaches

Direct competitive inhibitor drugs of the 12 studied genes were mined using PharmGKB software. The latter is widely used by researchers as a platform to detect the best therapies for specific gene groups.

2.5 GraphPad Prism

GraphPad Prism is a statistical software that is designed precisely for many studies including analysis of categorical and quantitative data. This makes it simpler to enter data accurately, select appropriate analysis, and produce appealing and understandable visualizations. The nineth version was used in this project.

CHAPTER THREE

RESULTS

3.1. Oncoprint

Samples from Invasive Breast Carcinoma (TCGA, PanCancer Atlas)

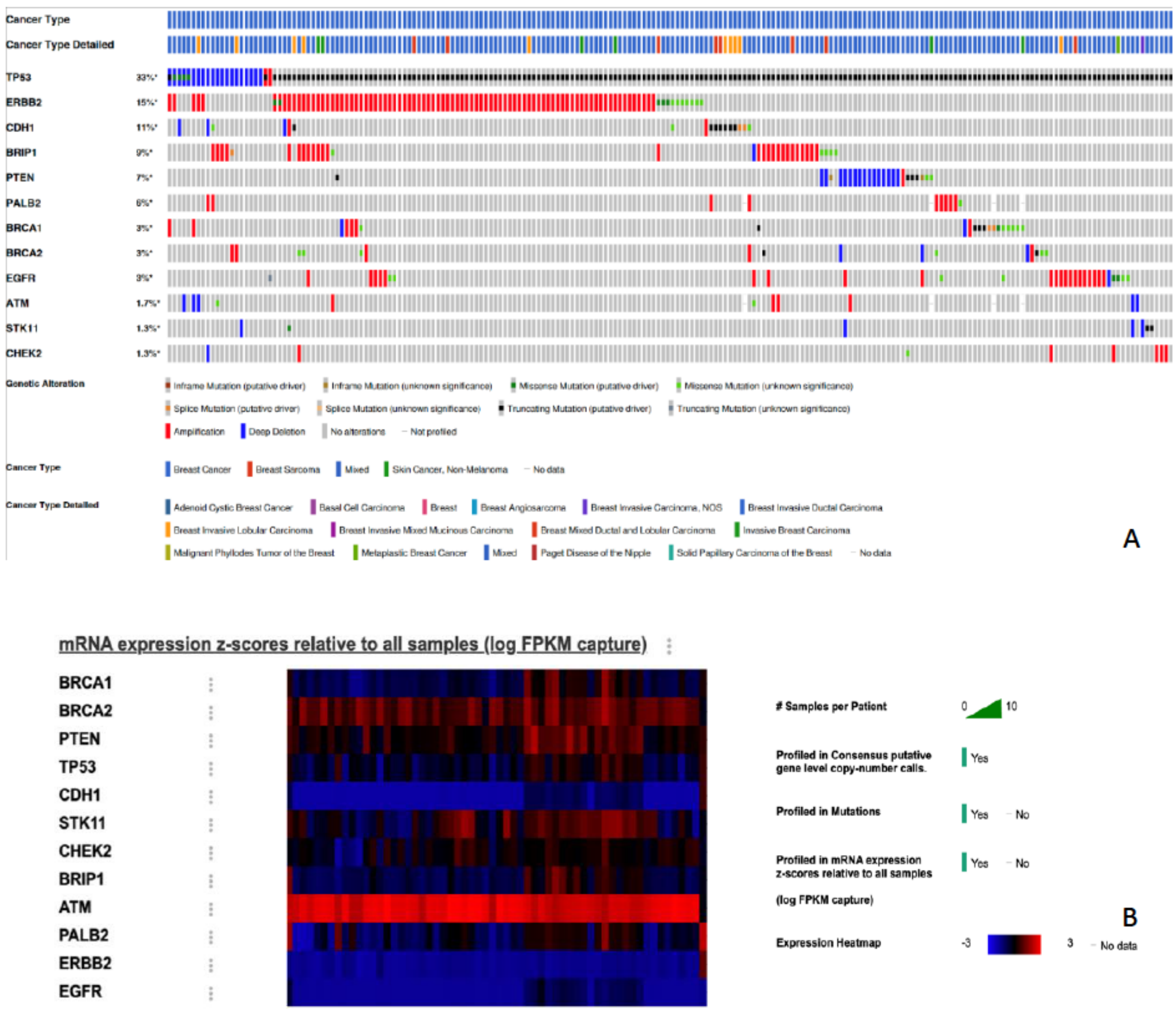


Figure 1 - *(A)Oncoprint of the 12 key role genes BRCA1, BRCA2, PTEN, TP53, CDH1, STK11, CHEK2, BRIP1, ATM, PALB2, ERBB2, EGFR in 3,617 samples of Invasive Breast Carcinoma. Each dash represents a patient and indicates the kind of mutation that person is carrying; (B) Heatmap showing the changes in expression of the 12 key role genes in relative to the various samples.*

The oncoprint of the 12 analyzed genes in 3,617 samples of invasive breast carcinoma reveals the sorts of mutations occurring in each gene along with their frequencies. The genes of interest and the percentage of alterations are displayed in Figure 1A.

TP53 scored the highest alteration percentage of 33%. This was followed by *ERBB2* with 15%, *CDH1* with 11%, *BRIP1* 9%, *PTEN* 7%, *PALB2* 6%, *BRCA1* 3%, *BRCA2* 3%, *EGFR* 3%, *ATM* 1.7%, *STK11* 1.3%, and *CHEK2* 1.3%.

The percentage of alteration in every of our twelve studied genes are shown in Figure 2. To dive deeper and see the type of mutations happened in every gene, a detailed alterations type was conducted and shown in Figure 3.

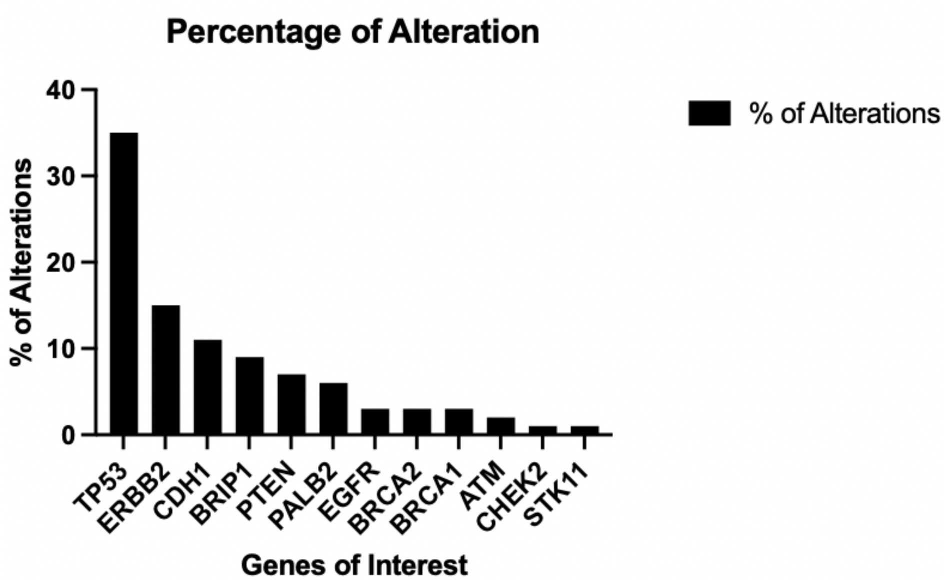


Figure 2 - *Percentage of alteration in the twelve queried genes of interest according to the Oncoprint.*

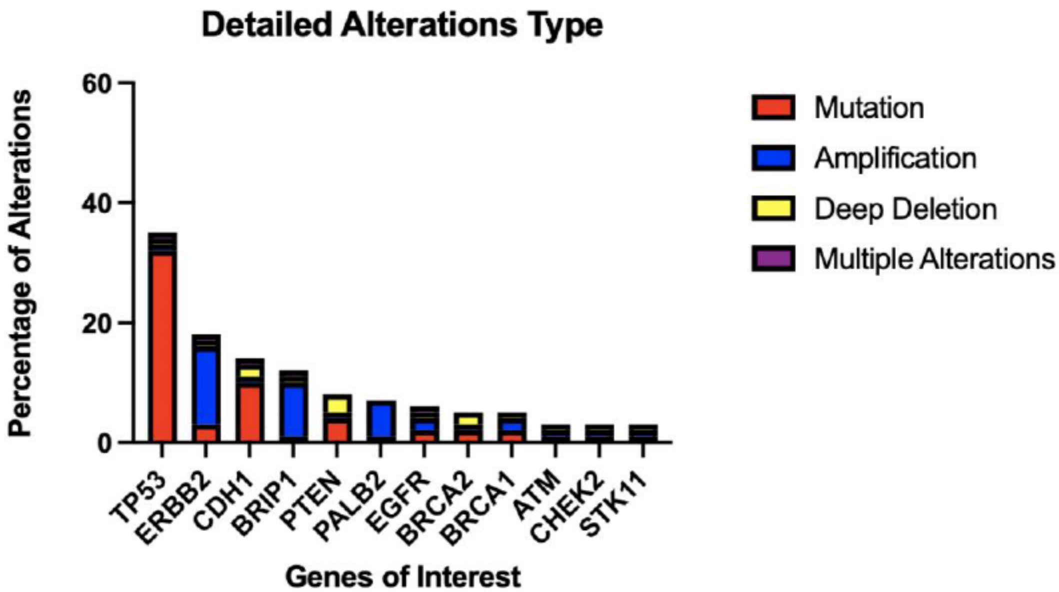


Figure 3 *- Detailed alterations type in the twelve queried genes of interest according to the Oncoprint. Each mutation type is shown in a different color, highlighted in the legend on the right of the figure.*

TP53 scored the highest alteration percentage followed by *ERBB2* and *CDH1*. A mixture of different mutations was observed in our 12 studied genes, such as amplifications, deep deletions and multiple alterations.

3.2. Mutual Exclusivity

Mutual exclusivity may reveal crucial details about cancer, including its various subtypes and different growth paths. Additionally, it can highlight crucial connections between genes and pathways. For example, if two genes share a cancer-driving pathway, a mutation in only one of them may be enough to dysregulate the system and cause cancer to progress.

A	B	Neither	A Not B	B Not A	Both	Log2 Odds Ratio [a]	p-Value [b]	q-Value [c]	Tendency [d]
TP53	*ERBB2*	1853	747	230	306	1.723	<0.001	<0.001	Co-occurrence
BRIP1	*ERBB2*	2408	192	405	131	2.020	<0.001	<0.001	Co-occurrence
TP53	*CDH1*	1778	1000	305	53	-1.694	<0.001	<0.001	Mutual exclusivity
BRCA1	*ERBB2*	2546	54	485	51	2.310	<0.001	<0.001	Co-occurrence

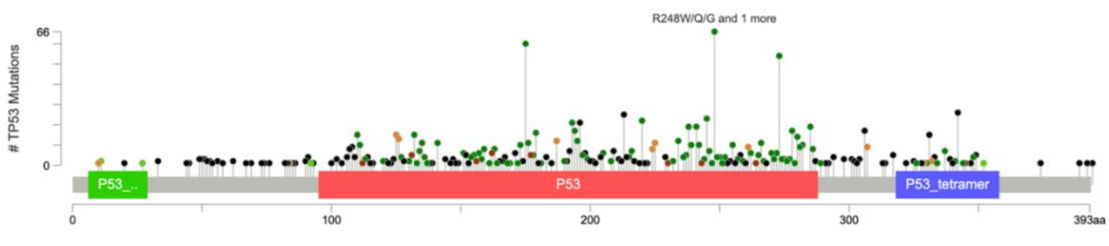
BRCA1	*BRIP1*	2742	71	289	34	2.184	<0.001	<0.001	Co-occurrence
TP53	*EGFR*	2040	983	43	70	1.756	<0.001	<0.001	Co-occurrence
PTEN	*ERBB2*	2383	217	521	15	-1.661	<0.001	<0.001	Mutual exclusivity
TP53	*STK11*	2068	1024	15	29	1.965	<0.001	<0.001	Co-occurrence
CDH1	*BRIP1*	2473	340	305	18	-1.220	<0.001	<0.001	Mutual exclusivity
BRCA2	*BRIP1*	2724	89	300	23	1.231	<0.001	0.005	Co-occurrence
BRCA2	*CDH1*	2690	88	334	24	1.135	0.001	0.008	Co-occurrence
CDH1	*EGFR*	2669	354	109	4	-1.854	0.002	0.013	Mutual exclusivity
PTEN	*TP53*	1949	134	955	98	0.578	0.003	0.013	Co-occurrence
BRCA1	*TP53*	2026	57	1005	48	0.764	0.006	0.027	Co-occurrence
ERBB2	*EGFR*	2516	507	84	29	0.777	0.012	0.055	Co-occurrence
BRCA2	*STK11*	2985	107	39	5	1.839	0.019	0.078	Co-occurrence
BRIP1	*ATM*	827	95	32	9	1.292	0.026	0.097	Co-occurrence
TP53	*CHEK2*	2061	1032	22	21	0.931	0.027	0.097	Co-occurrence
CHEK2	*BRIP1*	2779	34	314	9	1.228	0.028	0.097	Co-occurrence
BRCA2	*TP53*	2018	65	1006	47	0.537	0.037	0.121	Co-occurrence
CDH1	*ERBB2*	2292	308	486	50	-0.385	0.053	0.161	Mutual exclusivity
ATM	*ERBB2*	795	31	127	10	1.014	0.054	0.161	Co-occurrence
CHEK2	*ERBB2*	2568	32	525	11	0.750	0.103	0.296	Co-occurrence
BRCA2	*ATM*	882	40	37	4	1.253	0.112	0.308	Co-occurrence
TP53	*ATM*	632	290	24	17	0.626	0.121	0.308	Co-occurrence
BRCA2	*PTEN*	2804	100	220	12	0.613	0.121	0.308	Co-occurrence
BRCA2	*ERBB2*	2512	88	512	24	0.420	0.134	0.327	Co-occurrence
BRCA1	*BRCA2*	2925	99	106	6	0.742	0.171	0.399	Co-occurrence
BRCA1	*EGFR*	2924	99	107	6	0.728	0.176	0.399	Co-occurrence
BRCA1	*ATM*	889	33	38	3	1.089	0.194	0.414	Co-occurrence
BRCA2	*CHEK2*	2984	109	40	3	1.038	0.197	0.414	Co-occurrence
CHEK2	*EGFR*	2983	40	110	3	1.024	0.201	0.414	Co-occurrence
PTEN	*CHEK2*	2866	227	38	5	0.732	0.209	0.418	Co-occurrence
PTEN	*STK11*	2865	227	39	5	0.694	0.223	0.433	Co-occurrence
BRIP1	*PALB2*	817	101	42	3	-0.791	0.263	0.484	Mutual exclusivity
TP53	*BRIP1*	1874	939	209	114	0.122	0.264	0.484	Co-occurrence

Table 1 *- Mutual exclusivity of BRCA1, BRCA2, PTEN, TP53, CDH1, STK11, CHEK2, BRIP1, ATM, PALB2, ERBB2 and EGFR in 3617 samples of Invasive Breast Carcinoma.*
*[a] quantitative strength evaluation of the presence or absence of alterations in A, and their association with the presence or absence of alterations in B. the odds ratio is calculated based on (Neither * Both) / (A Not B * B Not A). [b] According to a one-sided Fisher Exact Test. [c] According to Benjamini-Hochberg FDR correction procedure. [d] co-occurrence and mutual exclusivity are determined based on the lo2 ratio being >0 or ≤0, respectively; a significant association is determined by a q-value < 0.05.*

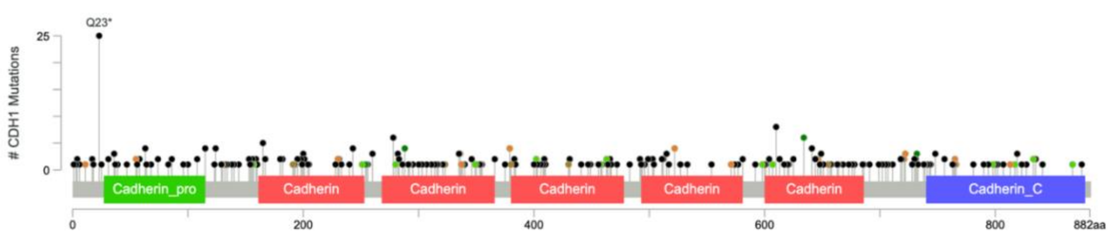
The majority of the gene combinations of our 12 studied genes are co-occurring. Only a few combinations showed mutual exclusivity.

3.3. Mutations

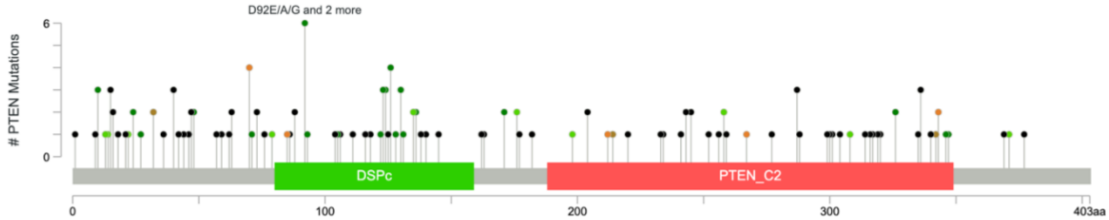
Mutations occurring in gene domain, their frequency, as well as the mutation type and annotations were detected and visualized through cBioportal. As shown in Figure 4 with the lollipop graphs, mutations detected in each domain of proteins encoded by each of the 12 genes were displayed. This annotation shows a variety of mutations occurring in protein domains, and the frequency of each mutation. Each unique type of protein change would lead to specific clinical consequences and has specific hotspots, and background data.



(a)

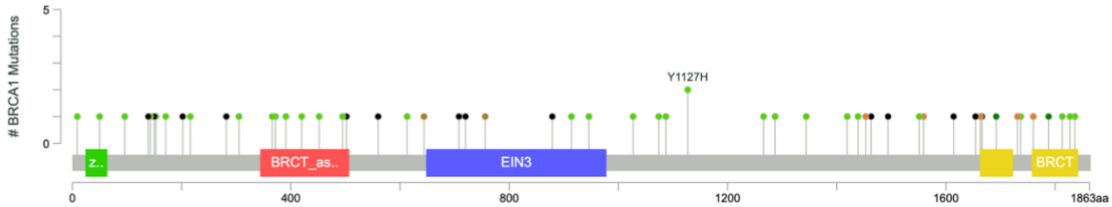


(b)



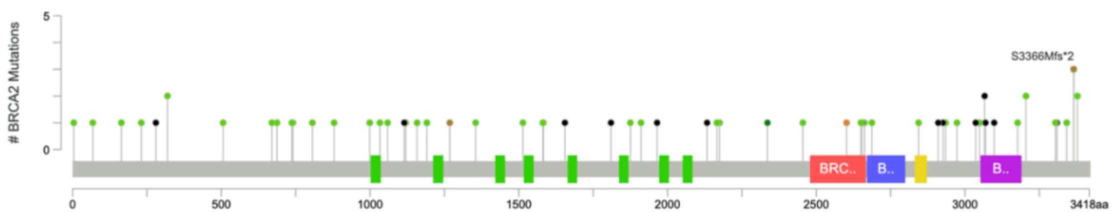
PTEN Mutations
D92E/A/G and 2 more
6
0
DSPc
PTEN_C2
0
100
200
300
403aa

(c)



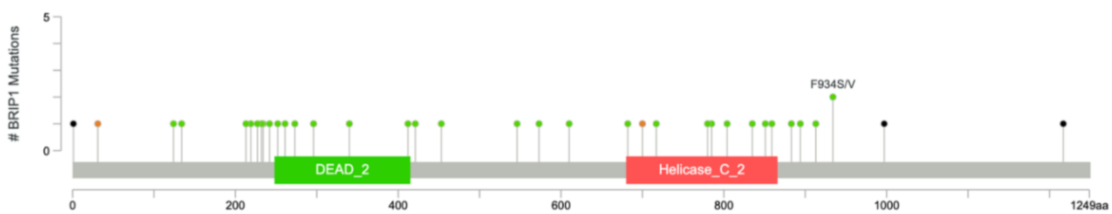
BRCA1 Mutations
Y1127H
5
0
Z..
BRCT_as..
EIN3
BRCT
0
400
800
1200
1600
1863aa

(d)



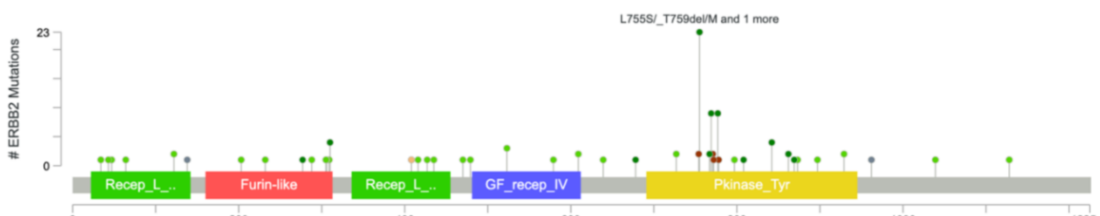
BRCA2 Mutations
S3366Mfs*2
5
0
BRC..
B..
B..
0
500
1000
1500
2000
2500
3000
3418aa

(e)



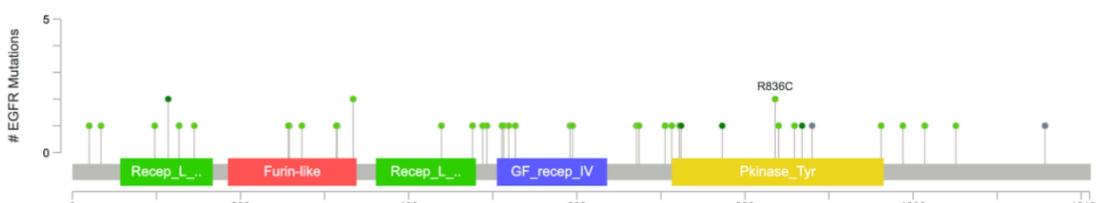
BRIP1 Mutations
F934S/V
5
0
DEAD_2
Helicase_C_2
0
200
400
600
800
1000
1249aa

(f)



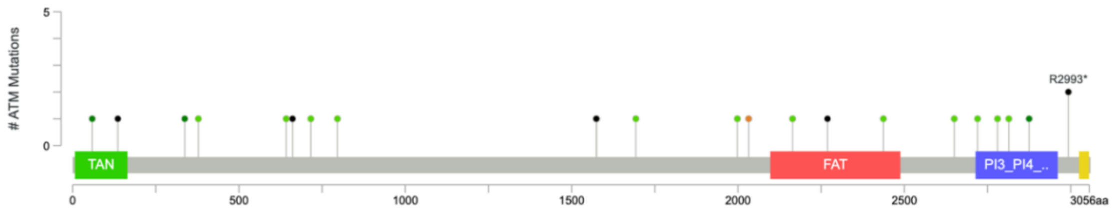
ERBB2 Mutations
L755S/_T759del/M and 1 more
23
0
Recep_L_..
Furin-like
Recep_L_..
GF_recep_IV
Pkinase_Tyr
0
200
400
600
800
1000
1225aa

(g)

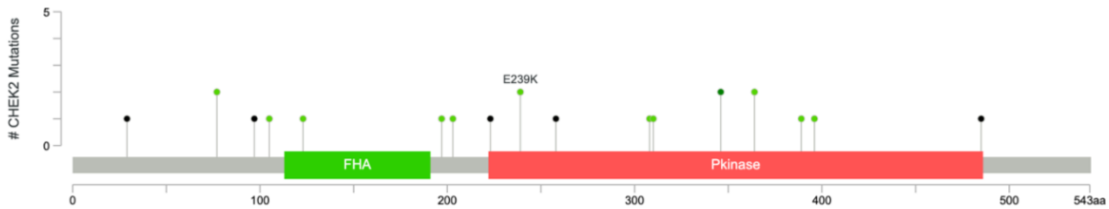


EGFR Mutations
R836C
5
0
Recep_L_..
Furin-like
Recep_L_..
GF_recep_IV
Pkinase_Tyr
0
200
400
600
800
1000
1210aa

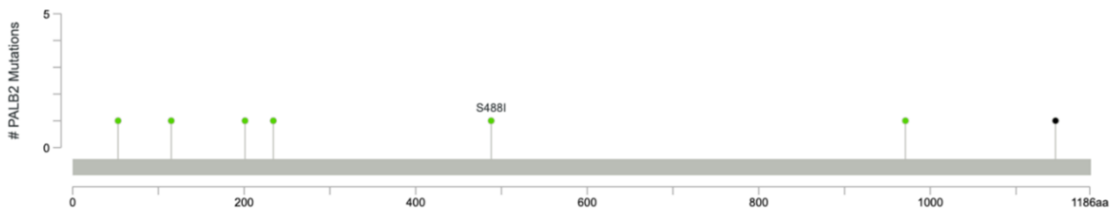
(h)



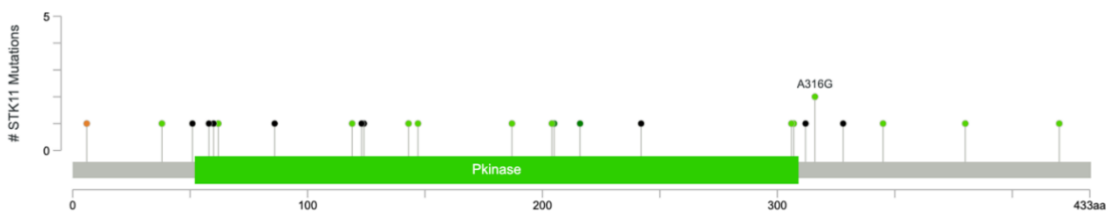
(i)



(j)



(k)



(l)

Figure 4 *(a-l) - The lollipop models of the twelve query genes showing respectively TP53, CDH11, PTEN, BRCA1, BRCA2, BRIP1, ERBB2, EGFR, ATM, CHEK2, PALB2, STK11.*

A higher dash reflects a more common mutation. Black circles represent nonsense mutations, whereas mutations with a green circle are missense, truncating mutations are shown in light blue color and splice mutations in orange color.

3.4. Mutations Oncogenicity

Oncogenicity status helps in understanding the origin and change that lead to carcinogenesis. Genes and their mutations are classified as oncogenic, likely oncogenic, neutral, or inconclusive based on the effect they leave once mutated, based on observational, clinical and experimental evidence (Hanahan and Weinberg, 2011). An oncogenic or likely status are granted in the case the change is known to or likely to promote cell proliferation or other cancer-related characteristics, respectively. As for a neutral classification, it is granted when the mutated gene produced in cells, the modification does not affect protein function or give a growth or survival benefit. As for the inconclusive status, the oncogenic effect of the mutant change is described by conflicting or insufficient. The table below (Table 2) shows our genes of interest with specific position mutations and their oncogenicity status.

Oncogenic	**Likely oncogenic**	**Neutral**	**Inconclusive**
G724S (EGFR) *E868G (EGFR)* *E114K (EGFR)* *H773D (EGFR)* *R130G (PTEN)* *R130Q (PTEN)* *C124S (PTEN)* *D92E (PTEN)* *C71F (PTEN)* *D769Y (ERBB2)* *L755S (ERBB2)*	*D1692N (BRCA1)* *Y1463 (BRCA1)* *C3069LFS5 (BRCA2)* *Y1655 (BRCA2)* *R181C (TP53)* *R175H (TP53)* *D288N (CDH1)* *S216F (STK11)* *R346H (CHEK2)* *W1217 (BRIP 1)* *R337H (ATM)* *Q1146 (PALB2)*	*R98Q (EGFR)* *V5921 (EGFR)* *L755M (ERBB2)* *A1039T (ERBB2)* *V1128I (ERBB2)* *T306M (ERBB2)*	*F2042 (STK11)* *G309A (ERBB2)* *G727A (ERBB2)*

Table 2 - *Oncogenicity status for specific mutations in Invasive Breast Carcinoma. Data from cBioPortal. Genes are indicated in the parenthesis.*

According to Douglas Hanahan and Robert Weinberg, the hallmarks of cancer start with continuous proliferation signaling followed by resistance to cell death, then escaping growth suppressors reaching to metastasis and finally angiogenesis.

3.5. Comparison and survival

Patient survival rates were assessed in two cases: presence and absence of the gene alterations in the 12 studied genes (Figure 5). The unmodified group consists of individuals who have no alterations in any of the studied genes.

Patients with invasive breast cancer who have no mutations in the relevant genes are included in the unaltered group. Nevertheless, they may still have alterations in unrelated genes that are outside the focus of the study. The altered group, shown in red color, have mutations in the 12 studied genes, which showed lower overall survival rates.

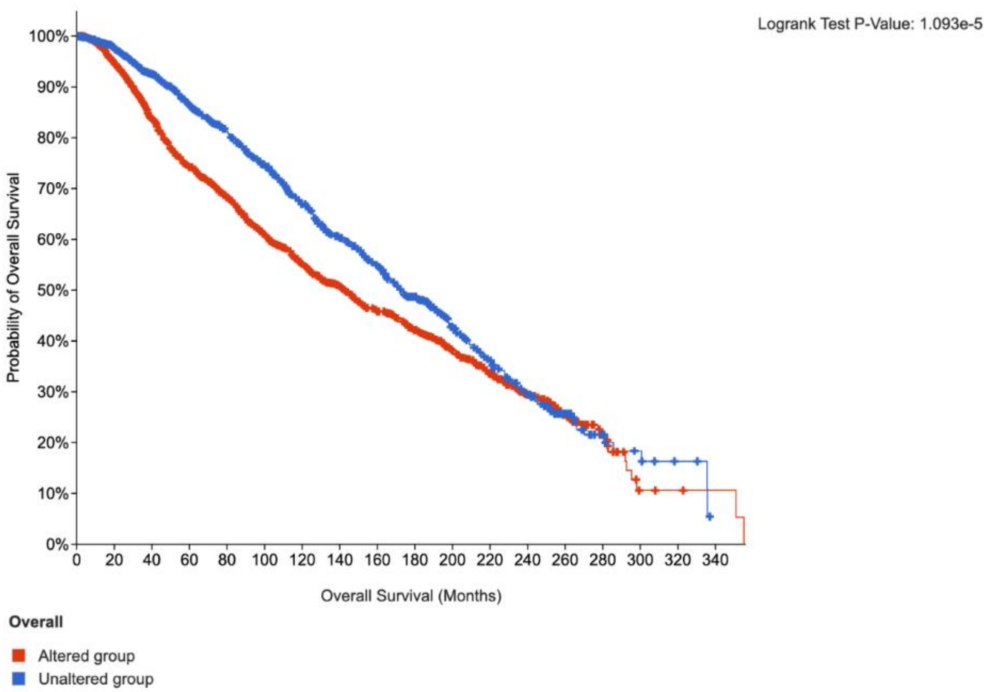


Figure 5 *- The overall survival of the unaltered group in comparison to the groups of the query altered genes.*

3.6. Therapies for Invasive Breast Carcinoma

Certain drugs are specified in table 3. using the mentioned bioinformatics tools to match the mutations. This means that every drug listed was tested to be effective for the specified mutations in our genes of interest.

Mutation	**Therapy**
D1692N (BRCA1)	Talazoparib, Olaparib
R130G (PTEN)	GSK2636771 and AZD8186
R346H (CHEK2)	Olaparib
Q1146 (PALB2)	Rucaparib
D7694 (ERBB2)	Neratinib
R181C (TP53)	NA
D288N (CDH1)	NA
S216F (STK11)	NA
D769Y (ERBB2)	Ado-Trastuzumab Emtansine
G724S (EGFR)	NA
H773D (EGFR)	NA
E114K (EGFR)	NA

Table 3 *- Therapies for specific mutations in Invasive Breast Carcinoma. Data from cBioPortal.*

CHAPTER FOUR

DISCUSSION

Cancer personalized medicine or precision medicine is based on the study of genes involved in cancer development. At the foundation of precision cancer medicine, is targeted cancer therapy. The latter would target proteins that control cellular replication and growth, as well as dispersion. In other cases, it can involve certain biomarkers or angiogenic factors. However, before any of these therapeutic approaches can be on the table, a thorough investigation of the genes involved in each cancer type needs to be done.

In classical approaches, this would require a long process of wet lab techniques, that would still be limited in sample number and precision. Thankfully, with the rise of the bioinformatic era, the scientific community is now in possession of a valuable arsenal of databases and tools. The latter are used to (i) mine the databases for samples related to a specific cancer type and see which genes were altered, (ii) highlight the mutations and map them against the predicted protein domains in each gene, (iii) evaluate the oncogenic potential of these mutations and (iv) predict which therapies would be better targeted in each alteration.

By 2020, breast cancer became the most frequent cancer diagnosed in women surpassing lung cancer, with invasive breast carcinoma accounting for 80% of all the diagnoses (CDC, 2020). This presses an issue to understand what is happening on a genetic level to customize therapies and help make patient treatments more precise.

In this study, we took interest in the gene alterations found in samples of Invasive Breast Carcinoma. A new understanding of breast cancer is provided by a variety of factors, including genomic changes, gene expression, histology, and immunological markers. These factors have prognostic and predictive significance in modern medicine. In fact, genomic alterations and gene expression patterns have enhanced the traditional care in early-stage tumorigenesis. Twelve genes were selected based on their roles as key players of significant pathways, crucial to the development and proliferation of invasive breast carcinoma: *BRCA1, BRCA2, PTEN, TP53, CDH1, STK11, CHEK2, BRIP1, ATM, PALB2, ERBB2 and EGFR.* In several malignancies, *TP53*, which is a tumor suppressor gene, has been found to be altered. Changes in the gene affect the expression of several genes that are either directly or indirectly regulated by p53. This causes DNA damage repair mechanisms to misfire, cell cycle arrest, chromatin remodeling, and apoptosis (Kaur, 2018). *TP53* is considered one of the important biomarkers and most frequently mutated gene in invasive breast carcinoma (Duffy MJ *et al.,* 2018).

In the 3,617 invasive breast carcinoma samples studied, *TP53* was shown to be the gene with the most mutations, with a percentage of 33% (Figure 1A). Most of the mutations in *TP53* were deep deletions which (shown in blue color in Figure 1A). They were coupled with truncating mutations which were putative drivers. Our findings matched those reported in the TCGA-Cancer Genome Atlas Network (cancer.gov, TCGA), citing *TP53* as the most frequently mutated gene in breast cancer. *TP53* encodes the transcription factor P53. Alterations in the gene affect the expression of several genes that are either directly or indirectly regulated by P53. This causes DNA damage repair mechanisms to collapse, cell cycle arrest, chromatin remodeling and apoptosis (Kaur, 2018).

As for other genes, numerous mutations of different types were detected. For instance, for the *ERBB2* gene, 15% of alterations were amplifications, as shown in blue color in Figure 3, with some missense mutations with unknown significance also recorded. As for the types of gene alterations in other genes, *CDH1* followed *ERBB2* and *TP53* with 11% gene alterations divided into deep deletions, amplification and missense mutations with either putative driver or unknown significance. *BRIP1* recorded 9% whereby the majority of mutations are amplification. *PTEN* recorded 7% with a majority of deep deletion. *PALB2* recorded 6%, *BRCA1* 3%, *BRCA2* 3%, *EGFR* 3%, *ATM* 1.7%, *STK* 11 1.3% and *CHEK2* 1.3%. Therefore, the genes that had the biggest percentage in amplification in their mutations were *ERBB2*, *BRIP1*, *PALB2*, *EGFR*, *BRCA1*, shown in blue color in Figure 3. *CDH1*, *PTEN*, and *BRCA2* are three genes with deep deletions alterations.

Based on these detected mutations, the subsequent analysis was to check for mutual exclusivity. Mutation patterns in gene pairs that are mutually exclusive refers to an epistatic interaction, where a mutation in one gene would have the same effects as a mutation in the other (Cisowski *et al.*, 2017). Different cooperating carcinogenic pathways are frequently activated when co-occurring driver genes are present. On the other hand, divergent and incompatible or redundant tumorigenic pathways are activated by driver genes that are mutually exclusive. According to our mutual exclusivity analysis (Table 1), *TP53* was found to be significantly mutually exclusive with *CDH1* and *PALB2* (p-value<0.001). This means that when there was a mutation in *TP53* it overrode the other mutations in *CDH1* and *PALB2*. Our findings also showed that *TP53* mutations can co-occur with *ERBB2* and *EGFR*. *BRIP1* co-occurs with *ERBB2* and *ATM*. Whereas it is mutually exclusive with *PALB2* and *EGFR* (Table 1). *BRCA1* co-occur with *ERBB2, BRIP1, TP53, BRCA2, EGFR and ATM*. Whereas it is mutually exclusive with *CDH1, PALB2* and *STK11*. *BRCA2* co-occur with *BRIP1, CDH1, STK11, TP53, ATM, PTEN* and *ERBB2*. *PTEN* co-occur with *TP53* only whereas it is mutually exclusive with *ERBB2* and *PALB2. ERBB2* co-occur with *EGFR* only and showed no mutual exclusivity with any of the 12 genes studied. *CHEK2* co-occur with *BRIP1* and *ERBB2* whereas it is mutually exclusive with *PALB2*, and *ATM*. *ATM* co-occur with *ERBB2* and is found to be mutually exclusive with *EGFR*.

Every mutation detected in the analyzed genes and categorized above were represented in a lollipop graph (Figure 4). A first example is *TP53*, in which four major mutations were the most prevalent, located at position 175, 248, 275 and 342. Based on the lollipop model, our results show that 66 out of 1,192 mutations from 66 patients were R248W/Q/G and 54 mutations were R273H. These two missense mutations are very likely oncogenic (cBioPortal). Further analysis with MutationTaster, a program that forecasts the impact of various gene mutations, and Ensembl proved our results for the specific R248W/Q/G TP53 mutation. An exact transcript (ENST 00000269305) also confirmed validity of our findings for the specific mutation R248W/Q/G. Using Ensembl, SNPs were detected in the *TP53* gene as well, such as SNP RS889884213. SNPs are considered to be possible markers of carcinogenesis (GaO *et al.*, 2019).

Nevertheless, and despite their reputation, not all mutations are oncogenic, as was also demonstrated by the findings of our study. For instance, the D1692N mutation in the *BRCA1* gene is thought to be likely, not surely, oncogenic (Table 2). Y1655 in *BRCA2* is also likely oncogenic. R181C in *TP53* is likely oncogenic. R175H in *TP53* is also likely oncogenic. Other mutations are neutral. For EGFR gene, R98Q and V5921 mutations were found to be neutral. For *ERBB2* gene, 4 mutations were found to be neutral: L755M, A1039T, V11281, and T306M. *ERBB2* has also 2 inconclusive mutations on the positions of G309A and G727A, along with F2042 mutation in *STK11*.

After highlighting the various key role genes, the alterations they might be subject to and their oncogenic potential, we sought to estimate and predict the overall survival rate in patients with mutations in the 12 genes (altered group). Survival rates were low for this group, dropping significantly from 100% in month 0 to 50% in month 130, reaching extremely low survival rates of 10% at the phase of 290 months to 350, until survival was at 0% survival after 360 months.

These low survival rates are indicative of the aggressivity of these mutations if they occur in the 12 key role genes.

Fortunately, however, a series of targeted therapies for specific mutations in our 12 studied genes were generated using PharmGKB (Table 3). Although many mutations still don't have a specific therapeutic agent to target them, many others do have. Starting by the mutation D1692N in *BRCA1* gene, Talazoparib and Olaparib are targeted drugs used. Talazoparib is an oral poly adenosine diphosphate polymerase inhibitor used in the treatment if advanced stages of breast cancer, specifically the tumors with mutations in *BRCA* genes (Rilliton *et al.*, 2018). It is similar to Olaparib, which is the first-class PARP inhibitor used also in BRCA mutated tumors, PARP being a cellular enzyme involved in numerous functions, including DNA repair. It is to be noted that Olaparib is also used in *CHEK2* gene mutations specifically for R346H mutations. Rucaparib is also a PARP inhibitor taken orally for the Q1146 mutation in *PALB2*. Another example to which a targeted therapy exists is the mutation R130G in *PTEN*, with GSK2636771 and AZD8186 as targeted therapy options (based on cBioPortal). Neratinib, a tyrosine kinase inhibitor, is used in the treatment of *ERBB2* mutations, along with others. The HER family of transmembrane tyrosine kinase receptors (HER1 and HER2) are irreversibly inhibited by Neratinib (Chan *et al.,* 2020) Ado-trastuzumab Emtansine, an antibody-drug conjugate targeting the human epidermal growth factor receptor 2 (HER2) with antitumor properties, found in the trastuzumab drug, with cytotoxic effect on the microtubule inhibitory agent DM1 (Peddi, & Hurvitz, 2014). Ado-trastuzumab Emtansine is shown to be efficient on D7694 mutations in the *ERBB2* gene according to our bioinformatic predictions. Many other mutations in *TP53* (R181C), *CDH1* (D288N), *STK11* (S216F) and *EGFR* (G724S), are still under targeted drug exploration (cBioPortal). Therefore, no specific assigned drug is incorporated in the database of neither cBioPortal nor PharmGKB.

CHAPTER FIVE

CONCLUSIONS AND PERSPECTIVES

The tumorigenesis of invasive breast cancer can be made clearer by the investigation of these different genes, their interactions and oncogenic potential. Even if their interactions are understood, it is difficult to specifically target them in the presence of complex interactions. Nevertheless, starting a customized treatment for each cancer patient can be made easier by being aware of the gene alterations this person carries, their types, mutual exclusivity and oncogenic potentials. The bioinformatics methods employed here made it possible to quickly access a large number of genes crucial and key players in breast cancer. The availability of large datasets and the tools to mine them has made faster and easier to research new treatments, validate the types of mutations detected in each gene, and examine the mutations in the genes of interest. Knowing the complexity of genetic interactions, further studies should be made by testing and analyzing more genes and their oncogenic implications on breast cancer patients. This will leave a new space for the discovery of more targeted cancer therapies in the hopes of lengthening the lifespan of cancer patients. Finally, geographical statistics can be made to compare the prevalence of the studied genes based on locations, ethnicity, and external environmental stressors. This will give a holistic view of Invasive Breast Carcinoma leading to a more successful diagnosis, prognosis, and treatment modalities.

REFERENCES

Abbas, Z., & Rehman, S. (2018). *An Overview of Cancer Treatment Modalities*.

Abdelwahad, Y. A. (2017). Male Breast Cancer: Epidemiology and Risk Factors. *Semin Oncol, 44*(4), 267–272.

Abdulrahman, G., & Rahman, G. (2012). Epidemiology of Breast Cancer in Europe and Africa. *J. Cancer Epidemiol., 20121*(2).

Arthur, R., Wang, Y., & Ye, K. (2017). Association between lifestyle, menstrual/reproductive history, and histological factors and risk of breast cancer in women biopsied for benign breast disease. *Breast Cancer Res Treat*.

Assi, H., Khoury, K., Dbouk, H., Khalil, L., Mouhieddine, T., & El Saghir, N. (2013). Epidemiology and prognosis of breast cancer in young women. *J Thorac Dis*, *5*(Suppl 1), S2–S8.

Azubuike, S., Muirhead, C., Hayes, L., & McNally, R. (2018). Rising global burden of breast cancer: The case of sub-Saharan Africa (with emphasis on Nigeria) and implications for regional development: A review. *World J Surg Oncology*, *16*(1), 63.

Balekouzou, A., & Yin, P. (2017). Reproductive risk factors associated with breast cancer in women in Bangui: A case–control study. *BMC Womens Health*.

Beaber, E., Buist, D., Barlow, W., & Malone, K. (2014). Recent oral contraceptive use by formulation and breast cancer risk among women 20 to 49 years of age. *Cancer Res*, *74*(15), 4078–4089.

Beral, V. (2003). Million Women Study Collaborators Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet*, 419–427.

Beral, V., Bull, D., & Doll, R. (1997). Breast cancer and hormone replacement therapy: Collaborative reanalysis of data from 51 epidemiological studies of 52, 705 women with breast cancer and 108, 411 women without breast cancer. *Lancet*, 1047–1059.

Bhadoria, A., & Kapil, U. (2013). Reproductive factors and breast cancer: A case-control study in tertiary care hospital of North India. *Indian J Cancer*, *50*(4), 316–321.

Bhosle, J., & Hall, G. (2009). Principles of cancer treatment by chemotherapy. *Surgery (Oxford)*, *27*(4), 173–177.

Brasky, T., & Li, Y. (2013). Pregnancy-related characteristics and breast cancer risk. *Cancer Causes Control*.

Calle, E., Rodriguez, C., Walker-Thurmond, K., & Thun, M. (2003). Overweight, obesity and mortality from cancer in a prospectively studied cohort of US adults. *New England Journal of Medicine*, *348*, 1625–1638.

Carioli, G., Malvezzi, M., Rodriguez, T., & Bertuccio, P. (2018). Trends and predictions to 2020 in breast cancer mortality: Americas and Australasia. *Breast*, *37*, 163–169.

Centers for Disease Control and Prevention. (2002). Annual smoking attributable mortality, years of potential life lost and economic costs- United States 1995-1999. *JAMA*, *287*(18), 2355–2356.

Chambers, A. F., Groom, A. C., & Macdonald, I. C. (2002). Dissemination and ghrowth of cancer cells in metastatic sites. *Nature Review Cancer*, *2*, 563–572.

Clavel-Chapelon, F., & Launoy, G. (2995). Reproductive factors and breast cancer risk: Effect of age at diagnosis. *Ann Epidemiol*, *5*(4), 315–320.

Cobain, E., & Milliron, K. (2016). Updates on breast cancer genetics: Clinical implications of detecting syndromes of inherited increased susceptibility to breast cancer. *Semin Oncol*, *43*(5).

Colditz, G., DeJong, W., Hunter, D., Trichopoulos, D., & Willett, W. (1996). Harvard report on cancer prevention. *Cancer Causes Control*, *7*, S1–S55.

Curry, S., Byers, T., & Hewitt, M. (2003). Fulfilling the Potential of Cancer Prevention and Early Detection. *National Academy Press*.

Dai, Q., & Liu, B. (2009). Meta-analysis of the risk factors of breast cancer concerning reproductive factors and oral contraceptive use. *Front Med China*.

Denoix, P. (1952). Nomenclature classification des cancers. *Bull Inst Nat Hyg*, *7*, 743–748.

DeSantis, C., Ma, J., Bryan, L., & Jemal, A. (2014). Breast Cancer Statistics, 2013. *CA Cancer J Clin*, *64*(1), 52–62.

Desantis, C., Ma, J., Goding Sauer, A., Newman, L., & Jemal, A. (2017). Breast Cancer Statistics, 2017, racial disparity in mortality by state. *CA Cancer J Clin*, *67*(6), 439–448.

DeVita, V., & Chu, E. (2008). A history of cancer chemotherapy. *Cancer Research*, *68*(21), 8643–8653.

Doll, R., Peto, R., Wheatley, K., Gray, R., & Sutherland, I. (1994). Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ*, *309*, 901–911.

Dorai, T., & Aggarwal, B. (2004). Role of chemopreventive agents in cancer therapy. *Cancer Letters*, *215*(2), 129–140.

Eger, A., & Mikulits, W. (2005). Models of epithelial-mesenchymal transition. *Drug Discovery Today: Dis Models*, *2*, 57–63.

Ellis, P. (2003). *WHO Classification of Tumours. Pathology and Genetics of Tumours of the Breast and Female Genital Organs*.

Ezzati, M., Lopez, A., Rodgers, A., Vander Hoorn, S., & Murray, C. (2002). Selected major risk factors and global and regional burden of disease. *Lancet*, *360*, 1347–1360.

Ferley, J., SoerjomataramI, I., Ervik, M., Dikshit, R., & Eser, S. (2013). Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11. *International Agency for Research on Cancer*, *1.0*.

Fioretti, F., Tavani, A., & Bosetti, C. (1999). Risk factors for breast cancer in nulliparous women. *Br J Cancer*.

Fogh, S. (2010). Hypofractionated stereotactic radiation therapy: An effective therapy for recurrent high-grade gliomas. *Journal of Clinical Oncology.*, *28*(18), 3048–3053.

Freund, C., Mirabel, L., Annane, K., & Mathelin, C. (2005). Breastfeeding and breast cancer. *Gynecol Obstet Fertil*, *33*(10).

Gabriel, N., James, L., & Carl, J. (2017). *AJCC cancer staging manual* (8th edition).

Ghoncheh, M., Mohammadian-Hafshejani, A., & Salehiniya, H. (2015). Incidence and Mortality of Breast Cancer and their Relationship to Development in Asia. *Asian Pacific Journal of Cancer Prev*, *16*(14), 6081–6087.

Ghoncheh, M., Pournamdar, Z., & Salehiniya, H. (2016). Incidence and Mortality and Epidemiology of Breast Cancer in the World. *Asian Pacific Journal of Cancer Prev*, *17*(s3), 43–46.

Giordano, S., Buzdar, A., & Hortobagyi, G. (2002). Breast Cancer in Men. *Ann Intern Med*, *137*(8), 678–687.

Giovannucci, E., Stampfer, M., Colditz, G., Rimm, E., & Willett, W. (1993). Folate, methionine, and alcohol intake and risk of colorectal adenoma. *Journal of National Cancer Institute*, *85*, 875–884.

Godet, I., & Gilkes, D. (2017). BRCA1 and BRCA2 mutations and treatment strategies for breast cancer. *Integr Cancer Sci Ther*, *4*(1), 1–17.

Goldblum, G., Weiss, S., & Foipe, A. (2013). *Enzinger and Weiss's Soft Tissue Tumors E-Book*. ElSevier Health Sciences.

Greene, F. L., & Sobin, L. H. (2008). The Staging of Cancer: A Retrospective and prospective Appraisal. *American Cancer Society*.

Hassen-Khodja. (2004). Gamma knife and linear accelerator stereotactic radiosurgery. *Agence d'évaluation Des Technologies et Des Modes d'intervention En Santé*.

Hortobagyi, G., de la Garza Salazar, J., & Pritchard, K. (2005). The global breast cancer burden: Variations in epidemiology and survival. *Clin Breast Cancer*, *6*(5), 391–401.

Innes, K., & Byers, T. (2004). First pregnancy characteristics and subsequent breast cancer risk among young women. *Int J Cancer*, *112*(2), 306–311.

International Agency for Research on Cancer. (2002). *Weight Control and Physical Activity*. *6*(IARC Handbook on Cancer Prevention-Lyon).

Jeong, S., An, Y., & Choi, J. (2017). Risk Reduction of Breast Cancer by Childbirth, Breastfeeding, and Their Interaction in Korean Women: Heterogeneous Effects Across Menopausal Status, Hormone Receptor Status, and Pathological Subtypes. *J Prev Med Public Health*, *50*(6), 401–410.

Jieun, K., & Min Jung, K. (2019). Introduction of a New Staging System of Breast Cancer for Radiologists: An Emphasis on the Prognostic Stage. *Korean J Radiol*, *20*(1), 69–82.

Johnstone, T. C. (2014). *The Crystal Structure of Oxaliplatin: A Case of Overlooked Pseudo Symmetry*.

Kano, H. (2017). Estimating the risks of adverse radiation effects after gamma knife radiosurgery for arteriovenous malformations. *Stroke*, *48*(1), 84–90.

Kelland, L. (2007). The resurgence of platinum-based cancer chemotherapy. *Nature Reviews*.

Kim, Y., Yoo, K.-Y., & Goodman, M. (2015). Difference in Incidence, Mortality and Survival of Breast Cancer by Regions and Countries in Asia and Contributing Factors. *Asian Pac J Cancer Prev*, *16*(7), 2857–2870.

Klein, C. A. (2009). Parallel progression of primary tumours and metastases. *Nature Review Cancer*, *9*, 302–312.

Laamiri, F., Bouayad, A., & Hasswane, N. (2015). Breast Cancer of Different Age Groups: Moroccan Data? *Open J Obstet Gynecol*, *5*(2), 79–87.

Ma, H., Henderson, K., & Sullivan-Halley, J. (2010). Pregnancy-related factors and the risk of breast carcinoma in situand invasive breast cancer among postmenopausal women in the California Teachers Study cohort. *Breast Cancer Research*.

Mahouri, K., Dehghani, Z. M., & Zare, S. (2007). Breast cancer risk factors in south of Islamic Republic of Iran: A case-control study. *East Mediterr Health J.*, *13*(6), 1265–1273.

Marchbanks, P., & Macdonald, J. (2002). Oral contraceptives and the risk of breast cancer. *N Engl Journal of MEd*.

Martinez, M., Heddens, D., Earnest, DL., Bogert, CL., Roe, D., & Einspahr J. (1999). Physical activity, body mass index, and prostaglandin E2 levels in rectal mucosa. *J Natl Cancer Inst*, *91*, 950–953.

Mastalarz, H., Mastalarz, A., Kochel, A., & Regiec, A. (2020). Synthesis of Platinum(II) Complexes with Some 1-Methylnitropyrazoles and In Vitro Research on Their Cytotoxic Activity. *Pharmaceuticals*, *13*(12), 433.

McKeown-Eyssen, G. (1994). Epidemiology of colorectal cancer revisited: Are serum triglycerides and/or plasma glucose associated with risk? *Cancer Epidemiology Biomarkers Prev*, *3*, 687–695.

Mehlen, P., & Puisieux, A. (2006). Metasasis: A question of life or death. *Nature Review Cancer*, *6*, 449–458.

Melbye, M., & Wohlfhart, J. (1999). Preterm delivery and risk of breast cancer. *Br J Cancer*, *80*(3–4), 609–613.

Metcalfe, K., Finch, A., & Poll, A. (2009). Breast cancer risks in women with a family history of breast or ovarian cancer who have tested negative for a BRCA1 or BRCA2 mutation. *Br J Cancer*, *100*(2), 421–425.

Momenimovahed, Z., & Salehiniya, H. (2019). Epidemiological characteristics of and risk factors for breast cancer in the world. *Dove Med Press : Breast Cancer*, *11*, 151–164.

Monneret, C. (2011). Platinum anticancer drugs. From serendipity to rational design. *Annales Pharmaceutiques Francaises*, *69*(6), 286–295.

National Center for Biotechnology Information. (2021). *PubChem Compound Summary for CID 441203, cis-Platin*.

Nguyen, J., Le, Q., & Duong, B. H. (2016). A matched case-control study of risk factors for breast cancer risk in Vietnam. *Int J. Breast Cancer*.

Okolsky, O. (2010). *Medical cyber physical systems*.

Olivier, M., Langerod, A., Carrieri, P., Bergh, J., Klaar, S., & Eyfjord, J. (2006). The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. *Clinical Cancer Reasearch*, *12*, 1157–1167.

Paffenbarger, J. R., Hyde, R., Wing, A., Lee, I.-M., & Jung, D. (1993). The association of changes in physical activity level and other lifestyle characteristics with mortality among men. *N England Journal of Medicine*, *328*, 538–545.

Parkin, D., & Fernandez, L. (2006). Use of statistics to assess the global burden of breast cancer. *Breast J*, *12*(s1), s70–s80.

Pate, R., & Wilmore, J. (1995). Physical activity and public health. A recommendaton from the centers for disease control and prevention and the American College of Sports Medicine. *JAMA*, *273*(5), 402–407.

Prat, A., & Pineda, E. (2015). Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast*, *24*(Suppl 2).

Rakha. (2008). *Histologic grading is an independent prognostic factor in invasive lobular carcinoma of the breast.*

Rojas, K., & Stuckey, A. (2016). Breast Cancer Epidemiology and Risk Factors. *Clin Obstetrics and Gynecol*, *59*(4), 651–672.

Roman, M., & Quintana, M. (2017). Cumulative risk of breast cancer screening outcomes according to the presence of previous benign breast disease and family history of breast cancer: Supporting personalised screening. *Br J Cancer*, *116*(11), 1480–1485.

Siddik, Z. H. (2003). Cisplatin: Mode of cytotoxic action and molecular basis of resistance. *Oncogene*, *22*.

Slattery, M., Potter, J., Samowitz, W., Schaffer, D., & Leppert, M. (1999). Methylenetetrahydrofolate reductase, diet, and risk of colon cancer. *Cancer Epidemiology Biomarkers Prev*, *8*(6), 513–518.

Tamakoshi, K., Yatsuya, H., & Wakai, K. (2005). Study Group Impact of menstrual and reproductive factors on breast cancer risk in Japan: Results of the JACC study. *Cancer Sci*, *96*(1), 57–62.

Thakur, P., Seam, R., Gupta, M., Sharma, M., & Fotedar, V. (2017). Breast cancer risk factor evaluation in a Western Himalayan state: A case–control study and comparison with the Western World. *South Asian J Cancer.*, *6*(3), 106–109.

Vatten, L., & Romundstad, P. (2002). Pre-eclampsia in pregnancy and subsequent risk for breast cancer. *Br J Cancer*.

Waks, A., & Winer, E. P. (2019). Breast Cancer Treatment: A Review. *JAMA*, *321*(3).

Weigelt, B., & Reis-Filho, J. S. (2009). *Histological and molecular tyes of breast cancer: Is there a unifying taxonomy?*

Wu, H.-C., Chang, D.-K., & Huang, C.-T. (2006). Targeted therapy for cancer. *Journal of Cancer Molecules.*, *2*(2), 57–66.

Youlden, D., Cramb, S., Yip, C., & Baade, P. (2014). Incidence and mortality of female breast cancer in the Asia-Pacific region. *Cancer Biol Med*, *11*(2), 101–115.

Zendehdel, M., Niakan, B., & Keshtkar, A. (2018). Subtypes of Benign Breast Disease as a Risk Factor for Breast Cancer: A Systematic Review and Meta-Analysis Protocol. *Iran J Med*, *43*(1).

Zhang, S., Hunter, D., Hankinson, S., & Colditz, G. (1999). A prospective study of folate intake and the risk of breast cancer. *JAMA*, *281*, 1632–1637.

Zhang, S., Willett, W., Selhub, J., & Hunter, D. (2003). Plasma folate, vitamin b6, vitamin b12, and homocysteine and risk of breast cancer. *Journal of National Cancer Institute*, *95*(5), 373–380.

World Health Organization Tobacco. http://www.who.int/mediacentre/factsheets/fs339/en/. Published 2013. Accessed June 2, 2014.

Park, J. H., Moon, J. H., Kim, H. J., Kong, M. H., & Oh, Y. H. (2020). Sedentary Lifestyle: Overview of Updated Evidence of Potential Health Risks. Korean journal of family medicine, 41(6), 365–373.

Brenner, D. R., Poirier, A. E., Grundy, A., Khandwala, F., McFadden, A., & Friedenreich, C. M. (2017). Cancer incidence attributable to inadequate physical activity in Alberta in 2012. CMAJ open, 5(2), E338–E344.

Scheffer, D. D. L., & Latini, A. (2020). Exercise-induced immune system response: Anti-inflammatory status on peripheral and central organs. Biochimica et biophysica acta. Molecular basis of disease, 1866(10), 165823. https://doi.org/10.1016/j.bbadis.2020.165823 Willett WC, Tamimi R, Hankinson SE, Hazra A, Eliassen AH, Colditz GA. Chapter 18: Nongenetic Factors in the Causation of Breast Cancer, in Harris JR, Lippman ME, Morrow M, Osborne CK. Diseases of the Breast, 5th edition, Lippincott Williams & Wilkins, 2014.

Kaur RP, Vasudeva K, Kumar R, Munshi A. Role of p53 Gene in Breast Cancer: Focus on Mutation Spectrum and Therapeutic Strategies. Curr Pharm Des. 2018;24(30):3566-3575. doi: 10.2174/1381612824666180926095709. PMID: 30255744.

Cisowski J, Bergo MO. What makes oncogenes mutually exclusive? Small GTPases. 2017 Jul 3;8(3):187-192. doi: 10.1080/21541248.2016.1212689. Epub 2016 Jul 14. PMID: 27416373; PMCID: PMC5584735.

Duffy MJ, Synnott NC, Crown J. Mutant p53 in breast cancer: potential as a therapeutic target and biomarker. Breast Cancer Res Treat. 2018 Jul;170(2):213-219. doi: 10.1007/s10549-018-4753-7. Epub 2018 Mar 21. PMID: 29564741.

Chan A, Moy B, Mansi J, Ejlertsen B, Holmes FA, Chia S, Iwata H, Gnant M, Loibl S, Barrios CH, Somali I, Smichkoska S, Martinez N, Alonso MG, Link JS, Mayer IA, Cold S, Murillo SM, Senecal F, Inoue K, Ruiz-Borrego M, Hui R, Denduluri N, Patt D, Rugo HS, Johnston SRD, Bryce R, Zhang B, Xu F, Wong A, Martin M; ExteNET Study Group. Final Efficacy Results of Neratinib in HER2-positive Hormone Receptor-positive Early-stage Breast Cancer From the Phase III ExteNET Trial. Clin Breast Cancer. 2021 Feb;21(1):80-91.e7. doi: 10.1016/j.clbc.2020.09.014. Epub 2020 Oct 6. PMID: 33183970.

Peddi, P. F., & Hurvitz, S. A. (2014). Ado-trastuzumab emtansine (T-DM1) in human epidermal growth factor receptor 2 (HER2)-positive metastatic breast cancer: latest evidence and clinical potential. Therapeutic advances in medical oncology, 6(5), 202–209.

Schon, K., & Tischkowitz, M. (2018). Clinical implications of germline mutations in breast cancer: TP53. Breast cancer research and treatment, 167(2), 417–423.

Slamon DJ, Godolphin W, Jones LA. et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science. 1989;244(4905):707–712.

Tateishi M, Ishida T, Mitsudomi T. et al. Prognostic value of c-erbB-2 protein expression in human lung adenocarcinoma and squamous cell carcinoma. Eur J Cancer. 1991;27(11):1372–1375.

Holbro T, Civenni G, Hynes NE. The ErbB receptors and their role in cancer progression. Exp Cell Res. 2003;284(1):99–110.

Corso G, Intra M, Trentin C, Veronesi P, Galimberti V. (2016) CDH1 germline mutations and hereditary lobular breast cancer. *Fam Cancer* 15:215–219.

Hamilton, M. T., Healy, G. N., Dunstan, D. W., Zderic, T. W., & Owen, N. (2008). Too Little Exercise and Too Much Sitting: Inactivity Physiology and the Need for New Recommendations on Sedentary Behavior. Current cardiovascular risk reports, 2(4), 292–298.

Apostolou, P., & Papasotiriou, I. (2017). Current perspectives on CHEK2 mutations in breast cancer. Breast cancer (Dove Medical Press), 9, 331–335.

Lynce, F., & Isaacs, C. (2016). How Far Do We Go with Genetic Evaluation? Gene, Panel, and Tumor Testing. American Society of Clinical Oncology educational book. American Society of Clinical Oncology. Annual Meeting, 35, e72–e78.

Hanauer,David A. *et al.*, Current Molecular Medicine, Volume 7, Number 1, 2007.
Elisabeth Remy; Sandra Rebouissou; Claudine Chaouiya; Andrei Zinovyev; François Radvanyi; Laurence Calzone, Cancer Res (2015), Volume 75.

Daub, J.T., Amini, S., Kersjes, D.J.E. et al. A systematic analysis of genetic interactions and their underlying biology in childhood cancer. Commun Biol 4, 1139 (2021).